2009 International Fuzzy Systems Association World Congress
2009 European Society for Fuzzy Logic and Technology Conference

Proceedings



Lisbon, Portugal, 20th - 24th July, 2009

Edited by J. P. Carvalho, D. Dubois, U. Kaymak and J. M. C. Sousa

IFSA/EUSFLAT 2009

INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

ERASMUS UNIVERSITEIT ROTTERDAM

idmec IST engenharia mecânica

inesc id lisboa

2009 International Fuzzy Systems Association WORLD CONGRESS
2009 European Society for Fuzzy Logic and Technology CONFERENCE

Proceedings



Lisbon, Portugal, 20th - 24th July, 2009

Edited by J. P. Carvalho, D. Dubois, U. Kaymak and J. M. C. Sousa

IFSA/EUSFLAT 2009

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR
TAP PORTUGAL
ANA Aeroportos de Portugal
FUNDAÇÃO LUSO-AMERICANA
BPI
Lisboa
Caixa Geral de Depositos
FUNDAÇÃO CALOUSTE GULBENKIAN
INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa
Erasmus
ERASMUS UNIVERSITEIT ROTTERDAM
idmec IST engenharia mecânica
inesc id lisboa



The Belém Tower (in Portuguese "Torre de Belém") is a fortified tower located in the Belém district of Lisbon, Portugal. The Belém Tower was built in the early sixteenth century in the Portuguese late-Gothic style, the Manueline, to commemorate Vasco da Gama's expedition (discovered the see way to India). This defensive, yet elegant construction has become one of the symbols of the city, a memorial to Portuguese power as it was during the Age of Discovery, which started in the 15th Century and continuing into the 17th Century. In 1983, it was classified, together with the nearby Mosteiro dos Jerónimos (Jerónimos Monastery), as a UNESCO World Heritage Site.

Cover design: Susana M. Vieira.

Proceedings of the 2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA–EUSFLAT 2009).

ISBN: 978-989-95079-6-8

This publication was supported by the Program "Fundo de Apoio à Comunidade Científica" from the "Fundação para a Ciência e a Tecnologia" (FCT).

© 2009, International Fuzzy Systems Association (IFSA)
2009, European Society for Fuzzy Logic and Technology (EUSFLAT)

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieven system, without permission in writing of the copyright holder.

Table of Contents

An Approach to Interval-Valued R-Implications and Automorphisms 1
Benjamin Bedregal, Graçaliz Dimuro and Renata Reiser

Recognition and Teaching of Robot Skills by Fuzzy Time-Modeling 7
Rainer Palm, Bourhane Kadmiry, Boyko Iliev and Dimiter Driankov

A Consensus Model for Group Decision Making with Incomplete Unbalanced Fuzzy Linguistic Preference Relations 13
Francisco Javier Cabrerizo, Juan Manuel Moreno and Enrique Herrera-Viedma

Real-Time Segmentation of Moving Objects in H.264 Compressed Domain with Dynamic Design of Fuzzy Sets 19
Cayetano Solana-Cipres, Luis Rodriguez-Benitez, Juan Moreno-Garcia, Luis Jimenez and Gerardo Fernandez-Escribano

On the Benefits of Representing Music Objects as Fuzzy Sets 25
Klaas Bosteels, Elias Pampalk and Etienne Kerre

Type-II Fuzzy Possibilistic C-Mean Clustering 30
M. H. Fazel Zarandi, Marzie Zarinbal and I. B. Turksen

Opposite Fuzzy Sets with Applications in Image Processing 36
Hamid R. Tizhoosh

Genetic Cooperative-Competitive Fuzzy Rule Based Learning Method using Genetic Programming for Highly Imbalanced Data-Sets 42
Alberto Fernández, Francisco José Berlanga, Maria Jose Del Jesus and Francisco Herrera

New Fuzzy Color Clustering Algorithm Based on hsl Similarity 48
Vasile Patrascu

Multi-Dimensional Fuzzy Transforms for Attribute Dependencies 53
Ferdinando di Martino, Vincenzo Loia and Salvatore Sessa

Interval-Valued Fuzzy Propositions. An Application to the L-Fuzzy Contexts with Absent Values 58
Cristina Alcalde, Ana Burusco and Ramon Fuentes-González

On the Structure of the k-Additive Fuzzy Measures 64
Elias F. Combarro and Pedro Miranda

Type-2 Fuzzy Trust Aggregation 70
John T. Rickard, Sam Hamilton and Wendy Hamilton

On Vertices of the k-Additive Monotone Core 76
Pedro Miranda and Michel Grabisch

Resolution of Fuzzy Relation Equations with Sup-Inf Composition over Complete Brouwerian Lattices — a review 82
Xue-ping Wang and Feng Sun

Aggregation Operators for Conditional Crispness 88
Doretta Vivona and Maria Divari

A New Conditioning Rule, Its Generalization and Evidential Reasoning 92
Koichi Yamada, Vilany Kimala and Muneyuki Unehara

Consistency of Pair-wise Comparison Matrix with Fuzzy Elements 98
Jaroslav Ramík

Fuzzy Continuous Evaluation in Training Systems Based on Virtual Reality 102
Ronei Moraes and Liliane Machado

A Comparative Analysis of Symbolic Linguistic Computational Models 108
Rosa M. Rodriguez, Luis Martinez and Macarena Espinilla

Universal Integrals Based on Level Dependent Capacities 114
Erich Peter Klement, Anna Kolesárová, Radko Mesiar and Andrea Stupňanová

A Conceptual Framework for Understanding a Fuzzy System 119
José M. Alonso and Luis Magdalena

An Experimental Study on the Interpretability of Fuzzy Systems 125
José M. Alonso and Luis Magdalena

Games and Capacities on Partitions 131
Michel Grabisch

The Minimization of the Risk of Falling in Portfolios under Uncertainty 137
Yuji Yoshida

Approximations by Interval, Triangular and Trapezoidal Fuzzy Numbers 143
Chi-Tsuen Yeh

Fuzzy c-Means Herding 149
Thomas A. Runkler

Interval AHP for Group of Decision Makers 155
Tomoe Entani

A Monge Algorithm for Computing the Choquet Integral on Set Systems 161
Ulrich Faigle and Michel Grabisch

Decomposition of Interval-Valued Fuzzy Morphological Operations by Weak [alpha_1,alpha_2]-cuts 167
Tom Mélange, Mike Nachtegael, Etienne Kerre and Peter Sussner

An Alternative Operationalization of Fuzzy Consideration Set. Application to Tourism 173
Luisa Lucila Lazzari, Patricia Inés Mouliá and Mariano Eriz

Ant Colony Optimization for Distributed Routing Path Optimization in Optical Burst-Switched Networks 178
Joao Pedro, Joao Pires and João Paulo Carvalho

Non-Conjunctive and Non-Disjunctive Uninorms in Atanassov's Intuitionistic Fuzzy Set Theory 184
Glad Deschrijver

Fuzzy Descriptions Logics with Fuzzy Truth Values 189
Fernando Bobillo and Umberto Straccia

A New Approach to Fuzzy Location of Cephalometric Landmarks in Craniofacial Superimposition 195
Oscar Ibáñez, Oscar Cordón, Sergio Damas, Sergio Guadarrama and José Santamaría

Incorporating Fuzziness in Spatial Susceptible-Infected Epidemic Models 201
Jan Baetens and Bernard De Baets

Recurrent Neural Kalman Filter Identification and Indirect Adaptive Control of a Continuous Stirred Tank Bioprocess 207
Ieroham Baruch and Carlos Román Mariaca Gaspar

A Multi-granular Linguistic Promethee Model 213
Nesrin Halouani, Luis Martínez, Habib Chabchoub, Jean-Marc Martel and Jun Liu

Generating Minimum Dispersion Densities from an Interval-Valued Fuzzy Set 219
Carol Walker, Elbert Walker and Ronald Yager

Approximate Reasoning in Surgical Decisions 225
Elisabeth Rakus-Andersson

Chaos Induced by Turbulent and Erratic Functions 231
Heriberto Román-Flores, Yurilev Chalco-Cano and Geraldo Nunes-Silva

Verbalizing Time-series Data: With an Example of Stock Price Trends 234
Ichiro Kobayashi and Naoko Okumura

Differential Equations based on Fuzzy Rules 240
Marina R. B. Dias and Laécio C. Barros

Sparsely Connected Autoassociative Fuzzy Implicative Memories for the Reconstruction of Large Gray-Scale Images ... 247
Marcos Eduardo Valle

A Restriction Level Approach for the Representation and Evaluation of Fuzzy Association Rules 253
Miguel Delgado, M. Dolores Ruiz and Daniel Sanchez

Semantic Proximity Between Queries and the Empty Answer Problem 259
Patrick Bosc, Carmen Brando, Allel Hadjali, Hélène Jaudoin and Olivier Pivert

A Fuzzy Delay Differential Equation Model for HIV Dynamics 265
Rosana M. Jafelice, Laécio C. Barros and Rodney C. Bassanezi

Dispersal of Leaf-Cutting Ants: fuzzy mathematical modeling, numerical approximation and simulations 271
Rosana Motta Jafelice, Cesar Guilherme Almeida, João Frederico C. A. Meyer and Heraldo L. Vasconcelos

Continuous R-Implications 277
Balasubramaniam Jayaram and Michal Baczynski

Fuzzy Voxel Object 282
Derek Anderson, Robert Luke, Erik Stone and James Keller

Fuzzy Logic-Based Image Processing Using Graphics Processor Units 288
Robert Luke, Derek Anderson, James Keller and Simon Coupland

Dominance-based Rough Set Analysis for Uncertain Data Tables 294
Tuan-Fang Fan, Churn-Jung Liau and Duen-Ren Liu

Overlap Index, Overlap Functions and Migrativity 300
Humberto Bustince, Javier Fernandez, Radko Mesiar, Javier Montero and Raul Orduna

On 'Family Resemblances' with Fuzzy Sets 306
Enric Trillas, Claudio Moraga and Alejandro Sobrino

Convergence Theorems for Generalized Random Variables and Martingales 312
Andrew Pinchuck

Production and Transportation Planning — A fuzzy Approach for Minimizing the Total Cost 317
Heinrich Rommelfanger

The Use of Interval-Valued Probability Measures in Fuzzy Linear Programming: a constraint set approach 321
Phantipa Thipwiwatpotjana and Weldon Lodwick

Human Perceptions versus Computational Perceptions in Computational Theory of Perceptions 327
Gracian Trivino and Alejandro Sobrino

A Characterization of Residual Implications Derived from Uninorms 333
Isabel Aguiló, Jaume Suñer and Joan Torrens

A Logic-Based Approach for Evaluating Interpretability of Fuzzy Rule-Based Classifiers 339
Corrado Mencar, Ciro Castiello and Anna M. Fanelli

Cloud Based Design Optimization 345
Martin Fuchs

Local Models for the Analysis of Spatially Varying Relationships in a Lignite Deposit....351
Bulent Tutmez, A. Erhan Tercan, Uzay Kaymak and Christopher D. Lloyd

Fuzzy Integrals over Complete Residuated Lattices 357
Antonin Dvorak and Michal Holcapek

A Link Between the 2-additive Choquet Integral and Belief Functions 363
Brice Mayag, Michel Grabisch and Christophe Labreuche

Agent-based Fuzzy Constraint-directed Negotiation Mechanism for Planning and Scheduling in Supply Chain....369
Bo-Ruei Kao, K. Robert Lai and Menq-Wen Lin

Learning and Classification of Events in Monitored Environments 375
Javier Albusac, Jose Jesús Castro, Lorenzo Manuel López López, David Vallejo and Luis Jiménez

A Hybrid Method for Soccer Video Events Retrieval Using Fuzzy Systems 381
Ehsan Lotfi, M. Akbarzadeh-T., Hamid Reza Pourreza, Mona Yaghoubi and F. Dehghan

Pointwise Defined CRI-Based Aggregation Distributive Operators Are Trivial 386
Siegfried Gottwald and Philipp Hüsam

A Possibilistic Approach to Bottleneck Combinatorial Optimization Problems with Ill-Known Weights....390
Adam Kasperski and Pawel Zielinski

Web Usage Mining: users' navigational patterns extraction from web logs using ant-based clustering method 396
Kobra Etminani, Mohammad Reza Akbarzadeh Tootoonchi and Noorali Raeeji Yanehsari

Multiplicative and Implicative Importance Weighted Averaging Aggregation Operators with Accurate Andness Direction 402
Henrik Legind Larsen

Training a Personal Alert System for Research Information Recommendation 408
German Hurtado Martin, Chris Cornelis and Helga Naessens

On Reduced Semantics for Fuzzy Predicate Logics....414
Pilar Dellunde

Fuzzy Clustering for Finding Fuzzy Partitions of Many-Valued Attribute Domains in a Concept Analysis Perspective 420
Yassine Djouadi, Basma Alouane and Henri Prade

On Possibilistic Sequencing Problems with Fuzzy Parameters....426
Adam Kasperski and Pawel Zielinski

The Bandler-Kohout Subproduct as a Suitable Inference 432
Martin Stepnicka and Balasubramaniam Jayaram

Early Start Can Inhibit Learning: Towards A New Explanation....438
Olga Kosheleva

Intensity Assessment of Pedestrian Collisions in City of Mashhad based on Fuzzy Probabilities 443
Farzaneh Tatari and Mohammad-R. Akbarzadeh-T

Number-free Mathematics Based on T-norm Fuzzy Logic....449
Libor Behounek

Uncertainty Can Decrease Privacy: an observation 455
Karen Villaverde and Olga Kosheleva

Decentralized Adaptive Fuzzy-Neural Control of an Anaerobic Digestion Bioprocess Plant 460
Ieroham Baruch and Rosalba Galvan-Guerra

Enforcing Local Properties in Online Learning First Order TS-fuzzy Systems by Incremental Regularization.....466
Nils Rosemann, Werner Brockmann and Bastian Neumann

Dynamic vs. Static Decision Strategies in Adversarial Reasoning 472
David Pelta and Ronald R. Yager

Generalized Atanassov's Intuitionistic Fuzzy Index. Construction Method.....478
Edurne Barrenechea, Humberto Bustince, Miguel Pagola, Javier Fernandez and Josean Sanz

Time Series Analysis and Prediction Based on Fuzzy Rules and the Fuzzy Transform 483
Martin Stepnicka, Viktor Pavliska, Vilem Novak, Irina Perfilieva, Lenka Vavrickova and Iva Tomanova

On the Use of Aggregation Operators for Location Privacy 489
Aida Valls, Jordi Nin and Vicenç Torra

Application of Fuzzy Vectors of Normalized Weights in Decision Making Models.....495
Ondrej Pavlacka and Jana Talasova

Integrating Complex User Preferences into a Route Planner: A Fuzzy-Set-Based Approach 501
Amine Mokhtari, Olivier Pivert and Allel Hadjali

Basic Properties of L-fuzzy Quantifiers of the Type <1> Determined by Fuzzy Measures.....507
Antonin Dvorak and Michal Holcapek

On the Definition of Extended Norms and Co-norms to Aggregate Fuzzy Bipolar Conditions 513
Ludovic Liétard and Daniel Rocacher

An Abstract Approach to Fuzzy Logics: implicational semilinear logics.....519
Petr Cintula and Carles Noguera

Polynomial Functions on Bounded Chains 525
Miguel Couceiro and Jean-Luc Marichal

Quasi-Polynomial Functions on Bounded Chains.....531
Miguel Couceiro and Jean-Luc Marichal

Novel Methods for the Design of General Type-2 Fuzzy Sets based on Device Characteristics and Linguistic Labels Surveys.....537
Christian Wagner and Hani Hagras

Flow Line Systems with Possibilistic Data: a System with Waiting Time in Line Uncertain 544
David de la Fuente and María José Pardo

Valuations on the Algebra of Intervals 550
Mirko Navara

Impute Missing Assessments by Opinion Clustering in Multi-Criteria Group Decision Making Problems.....555
Jun Ma, Guangquan Zhang, Jie Lu and Da Ruan

Application of Zadeh's Impossibility Principle to Approximate Explanation.....561
Vesa Niskanen

On the Completion Mechanism Produced by the Choquet Integral on Some Decision Strategies 567
Christophe Labreuche

Fuzzy Implications and the Weak Law of Importation 573
Sebastià Massanet and Joan Torrens

Fuzzy Transform and Smooth Functions 579
Luciano Stefanini

Towards F-transform of a Higher Degree 585
Irina Perfilieva and Martina Daňková

Some Chaotic and Mixing Properties of Zadeh's Extension 589
Jiri Kupka

Associativity of Triangular Norms in Light of Web Geometry 595
Milan Petrík and Peter Sarkoci

Fuzzy Arithmetic with Parametric LR Fuzzy Numbers 600
Luciano Stefanini and Laerte Sorini

Type-2 Fuzzy Arithmetic using Alpha-planes 606
Hussam Hamrawi and Simon Coupland

Operations for Real-Valued Bags and Bag Relations 612
Sadaaki Miyamoto

An Explicit Mapping for Kernel Data Analysis and Application to Text Analysis 618
Sadaaki Miyamoto, Yuichi Kawasaki and Keisuke Sawazaki

Distorted Probabilities and m-Separable Fuzzy Measures 624
Yasuo Narukawa and Vicenc Torra

Image Analysis Applications of Morphological Operators based on Uninorms 630
Manuel González Hidalgo, Arnau Mir Torres, Daniel Ruiz Aguilera and Joan Torrens-Sastre

A Value via Posets Induced by Graph-Restricted Communication Situations 636
Katsushige Fujimoto and Aoi Honda

Ranking of Fuzzy Numbers, Some Recent and New Formulas 642
Saeid Abbasbandy

Lattice Properties of Discrete Fuzzy Numbers under Extended Min and Max 647
Jaume Casasnovas and J. Vicente Riera

On Some Classes of Aggregation Functions that are Migrative 653
Janos Fodor and Imre J. Rudas

Intelligent Operation Support Method based on Time Change Fuzzy Sets 657
Seiji Yasunobu and Mitsuru Kokuhuda

Semantical Evaluators 663
Antoon Bronselaer and Guy De Tre

A Hybrid Meta-Heuristic to Solve the Portfolio Selection Problem 669
Jose M. Cadenas, Juan V. Carrillo, M. Carmen Garrido, Carlos Ivorra, Teresa Lamata and Vicente Liern

Information System for the valuation of Universities in Spain 675
M. Socorro Garcia-Cascales and M. Teresa Lamata

Multiple-Criteria Fuzzy Evaluation: The FuzzME Software Package 681
Jana Talasova and Pavel Holecek

A Computing with Words Path from Fuzzy Logic to Natural Language 687
Claudio Moraga and Enric Trillas

Approximation Properties of Inverse Fuzzy Transforms over the Residuated Lattice 693
Martina Dankova and Martin Stepnicka

An Interval Type-2 Fuzzy Distribution Network 697
Simon Miller, Viara Popova, Robert John and Mario Gongora

Fuzzy Differential Equation with π-Derivative 703
Yurilev Chalco-Cano, Heriberto Román-Flores and María Dolores Jiménez-Gamero

Fuzzy Logic Approach to Represent and Propagate Imprecision in Agri-Environmental Indicator Assessment 707
Zainab Assaghir, Philippe Girardin and Amedeo Napoli

Representation of States on MV-algebras by Probabilities on R-generated Boolean Algebras 713
Brunella Gerla and Tomas Kroupa

Contributions to Dynamic Multicriteria Decision Making Models 719
Tiago Pais and Rita Ribeiro

A Fuzzy Set Approach to Ecological Knowledge Discovery 725
Arkadiusz Salski

Numerical Solution of Interval Differential Equations with Generalized Hukuhara Differentiability 730
Barnabas Bede and Luciano Stefanini

Phi-Calculus Fuzzy Arithmetic: application to model based control 736
Laurent Vermeiren, Thierry M. Guerra and Hakim Lamara

Optimising the Fuzzy Granulation of Attribute Domains 742
Marcos Cintra, Heloisa Camargo and Trevor Martin

Discriminatory Components for Pattern Classification 748
Nick Pizzi and Witold Pedrycz

Image Reduction with Interval-Valued Fuzzy Sets and OWA Operators 754
Daniel Paternain, Humberto Bustince, Josean Sanz, Mikel Galar and Carlos Guerra

Fuzzy Transforms, Korovkin Theorems and the Durrmeyer Operator 760
Barnabas Bede and Imre J. Rudas

Fuzzy Logic Applications in Wireless Communications 763
Maria Erman, Abbas Mohammed and Elisabeth Rakus-Andersson

Learning Fuzzy Rule Based Classifier in High Performance Computing Environment 768
Vinicius da F. Vieira, Alexandre G. Evsukoff, Beatriz S. L. P. de Lima and Sylvie Galichet

The Use of Interval Type-2 Fuzzy Logic as a General Method for Edge Detection 774
Olivia Mendoza, Patricia Melin and Juan Ramon Castro

Comparison of Two Versions of the Ferrers Property of Fuzzy Interval Orders 780
Susana Díaz, Bernard De Baets and Susana Montes

Decomposition of the Transitivity for Additive Fuzzy Preference Structures 786
Susana Díaz, Susana Montes and Bernard De Baets

Fuzzy Tendency based Time Series Model for Forecasting Server Traffic 792
Nadezhda Yarushkina, Tagir Unusov and Tatyana Afanasyeva

Continuity and Additivity of the Trapezoidal Approximation Preserving the Expected Interval Operator 798
Adrian Ban and Lucian Coroianu

A Baseline Learning Genetic Fuzzy Classifier Based on Low Quality Data 803
Ana Palacios, Luciano Sanchez and Ines Couso

Pattern Recognition in Blur Motion Noisy Images using Fuzzy Methods for Response Integration in Ensemble Neural Networks 809
Miguel Lopez, Patricia Melin and Oscar Castillo

Fuzzy Logic Adaptation of a Hybrid Evolutionary Method for Pattern Recognition 815
Fevrier Valdez, Patricia Melin and Oscar Castillo

Fuzzy Subgroups and the Teichmüller Space 821
Paavo Kukkurainen

Reasoning about Actions in Fuzzy Environment 825
Emad Saad, Shamiaa Elmorsy, Mahmoud Gabr and Yasser Hassan

Interval-Valued Restricted Equivalence Functions Applied on Clustering Techniques 831
Aranzazu Jurio, Miguel Pagola, Daniel Paternain, Carlos Lopez-Molina and Pedro Melo-Pinto

Transitive Closure of Interval-valued Relations 837
Ramón González del Campo, Luis Garmendia and Jordi Recasens

Choquet Integral Based Evaluations by Fuzzy Rules....843
Eiichiro Takahagi

Fuzzy Modelling of Sensor Data for the Estimation of an Origin-Destination Matrix 849
Krystyna Biletska, Sophie Midenet, Marie-Hélène Masson and Thierry Denoeux

Classification with Reject Options in a Logical Framework: a fuzzy residual implication approach....855
Hoel Le Capitaine and Carl Frélicot

Fuzzy c-Lines for Data with Tolerance....861
Yuchi Kanzawa, Yasunori Endo and Sadaaki Miyamoto

Geographic Information Retrieval based on Two Orthogonal Criteria....867
Gloria Bordogna, Giorgio Ghisalberti, Marco Pagani and Giuseppe Psaila

A Semi-infinite Programming Approach to Possibilistic Optimization under Necessity Measure Constraints....873
Masahiro Inuiguchi

On Aristotle's NC and EM Principles in Three-valued Logics 879
Enric Trillas, Claudio Moraga and Eloy Renedo

Parametric Fuzzy Modelling Framework for Complex Data-Inherent Structures 885
Arne-Jens Hempel and Steffen F. Bocklisch

On Triangular Norm Based Fuzzy Description Logics....891
Àngel García-Cerdaña, Eva Armengol and Francesc Esteva

Evaluating Premises, Partial Consequences and Partial Hypotheses 897
Itziar García-Honrado, Enric Trillas, Sergio Guadarrama and Eloy Renedo

Flexible Retrieval of X-Ray Images Based on Shape Descriptors Using a Fuzzy Object-Relational Database 903
Juan Miguel Medina, Sergio Jaime-Castillo, Carlos D. Barranco and Jesús R. Campaña

Minimal Combination for Incremental Grammar Fragment Learning....909
Nurfadhlina Mohd Sharef, Trevor Martin and Yun Shen

Managing Imperfect Temporal Metadata in the Catalog services of Spatial Data Infrastructures Compliant with INSPIRE 915
Gloria Bordogna, Paola Carrara, Marco Pagani, Monica Pepe and Anna Rampini

On Some Fuzzy Categories of Many-Valued Topological Spaces 921
Ingrida Uljane

Valued Constraint Satisfaction Problems Applied to Functional Harmony....925
Nele Verbiest, Chris Cornelis and Yvan Saeys

Detecting and Reacting on Drifts and Shifts in On-Line Data Streams with Evolving Fuzzy Systems....931
Edwin Lughofer and Plamen Angelov

How to Learn Fuzzy User Preferences with Variable Objectives....938
Alan Eckhardt and Peter Vojtas

Multi-Dimensional Scaling applied to Hierarchical Rule Systems....944
Thomas Gabriel, Kilian Thiel and Michael Berthold

Rule Weight Optimization and Feature Selection in Fuzzy Systems with Sparsity Contraints....950
Edwin Lughofer and Stefan Kindermann

Increasing the Performance of Fuzzy Retrieval Using Impact Ordering....957
Carlos Barranco and Sven Helmer

Colour Image Segmentation using A-IFSs....963
Pedro Couto, Pedro Melo-Pinto, Humberto Bustince, Edurne Barrenechea and Miguel Pagola

SaM: A Split and Merge Algorithm for Fuzzy Frequent Item Set Mining....968
Christian Borgelt and Xiaomeng Wang

Online Recognition of Fuzzy Time Series Patterns....974
Gernot Herbst and Steffen F. Bocklisch

The Collapsing Method: Does the Direction of Collapse Affect Accuracy?....980
Sarah Greenfield, Francisco Chiclana and Robert John

Data Integration, Approximate Categorisation and Fuzzy Associations....986
Trevor Martin and Yun Shen

Around Fuzziness. Some Philosophical Thoughts....992
Alejandro Sobrino and Martín Pereira

Fuzziness as a Model of User Preference in Semantic Web Search....998
Veronika Vanekova and Peter Vojtas

Multi-Goal Aggregation of Reduced Preference Relations Based on Fuzzy Interactions between Decision Goals....1004
Rudolf Felix

Efficient Calculation of the Value Function in Fuzzy Real Option by Differential Evolution Algorithms....1009
Elettra Agliardi, Maria Letizia Guerra and Luciano Stefanini

A Dynamic Classification Method for the Discrimination of Evolving Data....1015
Laurent Hartert, Moamar Sayed Mouchaweh and Patrice Billaudel

Possibility Theory and Formal Concept Analysis in Information Systems....1021
Didier Dubois and Henri Prade

On the use of Measures of Separability of Classes to Characterise the Domains of Competence of a Fuzzy Rule Based Classification System....1027
Julián Luengo and Francisco Herrera

Learning Concurrently Granularity, Membership Function Parameters and Rules of Mamdani Fuzzy Rule-based Systems....1033
Michela Antonelli, Pietro Ducange, Beatrice Lazzerini and Francesco Marcelloni

Semantically-Driven Flexible Division in Fuzzy Object Oriented Models....1039
Nicolás Marin, Carlos Molina, Olga Pons and M. Amparo Vila

Rule Base and Inference System Cooperative Learning of Mamdani Fuzzy Systems with Multiobjective Genetic Algorithms....1045
Antonio Angel Márquez, Francisco Alfredo Márquez and Antonio Peregrín

On the Information Provided by Uncertainty Measures in the Classification of Remote Sensing Images 1051
Luisa M. S. Gonçalves, Cidália C. Fonte, Eduardo N. B. S. Júlio and Mario Caetano

On Interval and Fuzzy Calculations of Economic Uncertainty 1057
Hans Schjær-Jacobsen

A Method for Locating the Iliac Crests Based on the Fuzzy Hough Transform 1063
Sergio Jaime-Castillo, Juan Miguel Medina and Antonio Garrido

Membership Dependent Stability Analysis of TS Fuzzy Controlled Systems using Coupling Attenuation 1068
Jean-Yves Dieulot and Nesrine Elfelly

Characterization of L-Fuzzy Semi-Filters and Semi-Ideals 1073
Jorge Jiménez, Susana Montes, Branimir Šešelja and Andreja Tepavčević

Fuzzy Intersection and Difference Model for Topological Relations 1079
Ahed Alboody, Florence Sedes and Jordi Inglada

T-norm-based Fuzzy Logics and Logics for Human Reasoning 1085
Thomas Vetterlein and Klaus-Peter Adlassnig

On the Representation Theorem of Multi-Adjoint Concept Lattices 1091
Jesús Medina and Manuel Ojeda-Aciego

A Neuro-fuzzy System for Fraud Detection in Electricity Distribution 1096
Cyro Muniz, Marley Vellasco, Ricardo Tanscheit and Karla Figueiredo

A Fuzzy System for the Assessment of Human Reliability 1102
Dilza Szwarcman, Jesus Domech, Ricardo Tanscheit, Marley Vellasco, Marco Pacheco and Nicholas Ribeiro

Multivariate Segmentation of Time Series with Differential Evolution 1108
Daniel Graves and Witold Pedrycz

Coevolutionary Genetic Fuzzy System to Assess Multiagent Bidding Strategies in Electricity Markets 1114
Igor Walter and Fernando Gomide

A Type-2 Fuzzy Portfolio Selection Problem Considering Possibility Measure and Crisp Possibilistic Mean Value 1120
Takashi Hasuike and Hiroaki Ishii

Finding the Best Flexibility Strategies by Using an Integrated Method of FAHP and QFD 1126
Müjde Erol Genevois and Tuncay Gürbüz

Continuous OWA Operator and its Calculation 1132
Yasuo Narukawa and Vicenc Torra

A process Algebra Approach to Fuzzy Reasoning 1136
Liliana D'Errico and Michele Loreti

A Generalized Numerical Solution for Fuzzy Relation Equations 1142
Iman Mohammadi Ardehali and Milad Avazbeigi

A Multiobjective Evolutionary Algorithm for Tuning Fuzzy Rule Based Systems with Measures for Preserving Interpretability 1146
María José Gacto, Rafael Alcalá and Francisco Herrera

On Tolerant Fuzzy c-Means Clustering with L1-Regularization 1152
Yukihiro Hamasuna, Yasunori Endo and Sadaaki Miyamoto

A Parametric Approach to Solve Quadratic Programming Problems with Fuzzy Environment in the Set of Constraints .1158
Carlos Cruz, Ricardo C. Silva, José Luis Verdegay and Akebo Yamakami

Interval-Valued Fuzzy System for Segmentation of Prostate Ultrasound Images 1164
Miguel Pagola, Edurne Barrenechea, Aranzazu Jurio, Mikel Galar, Pedro Couto, Farhang Sahba and Hamid Tizhoosh

A toward Framework for Generic Uncertainty Management....1169
Ernesto Damiani, Paolo Ceravolo and Marcello Leida

Fuzzy Descriptions to Identify Temporal Substructure Changes of Cooccurrence Graphs 1177
Matthias Steinbrecher and Rudolf Kruse

Online Measurement of Oily-water Using Rule-based Fuzzy Concentration Map and Halogen Fluorescence....1183
Gyeongdong Baek, Moon-Ho Choi, Tae-Ryong Jeon, Jaewoo Cho and Sungshin Kim

Features Stereo Matching Based on Fuzzy Logic 1188
Grzegorz Nieradka and Bohdan Butkiewicz

Non-additive Robust Ordinal Regression with Choquet integral, Bipolar and Level Dependent Choquet integrals 1194
Silvia Angilella, Salvatore Greco and Benedetto Matarazzo

Towards a Unified Logical Framework of Fuzzy Implications to Compare Fuzzy Sets....1200
Hoel Le Capitaine and Carl Frélicot

Preservation of Properties of Interval-Valued Fuzzy Relations 1206
Barbara Pękala

A Fuzzy Approach for the Model of Sliding Window. An Application to Behaviour Patterns Mining 1211
María Ros, Miguel Delgado and Amparo Vila

Organizational Risk Assessment using Adaptive Neuro-Fuzzy Inference System 1217
Javad Jassbi and Sohrab Khanmohammadi

On the Use of Min-Based Revision Under Uncertain Evidence for Possibilistic Classifiers....1223
Salem Benferhat and Karim Tabia

Parallel Coordinates-based Fuzzy Color Distance and its Applications 1229
Constantin Vertan

Definition of Fuzzy Pareto-Optimality by Using Possibility Theory....1234
Ricardo C. Silva and Akebo Yamakami

Terrain Morphology Classification over Fuzzy Digital Elevation Models 1240
Jorge Santos

Modeling of Polyester Dyeing Using an Evolutionary Fuzzy System....1246
Maryam Nasiri and Stefan Berlik

Fuzzy Concept Lattice is Made by Proto-Fuzzy Concepts....1252
Ondrej Kridlo and Stanislav Krajci

A Preliminary Analysis for Improving Model Structure of Fuzzy Habitat Preference Model for Japanese Medaka (Oryzias latipes) 1258
Shinji Fukuda

Towards Intuitive Understanding of the Cauchy Deviate Method for Processing Interval and Fuzzy Uncertainty....1264
Vladik Kreinovich and Hung Nguyen

Fuzzy and PSO Based Algorithm for Driver's Behavior Modeling 1270
Sohrab Khanmohammadi, Mohammad Ali Tinati, Sehraneh Ghaemi and Mohammad Ali Badamchizadeh

Evaluating Firms' Gender Equity by Fuzzy Logic....1276
Tindara Addabbo, Gisella Facchinetti, Giovanni Mastroleo and Tiziana Lang

Morphisms in Categories of Sets with Similarity Relations 1282
Jiri Mockor

Bipolar Queries: An Approach and its Various Interpretations 1288
Slawomir Zadrozny and Janusz Kacprzyk

A Fuzzy Variant of the Rand Index for Comparing Clustering Structures 1294
Eyke Huellermeier and Maria Rifqi

Fuzzy Modeling of Labeled Point Cloud Superposition for the Comparison of Protein Binding Sites 1299
Thomas Fober and Eyke Huellermeier

An Interval Approach for Fuzzy Linear Regression with Imprecise Data 1305
Amory Bisserier, Reda Boukezzoula and Sylvie Galichet

A Predator-Prey Model with Fuzzy Initial Populations 1311
Muhammad Zaini Ahmad and Bernard De Baets

Local Semantic Context Analysis for Automatic Ontology Matching 1315
Giuseppe Fenza, Vincenzo Loia and Sabrina Senatore

Using Fuzzy Linguistic Summaries for the Comparison of Time Series: an application to the analysis of investment fund quotations 1321
Janusz Kacprzyk and Anna Wilbik

Exchanging Iterated Expectations of Random Upper Semicontinuous Functions: an application to decision theory 1327
Miguel López-Díaz and Luis J. Rodríguez-Muñiz

60 years "A Mathematical Theory of Communication" — Towards a "Fuzzy Information Theory" 1332
Rudolf Seising

On Classical, Fuzzy Classical, Quantum and Fuzzy Quantum Systems 1338
Rudolf Seising

Possibility and Necessity Evaluations based on Ordinal Comparability 1343
Masaaki Ida

Fuzzy Answer Set Programming with Literal Preferences 1347
Jeroen Janssen, Steven Schockaert, Dirk Vermeir and Martine De Cock

Two Results About Optimization of Fuzzy Variable Functions 1353
A. David Baez-Sanchez

Industrial Monitoring by Evolving Fuzzy Systems 1358
Lara Aires, Joao Araujo and Antonio Dourado

Probabilistic Decision Making with the OWA Operator and its Application in Investment Management 1364
José M. Merigó

Some Results on Lipschitz Quasi-Arithmetic Means 1370
Gleb Beliakov, Tomasa Calvo and Simon James

Modeling Position Specificity in Sequence Kernels by Fuzzy Equivalence Relations 1376
Ulrich Bodenhofer, Karin Schwarzbauer, Mihaela Ionescu and Sepp Hochreiter

Universal Approximation of a Class of Interval Type-2 Fuzzy Neural Networks Illustrated with the Case of Non-linear Identification 1382
Juan R. Castro, Oscar Castillo, Patricia Melin, Antonio Rodriguez-Diaz and Olivia Mendoza

Spectral Learning with Type-2 Fuzzy Numbers for Question/Answering System 1388
Asli Celikyilmaz and I. Burhan Turksen

Can We Learn Algorithms from People Who Compute Fast: an indirect analysis in the presence of fuzzy descriptions ..1394
Olga Kosheleva

A Parametric Method to Solve Quadratic Programming Problems with Fuzzy Costs ..1398
Ricardo C. Silva, Carlos Cruz and Akebo Yamakami

Comparison of Multicriteria Methods for Land-use Suitability Assessment ..1404
Jozo Dujmovic, Guy De Tre and Suzana Dragicevic

Computing with Actions: the case of driving a car in a simulated car race ..1410
Sergio Guadarrama

Analysis of Similarity Measures for Atanassov's Intuitionistic Fuzzy Sets ..1416
Eulalia Szmidt and Janusz Kacprzyk

Non-Commutative EQ-Logics and Their Extensions..1422
Vilem Novak and Martin Dyba

Discriminant Analysis for Fuzzy Random Variables Based on Nonparametric Regression..1428
Ana Colubi, M. Ángeles Gil, Gil González-Rodríguez and Wolfgang Trutschnig

Confidence Regions for the Mean of a Fuzzy Random Variable..1433
Gil González-Rodríguez, Wolfgang Trutschnig and Ana Colubi

A Fuzzy Multi-Criteria Approach to WEEE Treatment Strategy Selection..1439
İlke Bereketli, Tuncay Gürbüz and Müjde Erol Genevois

Hierarchic Approach in the Analysis of Tomographic Eye Image ..1444
Robert Koprowski and Zygmunt Wrobel

A Two Step Solution Procedure to a Fuzzy Medical Waste Disposal Facility Location Problem..1450
H. Ziya Ulukan and Yeşim Kop

Online Classification of Machine Operation Modes Based on Information Compression and Fuzzy Similarity Analysis..1456
Gancho Vachkov

Introducing Relation Compactness for Generating a Flexible Size of Search Results in Fuzzy Queries..1462
Xiaohui Tang, Guoqing Chen and Qiang Wei

Robot Cooperation without Explicit Communication by Fuzzy Signatures and Decision Trees..1468
Áron Ballagi, Laszlo T. Koczy and Tom Gedeon

A Fuzzy Filter for the Removal of Gaussian Noise in Colour Image Sequences..1474
Tom Mélange, Mike Nachtegael and Etienne Kerre

Visualising Rough Time Intervals in a Two-Dimensional Space ..1480
Yi Qiang, Katrin Asmussen, Matthias Delafontaine, Birger Stichelbaut, Guy De Tre, Philippe De Maeyer and Nico Van de Weghe

Repairing Infeasibility in Fuzzy Goal Programming ..1486
Mariano Jiménez, Amelia Bilbao, Mar Arenas and Maria Victoria Rodriguez-Uria

Sorted Kernel Matrices as Cluster Validity Indexes ..1490
Francisco Queiroz, Antonio Braga and Witold Pedrycz

Complementary Source Information Cooperation Within a Decision System for Crop Monitoring ..1496
Mahmoud El Hajj, Serge Guillaume and Agnes Begue

Soft Computing Confronting Philosophical and Sociological Critiques to Classical AI ..1502
Veronica Sanz

Optimizing Fuzzy Flip-Flop Based Neural Networks by Bacterial Memetic Algorithm....1508
Rita Lovassy, Laszlo T. Koczy and Laszlo Gal

Bi-criteria Genetic Selection of Bagging Fuzzy Rule-based Multiclassification Systems....1514
Krzysztof Trawinski, Arnaud Quirin and Oscar Cordon

A New Fuzzy Noise-Rejection Data Partitioning Algorithm with Revised Mahalanobis Distance....1520
Mohammad Hossein Fazel Zarandi, Milad Avazbeigi and I. Burhan Turksen

Determining OWA Weights by Maximizing Consensus....1527
José Luis García-Lapresta, Bonifacio Llamazares and Teresa Peña

Weighted Trapezoidal Approximations of Fuzzy Numbers....1531
Przemyslaw Grzegorzewski and Karolina Pasternak-Winiarska

Nonlinear-shaped Approximation of Fuzzy Numbers....1535
Przemyslaw Grzegorzewski and Luciano Stefanini

Characterizing Fuzzy Modal Semantics by Fuzzy Multimodal Systems with Crisp Accessibility Relations....1541
Félix Bou, Francesc Esteva, Lluis Godo and Ricardo Oscar Rodriguez

Fuzzy Optimization in Vehicle Routing Problems....1547
Julio Brito, José Andrés Moreno and José Luis Verdegay

Weighted Decisions in a Fuzzy Random Forest....1553
Piero Bonissone, José M. Cadenas, Carmen Garrido, Ramón Andrés Díaz and Raquel Martínez

Evolutionary Learning of Quantified Fuzzy Rules for Hierarchical Grouping of Laser Sensor Data in Intelligent Control....1559
Manuel Mucientes, Ismael Rodríguez-Fdez and Alberto Bugarín

Graded Equipollence of Fuzzy Sets....1565
Michal Holcapek and Matej Turcan

Managing the Absence of Items in Fuzzy Association Mining....1571
Carlos Molina, Daniel Sanchez, Jose-Maria Serrano and M. Amparo Vila

Towards an Interpretation of the Connectivity Analysis in the Frame of Fuzzy Equivalence Relations for Medical Diagnoses....1577
Tatiana Kiseliova

Identification of Web Information using Concept Hierarchies and On-line Updates of Concept Importance....1583
Zhan Li, Ronald R. Yager and Marek Reformat

Possibility Distribution: a unified representation for parameter estimation....1589
Gilles Mauris

Cooperative Partner Agent of Seven-Card-Stud Poker....1595
Keisuke Osone and Takehisa Onisawa

Systems of Fuzzy Relation Equations in a Space with Fuzzy Preorder....1601
Irina Perfilieva

On M-Approximative Operators and M-Approximative Systems....1606
Alexander Sostak

A Fuzzy Weight Representation for Inner Dependence Method AHP....1612
Shin-ichi Ohnishi, Takahiro Yamanoi and Hideyuki Imai

On Conceptual Indexing for Data Summarization....1618
Henrik Bulskov and Troels Andreasen

Fuzzy Analogical Model of Adaptation for Case-Based Reasoning 1625
Bernadette Bouchon-Meunier, Christophe Marsala and Maria Rifqi

Embedded Genetic Learning of Highly Interpretable Fuzzy Partitions 1631
Jorge Casillas

Non Additive Measures for Group Multi Attribute Decision Models 1637
Marta Cardin and Silvio Giove

A Desktop Calculator for Parametric Fuzzy Arithmetic 1643
Laerte Sorini

Discussions on Interpretability of Fuzzy Systems using Simple Examples 1649
Hisao Ishibuchi and Yusuke Nojima

Stability Analysis of Continuous PI-like Fuzzy Control Systems based on Vector Norms Approach 1655
Anis Sakly, Basma Zahra and Mohamed Benrejeb

The fuzzy Induced Generalized OWA Operator and its Application in Business Decision Making 1661
Jose M. Merigo and Anna M. Gil-Lafuente

Solution for Fuzzy Road Transport Traveling Salesman Problem Using Eugenic Bacterial Memetic Algorithm 1667
János Botzheim, Péter Földesi and László Kóczy

Feature Selection Using Fuzzy Objective Functions 1673
Susana Vieira, João M.C. Sousa and Uzay Kaymak

Clustering of Fuzzy Shapes by Integrating Procrustean Metrics and Full Mean Shape Estimation into K-Means Algorithm 1679
Vasile Georgescu

Development of a Fuzzy Expert System for a Nutritional Guidance Application 1685
Petri Heinonen, Marjo Mannelin, Hannu Iskala, Aki Sorsa and Esko Juuso

Possible and Necessary h-Indices 1691
Marek Gągolewski and Przemysław Grzegorzewski

Similarity Based Fuzzy Interpolation Applied to CAGD 1696
Joan Jacas, Amadeu Monreal and Jordi Recasens

KM-Fuzzy Approach Space 1702
Juan Martínez-Moreno, Antonio Roldán and Concepcion Roldán

Decision Making in Competitive Location using Fuzzy Sets 1706
Clara M. Campos, Lourdes Canós, María J. Canós, Vicente Liern, José A. Moreno and Dolores Santos

Using Syntactic Possibilistic Fusion for Modeling Optimal Optimistic Qualitative Decision 1712
Salem Benferhat, Faiza Haned, Aïcha Mokhtari and Ismahane Zeddigha

A Fuzzy-Semiotic Framework for Modeling Imprecision in the Assessment of Depression 1717
Mila Kwiatkowska, Krzysztof Kielan and Krzysztof Michalik

An Axiomatical Approach to T-Multi-indistinguishabilities 1723
Javier Martín and Gaspar Mayor

Mining Linguistic Information for Configuring a Visual Surface Inspection System 1729
David Martin, Domingo Miguel Guinea, Maria C. Garcia-Alegre and Domingo Guinea

A construction of an L-fuzzy valued measure of L-fuzzy sets 1735
Vechislav Ruzha and Svetlana Asmuss

Distributed Genetic Tuning of Fuzzy Rule-Based Systems 1740
Ignacio Robles, Rafael Alcalá, José Manuel Benítez and Francisco Herrera

Forecasting Exchange Rates: A Neuro-Fuzzy Approach 1745
Meysam Alizadeh, Roy Rada, Akram Khaleghei Ghoshe Balagh and Mir Mehdi Seyyed Esfahani

Genetic Learning of Serial Hierarchical Fuzzy Systems for Large-Scale Problems 1751
Alicia D. Benítez and Jorge Casillas

An Introduction to Parameterized IFAM Models with Applications in Prediction 1757
Peter Sussner, Rodolfo Miyasaki and Marcos Eduardo Valle

Defining A Fuzzy Partition for Coarseness Modelling in Texture Images 1763
Jesus Chamorro-Martinez, Pedro M. Martínez-Jimenez and Jose M. Soto-Hidalgo

A Fuzzy Particle Swarm Optimization Algorithm for a Cell Formation Problem 1768
Esmaeil Mehdizadeh and Reza Tavakkoli-Moghaddam

Multiple Negations in Fuzzy Interval Logic 1773
Eunjin Kim and Ladislav J. Kohout

Development of Multiple Linguistic Equation Models with Takagi-Sugeno Type Fuzzy Models 1779
Esko Juuso

Knowledge Discovery in the Prediction of Bankruptcy 1785
Rui Jorge Almeida, Susana Vieira, Viorel Milea, Uzay Kaymak and Joao Miguel Costa Sousa

Generalizations of Fuzzy C-Means Algorithm to Granular Feature Spaces, based on Underlying Metrics: Issues and Related Works 1791
Vasile Georgescu

Enhancing OLAP Querying with the aid of H-IFS 1797
Panagiotis Chountas, Ermir Rogova and Krassimir T. Atanassov

Modeling Investor Optimism with Fuzzy Connectives 1803
Milan Lovric, Rui Jorge Almeida, Uzay Kaymak and Jaap Spronk

Tail Point Density Estimation Using Probabilistic Fuzzy Systems 1809
Rui Jorge Almeida and Uzay Kaymak

Optimization of an Oil Production System using Neural Networks and Genetic Algorithms 1815
Guillermo Jimenez de la C., Jose A. Ruz-Hernandez, Evgen Shelomov and Ruben Salazar M.

Rule Based Fuzzy Cognitive Maps in Socio-Economic Systems 1821
Joao Paulo Carvalho and Jose Tome

Operational Semantics for a Fuzzy Logic Programming System with Defaults and Constructive Answers 1827
Hannes Strass, Susana Muñoz-Hernández and Victor Pablos Ceruelo

Applying Bacterial Memetic Algorithm for Training Feedforward and Fuzzy Flip-Flop based Neural Networks 1833
László Gál, János Botzheim, Laszlo T. Koczy and Antonio Ruano

Interactive Fuzzy Modeling by Evolutionary Multiobjective Optimization with User Preference 1839
Yusuke Nojima and Hisao Ishibuchi

Incremental Possibilistic Approach for Online Clustering and Classification 1845
Moamar Sayed Mouchaweh and Bernard Riera

Optimistic Fuzzy Weighted Average 1851
Sylvie Galichet and Reda Boukezzoula

Evolving Inverse Fuzzy Models for Uncalibrated Visual Servoing in 3D Workspace 1857
Paulo Goncalves, João Sousa and João Caldas Pinto

On the Reliability of the Color Gradient Vector Argument Approach 1863
Belén Prados-Suarez, Pilar Sobrevilla, Eduard Montseny and Santiago Romaní

Evolutionary Robot Vision and Fuzzy Evaluation for Natural Communication of Partner Robots 1869
Akihiro Yorita and Naoyuki Kubota

Combining Wavelets and Computational Intelligence Methods with Applications on Multi-class Classification Datasets 1875
Carlos Campos Bracho

Poincaré recurrence theorem in MV-algebras 1880
Beloslav Riečan

Notes on Q-probabilities on intuitionistic fuzzy events 1882
Magdaléna Renčová

A New Bridge Between Fuzzy Sets and Statistics .....
Pedro Teran

L-Prime Spectrum of Modules .....
Bayram Ali Ersoy and Dan Ralescu

An Approach to Interval-Valued R-Implications and Automorphisms

Benjamín C. Bedregal[1] Graçaliz P. Dimuro, Renata H. S. Reiser[2]

1.Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte
Campus Universitário, 59072-970 Natal, Brazil
2.Programa de Pós-Graduação em Informática, Universidade Católica de Pelotas
Rua Felix da Cunha, 96010-000 Pelotas, Brazil
Email: bedregal@dimap.ufrn.br, {liz,reiser}@ucpel.tche.br

Abstract— The aim of this work is to introduce an approach for interval-valued R-implications, which satisfy some analogous properties of R-implications. We show that the best interval representation of an R-implication that is obtained from a left continuous t-norm coincides with the interval-valued R-implication obtained from the best interval representation of such t-norm, whenever this is an inclusion monotonic interval function. This provides, under this condition, a nice characterization for the best interval representation of an R-implication, which is also an interval-valued R-implication. We also introduce interval-valued automorphisms as the best interval representations of automorphisms. It is shown that interval automorphisms act on interval R-implications, generating other interval R-implications.

Keywords— Interval-valued fuzzy logic, R-implications, automorphisms, interval representations.

1 Introduction

Interval analysis [1] has been playing an important role in the modeling of the uncertainty and the errors that occur in numerical computations, allowing the development of computational models, methods and tools for the automatic error analysis of numerical algorithms in digital computers, with application in technological and scientific computation.[1]

On the other hand, fuzzy set theory [3] has provided a more generical and complete mathematical model of uncertainty and vagueness, which has been considered as the oldest and most widely reported component of present-day soft computing, helping with the design of flexible information processing systems, with application in, e.g., control systems, decision making, expert systems, pattern recognition. [4]

However, there may be the case that we do not have precise knowledge about the membership function that should be taken in a certain application. These considerations have led to some extensions of fuzzy sets, such us type-2 fuzzy sets [5], which incorporate uncertainty about the membership function into fuzzy set theory. Type-2 fuzzy sets has been largely applied since the works of Jerry Mendel in the 90's [6].

Interval-valued fuzzy sets are a particular case of type-2 fuzzy sets, where the membership degree of each element of the fuzzy set is given by a closed subinterval of the unit interval $[0, 1]$, allowing to deal not only with vagueness (lack of sharp class boundaries), but also with the uncertainty (lack of information), intuitively [7, 8]. Interval-valued fuzzy sets were introduced independently by Zadeh [5] and other authors (e.g., [9, 10, 11, 12]) in the 70's. Among several papers connecting these areas (see, e.g., [13, 14, 15, 16]), we adopted Bedregal and Takahashi's work [17, 18].

Fuzzy implications [19, 20] play an important role in fuzzy logic. They are usually derived from t-norms and t-conorms in several ways, e.g, S-implications, R-implications, QL-implications and D-implications.[2] The importance of fuzzy implications is not only because they are used in representing "If ... then" rules in fuzzy systems, but also because their use in performing inferences in approximate reasoning and fuzzy control. This is the main reason for searching many different models to perform this kind of fuzzy connectives. In particular, R-implications [21], which are based on left continuous t-norms, are used in the modelling of fuzzy rules that satisfy natural properties of implications (see Sect. 4), leading to the residuation property and to an important family of fuzzy logics [22]. On the other hand, the use of automorphisms [19] allows to modify implications preserving their fundamental properties, and then preserving their axiomatic.

The aim of this work is to introduce an approach for interval-valued R-implications as interval generalizations for R-implications that satisfy some analogous properties of R-implications. We show that the best interval representation of an R-implication that is obtained from a left continuous t-norm coincides with the interval-valued R-implication obtained from the best interval representation of such t-norm, whenever this is an inclusion monotonic interval function. Then, it is possible to provide, under this condition, a nice characterization for the best interval representation of an R-implication, which is also an interval-valued R-implication. We also introduce interval-valued automorphisms [16] as the best interval representations of automorphisms, showing that interval automorphisms act on interval R-implications, generating other interval R-implications. Commutative diagrams are used to analyze the relationship between the process for obtaining R-implications from t-norms and the process for obtaining interval-valued R-implications from interval-valued t-norms and those canonical interval constructions. Also, we show that the use of automorphism over R-implications and the use of interval-valued automorphisms over interval-valued

[1]For a survey on applications of Interval Mathematics, see, e.g., [2] and http://www.cs.utep.edu/interval-comp/.

[2]There are also some other methods for the generation of fuzzy implications (see, e.g, [20]).

ISBN: 978-989-95079-6-8

R-implications also comute when the interval constructor is applied.

The paper is organized as follows. In Sect. 2, the main concepts related to interval representations are presented. Section 3 considers interval-valued fuzzy t-norms. Fuzzy implications and R-implications, together with their properties, are presented in Sect. 4. Interval-valued fuzzy implications and R-implications are introduced in Sect. 5, and interval-valued automorphism in Sect. 6, as well as the main related results of the paper. Section 6 is the Conclusion and final remarks.

2 Interval Representations

Consider the real unit interval $U = [0, 1] \subseteq \mathbb{R}$ and let $\mathbb{U}$ be the set of subintervals of U, that is, $\mathbb{U} = \{[a, b] \mid 0 \leq a \leq b \leq 1\}$. The interval set has two projections $l, r : \mathbb{U} \to U$, defined by $l([a, b]) = a$ and $r([a, b]) = b$, respectively. For $X \in \mathbb{U}$, $l(X)$ and $r(X)$ are also denoted by $\underline{X}$ and $\overline{X}$, respectively.

The partial orders that are considered in this paper are the inclusion relation and the component-wise *Kulisch-Miranker order* (also called *product order*), defined by:

$$\forall X, Y \in \mathbb{U} : X \leq Y \Leftrightarrow \underline{X} \leq \underline{Y} \wedge \overline{X} \leq \overline{Y}. \quad (1)$$

An interval $X \in \mathbb{U}$ is said to be an interval representation of a real number α if $\alpha \in X$. Considering two interval representations X and Y of a real number α, X is said to be a better representation of α than Y if $X \subseteq Y$. This notion can be easily extended for tuples of n intervals $\overrightarrow{X} = (X_1, \ldots, X_n)$.

Definition 1 *A function $F : \mathbb{U}^n \longrightarrow \mathbb{U}$ is an* interval representation *of a function $f : U^n \longrightarrow U$ if, for each $\overrightarrow{X} \in \mathbb{U}^n$ and $\overrightarrow{x} \in \overrightarrow{X}$, $f(\overrightarrow{x}) \in F(\overrightarrow{X})$ [23].*[3]

Definition 2 *Let $F : \mathbb{U}^n \longrightarrow \mathbb{U}$ and $G : \mathbb{U}^n \longrightarrow \mathbb{U}$ be two interval representations of the function $f : U^n \longrightarrow U$. F is a* better interval representation *of f than G, denoted by $G \sqsubseteq F$, if, for each $\overrightarrow{X} \in \mathbb{U}^n$, the inclusion $F(\overrightarrow{X}) \subseteq G(\overrightarrow{X})$ holds.*

Definition 3 *For each real function $f : U^n \longrightarrow U$, the interval function $\widehat{f} : \mathbb{U}^n \longrightarrow \mathbb{U}$, defined by*

$$\widehat{f}(\overrightarrow{X}) = \left[\inf\{f(\overrightarrow{x}) \mid \overrightarrow{x} \in \overrightarrow{X}\}, \sup\{f(\overrightarrow{x}) \mid \overrightarrow{x} \in \overrightarrow{X}\}\right], \quad (2)$$

is called the best interval representation of *f [23].*[4]

The interval function $\widehat{f}$ is well defined and for any other interval representation F of f, $F \sqsubseteq \widehat{f}$. The interval function $\widehat{f}$ returns an interval that is narrower than any other interval representation of f. Thus, $\widehat{f}$ presents the *optimality property* of interval algorithms mentioned by Hickey et al. [24], when it is seen as an algorithm to compute a real function f.

Notice that if f is continuous in the usual sense, then for each $\overrightarrow{X} \in \mathbb{U}^n$, the interval function $\widehat{f}$ applied to $\overrightarrow{X}$ coincides with the image of f when applied to $\overrightarrow{X}$, that is, $\widehat{f}(\overrightarrow{X}) = f(\overrightarrow{X})$, where $f(\overrightarrow{X}) = \{f(\overrightarrow{x}) \mid \overrightarrow{x} \in \overrightarrow{X}\}$.

In this paper we will take into consideration the following notions of continuity for interval functions:

- *Moore* continuity: It is obtained as an extension of the continuity in the reals, considering the metric $d(X, Y) = \max(|\ \underline{X} - \underline{Y}\ |, |\ \overline{X} - \overline{Y}\ |)$ defined over $\mathbb{U}$.
- *Scott* continuity: The set $\mathbb{U}$ with reverse inclusion order is defined as a continuous domain [25], and a function $f : (\mathbb{U}, \supseteq) \to (\mathbb{U}, \supseteq)$ is Scott continuous if it is monotonic and preserves the least upper bound of directed sets [25].

The main result in [23] can be adapted to our context, considering U^n instead of $\mathbb{R}$, as shown in the following:

Theorem 4 *Let $f : U^n \longrightarrow U$ be a function. The following statements are equivalent: (i) f is continuous; (ii) $\widehat{f}$ is Scott continuous; (iii) $\widehat{f}$ is Moore continuous.*

3 Interval t-norms

A *triangular norm* (t-norm, for short) is a function $T : U^2 \to U$ that is commutative, associative, monotonic and has 1 as neutral element [26]. In the following definition, a natural extension of the t-norm notion for $\mathbb{U}$ is considered, following the approach introduced in [17].

Definition 5 *A function $\mathbb{T} : \mathbb{U}^2 \to \mathbb{U}$ is an* interval t-norm *if it is commutative, associative, monotonic with respect to the product and inclusion order and $[1, 1]$ is a neutral element.*

An interval t-norm may be considered as an interval representation of a t-norm. This generalization fits with the fuzzy principle, which means that the interval membership degree may be thought as an approximation of the exact degree.

See [18] for the proofs of the next two propositions in this section. The following result shows how an interval t-norm can be constructed from two given t-norms.

Proposition 6 *A function $\mathbb{T} : \mathbb{U}^2 \to \mathbb{U}$ is an interval t-norm if and only if there exist t-norms $\underline{\mathbb{T}}$ and $\overline{\mathbb{T}}$ with $\underline{\mathbb{T}} \leq \overline{\mathbb{T}}$ and*

$$\mathbb{T}(X, Y) = [\underline{\mathbb{T}}(\underline{X}, \underline{Y}), \overline{\mathbb{T}}(\overline{X}, \overline{Y})]. \quad (3)$$

The next proposition states that the best interval representation of a t-norm is an interval t-norm. It also shows that the interval representation of a t-norm coincides with the interval construction of Prop. 6 when both t-norms are the same.

Proposition 7 *Let T be a t-norm. Then $\widehat{T} : \mathbb{U}^2 \to \mathbb{U}$ is an interval t-norm, such that*

$$\widehat{T}(X, Y) = [T(\underline{X}, \underline{Y}), T(\overline{X}, \overline{Y})]. \quad (4)$$

4 Fuzzy Implications and R-Implications

Several definitions for fuzzy implication together with related properties have been given (see, e.g., [21, 19, 20]). The unique consensus in these definitions is that the fuzzy implication should present the same behavior of the classical implication for the crisp case. Thus, a function $I : U^2 \longrightarrow U$ is a *fuzzy implication* if it satisfies the minimal boundary conditions:

$$I(1, 1) = I(0, 1) = I(0, 0) = 1 \text{ and } I(1, 0) = 0.$$

Several reasonable properties may be required for fuzzy implications. The properties considered in this paper are:

I1 : If $y \leq z$ then $I(x, y) \leq I(x, z)$; (2-place monotonicity)

I2 : $I(x, I(y, z)) = I(y, I(x, z))$ (exchange principle);

I3 : $I(x, y) = 1$ if and only if $x \leq y$;

I4 : $\lim_{n\to\infty} I(x, y_n) = I(x, \lim_{n\to\infty} y_n)$ (right-continuity);

I5 : If $x \leq z$ then $I(x, y) \geq I(z, y)$; (1-place antitonicity)

Proposition 8 *[19, Lemma 1 (xi)] If I is a fuzzy implication satisfying* **I1**, **I2** *and* **I3**, *then I also satisfies* **I5**.

[3] Notice that the concept of interval representation is different from interval extension and natural extension. [1, page 21]

[4] Notice that $\widehat{f}$ is the interval hull of the range of f.

ISBN: 978-989-95079-6-8

Considering a t-norm T, the equation

$$I_T(x,y) = \sup\{z \in [0,1] \mid T(x,z) \leq y\}, \quad (5)$$

defines a fuzzy implication, called *R-implication* or *residuum* of T [21]. R-implications arise from the notion of residuum in Intuitionistic Logic [21] or, equivalently, from the notion of residue in the theory of lattice-ordered semigroups [27]. Observe that an R-implication is well-defined only if the t-norm is left-continuous[5] [19]. This justifies the name "residuum of T", since an R-implication satisfies the residuation condition when the underlying t-norm is left continuous:

$$T(x,z) \leq y \text{ if and only if } I_T(x,y) \geq z. \quad (6)$$

Moreover, a t-norm T is left-continuous if and only if it satisfies the residuation condition [28].

The main results relating R-implications and the properties **I1 – I5** are presented in the following.

Theorem 9 *[29, Theorem 1.14] A fuzzy implication $I : U^2 \to U$ is an R-implication with a left-continuous underlying t-norm if and only if I satisfies the properties* **I1–I4**.

Since, by Prop. 8, **I5** follows directly from **I1 – I3**, then R-implications with left-continuous underlying t-norms satisfy the properties **I1 – I5**.

5 Interval-valued Fuzzy R-Implications

The minimal properties required for fuzzy implications can be naturally extended to consider interval fuzzy degrees, by using degenerate intervals. Thus, a function $\mathbb{I} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an *interval fuzzy implication* if the following conditions hold:

$$\mathbb{I}([1,1],[1,1]) = \mathbb{I}([0,0],[0,0]) = \mathbb{I}([0,0],[1,1]) = [1,1]; \quad (7)$$

$$\mathbb{I}([1,1],[0,0]) = [0,0]. \quad (8)$$

Other properties (see Sect. 4) can be naturally extended:

$\mathbb{I}$**1** : If $Y \leq Z$ then $\mathbb{I}(X,Y) \leq \mathbb{I}(X,Z)$;

$\mathbb{I}$**2** : $\mathbb{I}(X,\mathbb{I}(Y,Z)) = \mathbb{I}(Y,\mathbb{I}(X,Z))$;

$\mathbb{I}$3**a** : If $\mathbb{I}(X,Y) = [1,1]$ then $X \leq Y$;

$\mathbb{I}$3**b** : If $\overline{X} \leq \underline{Y}$ then $\mathbb{I}(X,Y) = [1,1]$;

$\mathbb{I}$**4** : $\mathbb{I}_Y(X) = \mathbb{I}(X,Y)$ is (Moore, Scott) continuous;

$\mathbb{I}$**5** : If $X \leq Z$ then $\mathbb{I}(X,Y) \geq \mathbb{I}(Z,Y)$,

It is always possible to obtain an interval fuzzy implication from any fuzzy implication canonically. Interval fuzzy implications also meet the optimality property and preserve the same properties satisfied by fuzzy implications. The proof of the following proposition is straightforward.

Proposition 10 *If I is a fuzzy implication then $\widehat{I}$ is an interval fuzzy implication. [18]*

The next theorem states that the best interval representation of a fuzzy implication preserves, in some sense, the properties **I1–I5** listed in Sect. 4.

Theorem 11 *Let I be a fuzzy implication satisfying* **I1**. *If I satisfies a property* **Ik**, *for $k = 1,\ldots,5$, then $\widehat{I}$ satisfies* $\mathbb{I}$**k**.

Proof:

[5] A t-norm T is said to be left-continuous whenever $\lim_{n\to\infty} T(x_n,y) = T(\lim_{n\to\infty} x_n, y)$. [26]

$\mathbb{I}$**1**: If I satisfies **I1**, then $\widehat{I}(X,Y) = [\inf\{I(x,\underline{Y})|x \in X\}, \sup\{I(x,\overline{Y})|x \in X\}]$ and $\widehat{I}(X,Z) = [\inf\{I(x,\underline{Z})|x \in X\}, \sup\{I(x,\overline{Z})|x \in X\}]$. Thus, if $Y \leq Z$ then, for each $x \in X$, it holds that $I(x,\underline{Y}) \leq I(x,\underline{Z})$ and $I(x,\overline{Y}) \leq I(x,\overline{Z})$. It follows that $\inf\{I(x,\underline{Y})|x \in X\} \leq \inf\{I(x,\underline{Z}) \mid x \in X\}$, $\sup\{I(x,\overline{Y})|x \in X\} \leq \sup\{I(x,\overline{Z})|x \in X\}$, and, thus, $\widehat{I}(X,Y) \leq \widehat{I}(X,Z)$.

$\mathbb{I}$**2**: If I satisfies **I2**, then, for each $x,y,z \in U$, one has that $I(x,I(y,z)) = I(y,I(x,z))$. Then, since I satisfies **I1**:

$$\begin{aligned}
&\widehat{I}(X,\widehat{I}(Y,Z))\\
&= \widehat{I}(X,[\inf\{I(y,z)) \mid y \in Y, z \in Z\},\\
&\qquad \sup\{I(y,z)) \mid y \in Y, z \in Z\}])\\
&= \widehat{I}(X,[\inf\{I(y,\underline{Z}))|y \in Y,\},\sup\{I(y,\overline{Z}))|y \in Y\}])\\
&= [\inf\{I(x,\inf\{I(y,\underline{Z}) \mid y \in Y\}) \mid x \in X\},\\
&\quad \sup\{I(x,\sup\{I(y,\overline{Z}) \mid y \in Y\}) \mid x \in X\}]\\
&= [\inf\{I(x,I(y,\underline{Z})) \mid y \in Y, x \in X\},\\
&\qquad \sup\{I(x,I(y,\overline{Z})) \mid y \in Y, x \in X\}]\\
&= [\inf\{I(y,I(x,\underline{Z})) \mid y \in Y, x \in X\},\\
&\qquad \sup\{I(y,I(x,\overline{Z})) \mid y \in Y, x \in X\}]\\
&= \widehat{I}(Y,\widehat{I}(X,Z))
\end{aligned}$$

$\mathbb{I}$3a: One has that if $\widehat{I}(X,Y) = [1,1]$ then $\inf\{I(x,y)|x \in X, y \in Y\} = 1 = \sup\{I(x,y)|x \in X, y \in Y\}$. However, this is only possible if $\{I(x,y)|x \in X, y \in Y\} = \{1\}$, and, thus, $I(x,y) = 1$, for each $x \in X, y \in Y$. Since I satisfies **I3**, then, for each $x \in X, y \in Y$, it is valid that $x \leq y$. This is only possible if $\overline{X} \leq \underline{Y}$, that is, $X \leq Y$.

$\mathbb{I}$3b: If $\overline{X} \leq \underline{Y}$ then, for each $x \in X$ and $y \in Y$, it is valid that $x \leq y$. So, one has that $\{I(x,y)|x \in X, y \in Y\} = \{1\}$, $\inf\{I(x,y)|x \in X, y \in Y\} = 1 = \sup\{I(x,y)|x \in X, y \in Y\}$. It follows that $\widehat{I}(X,Y) = [1,1]$.

$\mathbb{I}$**4**: For each $x \in X$, let $I_x : U \longrightarrow U$ be defined by $I_x(y) = I(x,y)$. Thus, I is right-continuous if and only if, for each $x \in X$, I_x is continuous. If I_x is continuous then, by Theorem 4, $\widehat{I}_x$ is Scott (Moore) continuous. Thus:

$$\begin{aligned}
&\widehat{I}(X,Y)\\
&= [\inf\{I(x,y)|x \in X, y \in Y\}, \sup\{I(x,y)|x \in X, y \in Y\}]\\
&= [\inf\{\inf\{I(x,y)|y \in Y\} \mid x \in X\},\\
&\qquad \sup\{\sup\{I(x,y)|y \in Y\}|x \in X\}]\\
&= \bigcup_{x\in X} [\inf\{I(x,y)|y \in Y\}, \sup\{I(x,y)|y \in Y\}] = \bigcup_{x\in X} \widehat{I}_x(Y)
\end{aligned}$$

and then, considering that I_x is (topologically) continuous and the union preserves continuity, one concludes that $\widehat{I}$ is also Scott (Moore) continuous.

$\mathbb{I}$**5**: If I satisfies **I5**, then $\widehat{I}(X,Y) = [\inf\{I(\overline{X},y)|y \in Y\}, \sup\{I(\underline{X},y)|y \in Y\}]$ and $\widehat{I}(Z,Y) = [\inf\{I(\overline{Z},y)|y \in Y\}, \sup\{I(\underline{Z},y)|y \in Y\}]$. Thus, if $X \leq Z$ then, for each $y \in Y$, it holds that $I(\underline{X},y) \leq I(\underline{Z},y)$ and $I(\overline{X},y) \leq I(\overline{Z},y)$. It follows that $\inf\{I(\overline{Z},y)|y \in Y\} \leq \inf\{I(\overline{X},y) \mid y \in Y\}$ and $\sup\{I(\underline{Z},y)|y \in Y\} \leq \inf\{I(\underline{X},y)|y \in Y\}$, and, thus, $\widehat{I}(X,Y) \geq \widehat{I}(Z,Y)$.

□

The next corollary indicates that the best interval representation of a fuzzy implication satisfying **I1–I3** satisfies the properties $\mathbb{I}$**1**–$\mathbb{I}$**3** and $\mathbb{I}$**5**.

Corollary 12 *Let $I : U^2 \longrightarrow U$ be a fuzzy implication satisfying* **I1**, **I2** *and* **I3**. *Then $\widehat{I}$ satisfies* $\mathbb{I}$**1**–$\mathbb{I}$**3** *and* $\mathbb{I}$**5**.

ISBN: 978-989-95079-6-8

Proof: It follows from Prop. 8 and Theorem 11. □

Next proposition provides, for the best interval representation of a fuzzy implication satisfying the properties **I1**, **I2** and **I3**, a more concrete and simpler characterization than Eq. (2).

Proposition 13 *Let $I : U^2 \longrightarrow U$ be a fuzzy implication satisfying the properties* **I1**, **I2** *and* **I3**. *Then a characterization of $\widehat{I}$ can be obtained as*

$$\widehat{I}(X,Y) = [I(\overline{X},\underline{Y}), I(\underline{X},\overline{Y})]. \quad (9)$$

Proof: By Prop. 8, I also satisfies **I5** and, therefore, it holds that $\widehat{I}(X,Y) = [I(\overline{X},\underline{Y}), I(\underline{X},\overline{Y})]$ (see also [18]). □

Theorem 14 *Let $\mathbb{I} : \mathbb{U}^2 \to \mathbb{U}$ be an interval fuzzy implication. If $\mathbb{I}$ satisfies $\mathbb{I}$**1** and $\mathbb{I}$**5** and is inclusion monotonic, then $\mathbb{I}$ is representable by the functions $\underline{\mathbb{I}}(x,y) = l(\mathbb{I}([x,x],[y,y]))$ and $\overline{\mathbb{I}} = r(\mathbb{I}([x,x],[y,y]))$, that is, $\mathbb{I}(X,Y) = [\underline{\mathbb{I}}(\overline{X},\underline{Y}), \overline{\mathbb{I}}(\underline{X},\overline{Y})]$.*

Proof: Given $X, Y \in \mathbb{U}$, since $[\overline{X},\overline{X}] \subseteq X, [\underline{Y},\underline{Y}] \subseteq Y$, then, by the $\subseteq$-monotonicity of $\mathbb{I}$, we have that $\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}]) \subseteq \mathbb{I}(X,Y)$, and, thus, $l(\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}])) \geq l(\mathbb{I}(X,Y))$. On the other hand, since $X \leq [\overline{X},\overline{X}]$ and $[\underline{Y},\underline{Y}] \leq Y$, then, by $\mathbb{I}$**1** and $\mathbb{I}$**5**, it holds that $\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}]) \leq \mathbb{I}(X,Y)$, and, thus, $l(\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}])) \leq l(\mathbb{I}(X,Y))$. It follows that $l(\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}])) = l(\mathbb{I}(X,Y))$. Analogously, since $[\underline{X},\underline{X}] \subseteq X$ and $[\overline{Y},\overline{Y}] \subseteq Y$, then, by the inclusion monotonicity of $\mathbb{I}$, we have that $\mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}]) \subseteq \mathbb{I}(X,Y)$, and, thus, $r(\mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}])) \leq r(\mathbb{I}(X,Y))$. On the other hand, since $[\underline{X},\underline{X}] \leq X$ and $Y \leq [\overline{Y},\overline{Y}]$, then, by $\mathbb{I}$**1** and $\mathbb{I}$**5**, it holds that $\mathbb{I}(X,Y) \leq \mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}])$, and, thus, $r(\mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}])) \geq r(\mathbb{I}(X,Y))$. One has that $r(\mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}])) = r(\mathbb{I}(X,Y))$, and then $\mathbb{I}(X,Y) = [\underline{\mathbb{I}}(\overline{X},\underline{Y}), \overline{\mathbb{I}}(\underline{X},\overline{Y})]$. □

Definition 15 *An interval fuzzy implication $\mathbb{I}$ is an* interval R-implication *if there is an interval t-norm $\mathbb{T}$ such that*

$$\mathbb{I} = \mathbb{I}_{\mathbb{T}}(X,Y) = \sup\{Z \in \mathbb{U} | \mathbb{T}(X,Z) \leq Y\}. \quad (10)$$

Analogously, given an interval fuzzy implication $\mathbb{I}$, one has that $\mathbb{T}_{\mathbb{I}}(X,Y) = \inf\{Z \in \mathbb{U} \mid \mathbb{I}(X,Z) \geq Y\}$.

In Eq. (10), the supremum is determined considering the product order, and, thus, it results from the supremum considering the usual order on the real numbers (the interval endpoints).

The following theorem provides necessary conditions for an interval implication to be an interval R-implication.

Theorem 16 *Let $\mathbb{T}$ be a (Moore, Scott) left-continuous interval t-norm. Then $\mathbb{I}_{\mathbb{T}}$ satisfies $\mathbb{I}$**1**, and $\mathbb{I}$**3**–$\mathbb{I}$**5**.*

Proof: It follows that :

$\mathbb{I}$**1**: If $Y \leq Z$ and $\mathbb{T}(X,Z') \leq Y$ then it holds that $\mathbb{T}(X,Z') \leq Z$. It follows that $\{Z' \in \mathbb{U} \mid \mathbb{T}(X,Z') \leq Y\} \subseteq \{Z' \in \mathbb{U} \mid \mathbb{T}(X,Z') \leq Z\}$, and, thus, $\mathbb{I}_{\mathbb{T}}(X,Y) \leq \mathbb{I}_{\mathbb{T}}(X,Z)$.

$\mathbb{I}$**3**: $\mathbb{I}_{\mathbb{T}}(X,Y)=[1,1] \Leftrightarrow \sup\{Z \in \mathbb{U} | \mathbb{T}(X,Z) \leq Y\}=[1,1] \Leftrightarrow \{Z \in \mathbb{U} | \mathbb{T}(X,Z) \leq Y\}=\mathbb{U} \Leftrightarrow X=\mathbb{T}(X,[1,1]) \leq Y$.

$\mathbb{I}$**4**: It follows from the Moore (Scott) left-continuity of $\mathbb{T}$.

$\mathbb{I}$**5**: If $X \leq X'$ then $\mathbb{T}(X,Z) \leq Y$ implies that $\mathbb{T}(X',Z) \leq Y$. Thus, one has that $\{Z \in \mathbb{U} \mid \mathbb{T}(X',Z) \leq Y\} \subseteq \{Z \in \mathbb{U} \mid \mathbb{T}(X,Z) \leq Y\}$ and then it holds that $\sup\{Z \in \mathbb{U} \mid \mathbb{T}(X',Z) \leq Y\} \leq \sup\{Z \in \mathbb{U} \mid \mathbb{T}(X,Z) \leq Y\}$. It follows that $\mathbb{I}_{\mathbb{T}}(X',Y) \leq \mathbb{I}_{\mathbb{T}}(X,Y)$.

□

In [21, Lemma 5.16], it was proved that $T = T_{I_T}$, where $T_{I_T}(x,y) = \inf\{z \in U \mid I_T(x,z) \geq y\}$. The following two lemmas and proposition shows that this result can be extended for interval R-implications.

Lemma 17 *Let T be a left-continuous t-norm and $\mathbb{T}$ be a Moore left-continuous interval t-norm. If $\widehat{I_T} = \mathbb{I}_{\mathbb{T}}$ then $\mathbb{I}_{\mathbb{T}}([x,x],Z) \geq [y,y]$ if and only if $\mathbb{I}_{\mathbb{T}}([x,x],[\underline{Z},\underline{Z}]) \geq [y,y]$.*

Proof: It follows that: $\mathbb{I}_{\mathbb{T}}([x,x],Z) \geq [y,y] \Leftrightarrow \widehat{I_T}([x,x],Z) \geq [y,y] \Leftrightarrow \{I_T(x,z) \mid z \in Z\} \geq [y,y] \Leftrightarrow I_T(x,\underline{Z}) \geq y \Leftrightarrow \widehat{I_T}([x,x],[\underline{Z},\underline{Z}]) \geq [y,y] \Leftrightarrow \mathbb{I}_{\mathbb{T}}([x,x],[\underline{Z},\underline{Z}]) \geq [y,y]$. □

Lemma 18 *Let T be a left-continuous t-norm and $\mathbb{T}$ be a Moore left-continuous interval t-norm. If $\widehat{I_T} = \mathbb{I}_{\mathbb{T}}$ then it holds that $\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([x,x],[y,y]) = \mathbb{T}([x,x],[y,y])$.*

Proof: It follows that:

$$\begin{aligned}
&\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([x,x],[y,y])\\
&= \inf\{Z \in \mathbb{U} \mid \mathbb{I}_{\mathbb{T}}([x,x],Z) \geq [y,y]\}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} \mid \mathbb{I}_{\mathbb{T}}([x,x],[\underline{Z},\underline{Z}]) \geq [y,y]\} \text{ by Lemma 17}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} | \sup\{Z' \in \mathbb{U} | \mathbb{T}([x,x],Z') \leq [\underline{Z},\underline{Z}]\} \geq [y,y]\}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} \mid \sup\{[\overline{Z'},\overline{Z'}] \in \mathbb{U} \mid \mathbb{T}([x,x],[\overline{Z'},\overline{Z'}])\\
&\qquad \leq [\underline{Z},\underline{Z}]\} \geq [y,y]\} \text{ by } \subseteq\text{-monotonicity of } \mathbb{T}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} \mid I_{\mathbb{T}}(x,\underline{Z}) \leq y \wedge I_{\overline{\mathbb{T}}}(x,\underline{Z}) \leq y\}\\
&= [T_{I_{\mathbb{T}}}(x,y), T_{I_{\overline{\mathbb{T}}}}(x,y)]=[\underline{\mathbb{T}}(x,y), \overline{\mathbb{T}}(x,y)]=\mathbb{T}([x,x],[y,y])
\end{aligned}$$

□

Proposition 19 *Let T be a left-continuous and $\mathbb{T}$ be a Moore left-continuous interval t-norms. If $\widehat{I_T} = \mathbb{I}_{\mathbb{T}}$ then $\mathbb{T}_{\mathbb{I}_{\mathbb{T}}} = \mathbb{T}$.*

Proof: It follows that:

$$\begin{aligned}
&\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}(X,Y)\\
&= [\underline{\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}}(\underline{X},\underline{Y}), \overline{\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}}(\overline{X},\overline{Y})]\\
&= [l(\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([\underline{X},\underline{X}],[\underline{Y},\underline{Y}]), r(\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([\overline{X},\overline{X}],[\overline{Y},\overline{Y}])\\
&= [l(\mathbb{T}([\underline{X},\underline{X}],[\underline{Y},\underline{Y}])), r(\mathbb{T}([\overline{X},\overline{X}],[\overline{Y},\overline{Y}]))] \text{ by Lemma 18}\\
&= [\underline{\mathbb{T}}([\underline{X},\underline{X}],[\underline{Y},\underline{Y}]), \overline{\mathbb{T}}([\overline{X},\overline{X}],[\overline{Y},\overline{Y}])] = \mathbb{T}(X,Y)
\end{aligned}$$

□

Proposition 20 *Let T be a left-continuous t-norm and $\mathbb{T}$ be a Moore left-continuous interval t-norm. If $\widehat{I_T} = \mathbb{I}_{\mathbb{T}}$ then $\mathbb{T}$ is an interval representation of T.*

Proof: It holds that

$$\begin{aligned}
&\mathbb{T}([x,x],[y,y])\\
&= \mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([x,x],[y,y]) \text{by Prop. 19}\\
&= \inf\{Z \in \mathbb{U} \mid \mathbb{I}_{\mathbb{T}}([x,x],Z) \geq [y,y]\}\\
&= \inf\{Z \in \mathbb{U} \mid \widehat{I_T}([x,x],Z) \geq [y,y]\} \text{ by hypothesis}\\
&= \inf\{Z \in \mathbb{U} \mid I_T(x,z) \geq y, \forall z \in Z\}\\
&= \inf\{Z \in \mathbb{U} \mid I_T(x,\underline{Z}) \geq y\} \text{ by } \mathbf{I1} \text{ and Theorem 9}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} \mid I_T(x,\underline{Z}) \geq y\}\\
&= [\inf\{\underline{Z} \in U \mid I_T(x,\underline{Z}) \geq y\}, \inf\{\underline{Z} \in U \mid I_T(x,\underline{Z}) \geq y\}]\\
&= [T_{I_T}(x,y), T_{I_T}(x,y)] = [T(x,y), T(x,y)].
\end{aligned}$$

By the $\subseteq$-monotonicity of $\mathbb{T}$, we have that $\mathbb{T}$ represents T. □

Theorem 21 *If T is a left continuous t-norm, then $\mathbb{I}_{\widehat{T}} \subseteq \widehat{I_T}$.*

Proof: It follows that:

$$\begin{aligned}
&\mathbb{I}_{\widehat{T}}(X,Y)\\
&= \sup\{Z \in \mathbb{U} | \widehat{T}(X,Z) \leq Y\} \text{ by Eq. (10)}\\
&= \sup\{[\underline{Z},\overline{Z}] \in \mathbb{U} | [T(\underline{X},\underline{Z}), T(\overline{X},\overline{Z})] \leq [\underline{Y},\overline{Y}]\} \text{ by Eq. (4)}\\
&= \sup\{[\underline{Z},\overline{Z}] \in \mathbb{U} | T(\underline{X},\underline{Z}) \leq \underline{Y} \wedge T(\overline{X},\overline{Z}) \leq \overline{Y}\} \text{ by Eq. (1)}\\
&= [\min\{\sup\{\underline{Z} \in U | T(\underline{X},\underline{Z}) \leq \underline{Y}\},\\
&\quad \sup\{\overline{Z} \in U | T(\overline{X},\overline{Z}) \leq \overline{Y}\}\}, \sup\{\overline{Z} \in U | T(\overline{X},\overline{Z}) \leq \overline{Y}\}]\\
&= [\min\{I_T(\underline{X},\underline{Y}), I_T(\overline{X},\overline{Y})\}, I_T(\overline{X},\overline{Y})] \text{ by Eq. (5)}\\
&\subseteq [I_T(\overline{X},\underline{Y}), I_T(\underline{X},\overline{Y})] = \widehat{I_T}(X,Y) \text{ by Theorem 9 and Eq. (9)}
\end{aligned}$$

ISBN: 978-989-95079-6-8

□

The following results states a sufficient condition to guarantee the desired equivalence $\mathbb{I}_{\widehat{T}} = \widehat{I_T}$.

Lemma 22 *Let T be left-continuous t-norm. Then, it holds that $\underline{\mathbb{I}_{\widehat{T}}} = I_T = \overline{\mathbb{I}_{\widehat{T}}}$.*

Proof: It follows that

$$\begin{aligned}\underline{\mathbb{I}_{\widehat{T}}}(x,y) &= l(\mathbb{I}_{\widehat{T}}([x,x],[y,y]))\\ &= l(\sup\{Z \in \mathbb{U} \mid \widehat{T}([x,x],Z) \leq [y,y]\})\\ &= l(\sup\{[\overline{Z},\overline{Z}] \in \mathbb{U} \mid \widehat{T}([x,x],[\overline{Z},\overline{Z}]) \leq [y,y]\})\\ &\qquad \text{by } \leq \text{ and } \subseteq\text{-monotonicity}\\ &= \sup\{\overline{Z} \in U \mid T(x,\overline{Z}) \leq y\} = I_T(x,y).\end{aligned}$$

The proof of the other equality is analogous. □

Proposition 23 *If T is left-continuous t-norm then $\widehat{T}$ is a Moore left-continuous interval t-norm.*

Proof: It is analogous to the proof of Theorem 11 ($\mathbb{I}$**4**). □

Theorem 24 *Let T be a left-continuous t-norm. If $\mathbb{I}_{\widehat{T}}$ is $\subseteq$-monotonic then*

$$\mathbb{I}_{\widehat{T}} = \widehat{I_T}. \tag{11}$$

Proof: By Theorem 9 and Prop. 8, I_T satisfies **I1** and **I5** and $\widehat{I_T}(X,Y) = [\inf\{I_T(x,y) \mid x \in X \wedge y \in Y\}, \sup\{I_T(x,y) \mid x \in X \wedge y \in Y\}] = [I_T(\overline{X},\underline{Y}), I_T(\underline{X},\overline{Y})]$. On the other hand, since $\mathbb{I}_{\widehat{T}}$ is $\subseteq$-monotonic, T is (Moore, Scott) left-continuous (by Prop. 23), and, by Theorem 16, it satisfies $\mathbb{I}$**1** and $\mathbb{I}$**5**. By Theorem 14, it follows that $\mathbb{I}_{\widehat{T}}(X,Y) = [\underline{\mathbb{I}_{\widehat{T}}}(\overline{X},\underline{Y}), \overline{\mathbb{I}_{\widehat{T}}}(\underline{X},\overline{Y})]$. Then, by Lemma 22, one has that $\mathbb{I}_{\widehat{T}}(X,Y) = [I_T(\overline{X},\underline{Y}), I_T(\underline{X},\overline{Y})]$. Therefore, it holds that $\mathbb{I}_{\widehat{T}}(X,Y) = \widehat{I_T}(X,Y)$. □

The above results state the commutativity of the diagram of Fig. 1, where $\mathcal{C}(T)$ ($\mathcal{C}(\mathbb{T})$) is the class of left-continuous (interval) t-norms and $\mathcal{C}(I)$ ($\mathcal{C}(\mathbb{I})$) is the class of ($\subseteq$-monotonic interval) R-implications.

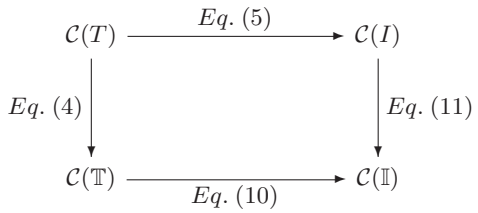


Figure 1: Commutative diagram relating R-implications with interval R-implications

6 Interval-valued Automorphisms

Definition 25 *[30] $\rho : U \longrightarrow U$ is an* automorphism *if it is bijective and monotonic.*[6] *The* action *of an automorphism ρ on a function $f : U^2 \to U$, denoted by f^ρ, is defined as $f^\rho(x,y) = \rho^{-1}(f(\rho(x),\rho(y)))$.*

As it is well known, if T is a t-norm then T^ρ is also a t-norm, and if I is an R-implication then I^ρ is also an R-implication.

$\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ is an *interval automorphism* if it is bijective and monotonic with respect to the product order [16]. The set of interval automorphisms is denoted by $Aut(\mathbb{U})$.

The next theorem shows that each interval automorphism can be constructed from an automorphism.

[6] In [19], $\rho : U \longrightarrow U$ is an automorphism if it is a continuous and strictly increasing function such that $\rho(0) = 0$ and $\rho(1) = 1$.

Theorem 26 *[16, Theorems 2 and 3] $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ is an interval automorphism if and only if there exists an automorphism $\rho : U \longrightarrow U$ such that $\varrho(X) = [\rho(\underline{X}), \rho(\overline{X})]$.*

Clearly, if $\rho : U \to U$ is an automorphism then $\widehat{\rho}$ can be obtained as $\widehat{\rho}(X) = [\rho(\underline{X}), \rho(\overline{X})]$. Therefore, interval automorphisms are the best interval representations of automorphisms.

Remark 1 *Let ϱ be an interval automorphism. Then: (i) ϱ^{-1} is an interval automorphism; (ii) $X \leq Y \Leftrightarrow \varrho(X) \leq \varrho(Y)$.*

Definition 27 *The* action *of an interval automorphism ϱ on an interval function $\mathbb{F}:\mathbb{U}^2 \longrightarrow \mathbb{U}$, denoted by $\mathbb{F}^\varrho$, is defined as*

$$\mathbb{F}^\varrho(X,Y) = \varrho^{-1}(\mathbb{F}(\varrho(X),\varrho(Y))). \tag{12}$$

In the following, we show how interval automorphisms act on interval t-norms and interval R-implications.

Proposition 28 *[17, Theorem 6.1] Let $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ be an interval automorphism and $\mathbb{T} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ be an interval t-norm. Then the mapping $\mathbb{T}^\varrho : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an interval t-norm.*

Theorem 29 *Let $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ be an interval automorphism and $\mathbb{T} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ be an interval t-norm. Then the mapping $\mathbb{I}_{\mathbb{T}}^\varrho : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is defined by $\mathbb{I}_{\mathbb{T}}^\varrho(X,Y) = \mathbb{I}_{\mathbb{T}^\varrho}(X,Y)$.*

Proof: Since ϱ is bijective and ϱ is monotonic, it follows that:

$$\begin{aligned}&\mathbb{I}_{\mathbb{T}}^\varrho(X,Y)\\ &= \varrho^{-1}(\mathbb{I}_{\mathbb{T}}(\varrho(X),\varrho(Y))) \text{ by Eq. (12)}\\ &= \varrho^{-1}(\sup\{Z \in \mathbb{U} | \mathbb{T}(\varrho(X),Z) \leq \varrho(Y)\}) \text{ by Eq. (10)}\\ &= \varrho^{-1}(\sup\{\varrho(Z') \in \mathbb{U} | \mathbb{T}(\varrho(X),\varrho(Z')) \leq \varrho(Y)\})\\ &= \sup\{\varrho^{-1}(\varrho(Z')) \in \mathbb{U} | \mathbb{T}(\varrho(X),\varrho(Z')) \leq \varrho(Y)\}\\ &= \sup\{Z' \in \mathbb{U} | \varrho^{-1}(\mathbb{T}(\varrho(X),\varrho(Z'))) \leq \varrho^{-1}(\varrho(Y))\} \text{ by Rem. 1}\\ &= \sup\{Z' \in \mathbb{U} | \varrho^{-1}(\mathbb{T}(\varrho(X),\varrho(Z'))) \leq Y\} \text{ by Rem. 1}\\ &= (\sup\{Z \in \mathbb{U} | \mathbb{T}^\varrho(X,Z) \leq Y\}) \text{ by Eq. (12)}\\ &= \mathbb{I}_{\mathbb{T}^\varrho}(X,Y) \text{ by Eq. (10)} .\end{aligned}$$

□

Corollary 30 *Let $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ be an interval automorphism and $\mathbb{T} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ be an interval t-norm. Then the mapping $\mathbb{I}_{\mathbb{T}}^\varrho : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an interval R-implication.*

Theorem 29 and Cor. 30 state that to apply an interval automorphism to an interval R-implication is the same that to apply it to an interval t-norm, and then to obtain an interval R-implication. Whenever an interval R-implication is submitted to an interval automorphism, a new interval R-implication is generated, which means that interval automorphisms may be applied in order to generate new interval R-implications. So, the commutative diagram pictured in Fig. 2 holds.

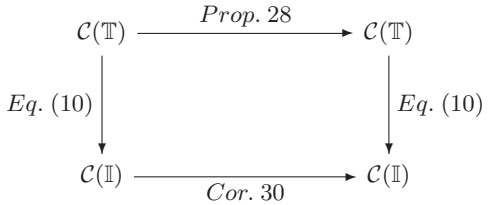


Figure 2: Commutative diagram relating interval t-norms, interval R-implications and interval automorphisms

ISBN: 978-989-95079-6-8

7 Conclusion and Final Remarks

The results presented in this paper extend our previous work (e.g., [17, 31]). We show that interval R-implications satisfy some analogous properties of R-implications. The best interval representation of an R-implication that is obtained from a left continuous t-norm coincides with the interval-valued R-implication obtained from the best interval representation of such t-norm, whenever this is a $\subseteq$-monotonic interval function. This provided, under this condition, a nice characterization for the best interval representation of an R-implication. Interval automorphisms are presented as best interval representations of automorphisms, showing that interval automorphisms act on interval R-implications, generating other interval R-implications.

Future work will consider the analysis of other important properties of interval-valued R-implications, in other to obtain a stronger relation between R-implications and interval-valued R-implications, establishing under which conditions interval R-implications are representable in the sense of [32].

Acknowledgment

This work is supported by CNPq (473201/07-0, 307185/07-9, 307879/06-2). We are grateful to the referees for their valuable suggestions, and also to Xiaohong Zhang for his comments.

References

[1] R.E. Moore. *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, 1979.

[2] R. B. Keafort and V. Kreinovich, editors. *Applications of Interval Computations*. Kluwer, Boston, 1996.

[3] L. A. Zadeh. Fuzzy sets. *Information and Control*, pages 338–353, 1965.

[4] S. Mitra and S. K. Pal. Fuzzy sets in pattern recognition and machine intelligence. *Fuzzy Sets and Systems*, 156:381–386, 2005.

[5] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - I. *Information Sciences*, 6:199–249, 1975.

[6] J. M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall, Upper Saddle River, 2001.

[7] W.A. Lodwick. Preface. *Reliable Computing*, 10(4):247–248, 2004.

[8] D. Dubois and H. Prade. Interval-valued fuzzy sets, possibility theory and imprecise probability. In *Proc. Intl. Conf. on Fuzzy Logic and Technology*, pages 314–319, Barcelona, 2005.

[9] I. Grattan-Guiness. Fuzzy membership mapped onto interval and many-valued quantities. *Z. Math. Logik. Grundladen Math.*, 22:149–160, 1975.

[10] K.U. Jahn. Intervall-wertige mengen. *Math. Nach.*, 68:115–132, 1975.

[11] H. T. Nguyen, V. Kreinovich, and Q. Zuo. Interval-valued degrees of belief: applications of interval computations to expert systems and intelligent control. *Intl. Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 5(3):317–358, 1997.

[12] R. Sambuc. *Fonctions ϕ-floues. Application l'aide au diagnostic en pathologie thyroidienne*. PhD thesis, Univ. Marseille, Marseille, 1975.

[13] R. E. Moore and W. Lodwick. Interval analysis and fuzzy set theory. *Fuzzy Sets and Systems*, 135(1):5–9, 2003.

[14] G. Cornelis, G. Deschrijver, and E. E. Kerre. Advances and challenges in interval-valued fuzzy logic. *Fuzzy Sets and Systems*, 157:622–627, 2006.

[15] D. Dubois and H. Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, 12:87–101, 1991.

[16] M. Gehrke, C. Walker, and E. Walker. Some comments on interval valued fuzzy sets. *International Journal of Intelligent Systems*, 11:751–759, 1996.

[17] B. C. Bedregal and A. Takahashi. The best interval representation of t-norms and automorphisms. *Fuzzy Sets and Systems*, 157(24):3220–3230, 2006.

[18] B. C. Bedregal and A. Takahashi. Interval valued versions of t-conorms, fuzzy negations and fuzzy implications. In *Proc. of the IEEE Intl. Conf. on Fuzzy Systems, Vancouver, 2006*, pages 553–559. IEEE, 2006.

[19] H. Bustince, P. Burilo, and F. Soria. Automorphism, negations and implication operators. *Fuzzy Sets and Systems*, 134:209–229, 2003.

[20] R.R. Yager. On some new classes of implication operators and their role in approximate reasoning. *Information Sciences*, 167:193–216, 2004.

[21] M. Baczynski and B. Jayaramb. (S,N)- and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems*, 159:1836–1859, 2008.

[22] S. Gottwald and P. Hajek. Triangular norm-based mathematical fuzzy logics. In E. P. Klement and R. Mesiar, editors, *Logical, Algebraic, Analytic and Probabilistic Aspects of Triangular Norms*, pages 275–299. Elsevier, Amsterdam, 2005.

[23] R. H. N. Santiago, B. C. Bedregal, and B. M. Acióly. Formal aspects of correctness and optimality in interval computations. *Formal Aspects of Computing*, 18(2):231–243, 2006.

[24] T. Hickey, Q. Ju, and M. Emdem. Interval arithmetic: from principles to implementation. *Journal of the ACM*, 48(5):1038–1068, 2001.

[25] G. Gierz, K. Hofmann, K. Keimel, J. D. Lawson, M. Mislove, and D. Scott. *Continuous Lattices and Domains*. Cambridge Press, Cambridge, 2003.

[26] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer, Dordrecht, 2000.

[27] J. C. Fodor. Contrapositive symmetry of fuzzy implications. *Fuzzy Sets and Systems*, 69:141–156, 1995.

[28] J. C. Fodor. Left-continuous t-norms in fuzzy logics: an overview. *Journal of Applied Sciences at Budapest Tech Hungary*, 1(2):141–156, 2004.

[29] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, Dordrecht, 1994.

[30] E.P. Klement and M. Navara. A survey on different triangular norm-based fuzzy logics. *Fuzzy Sets and Systems*, 101:241–251, 1999.

[31] R. H. S. Reiser, G. P. Dimuro, B.C. Bedregal, and R.H.N. Santiago. Interval valued QL-implications. In D. Leivant and R. Queiroz, editors, *Logic, Language, Information and Computation*, number 4576 in LNCS, pages 307–321. Springer, Berlin, 2007.

[32] G. Deschrijver. A representation of t-norms in interval-valued l-fuzzy set theory. *Fuzzy Sets and Systems*, 159, 2008.

ISBN: 978-989-95079-6-8

Recognition and Teaching of Robot Skills by Fuzzy Time-Modeling

Rainer Palm, Bourhane Kadmiry, Boyko Iliev, and Dimiter Driankov
AASS, Dept. of Technology Orebro University SE-70182 Orebro, Sweden
Emails: rub.palm@t-online.de, {bourhane.kadmiry, boyko.iliev, dimiter.driankov}@aass.oru.se

Abstract – Robot skills are low-level motion and/or grasping capabilities that constitute the basic building blocks from which tasks are built. Teaching and recognition of such skills can be done by Programming-by-Demonstration approach. A human operator demonstrates certain skills while his motions are recorded by a data-capturing device and modeled in our case via fuzzy clustering and Takagi-Sugeno modeling technique. The resulting skill models use the time as input and the operator's actions and reactions as outputs. Given a test skill by the human operator the robot control system recognizes the individual phases of skills and generates the type of skill shown by the operator.

Keywords – Fuzzy modeling, time clustering, robot skills, Programming-by-Demonstration

1 Introduction

Robot skills are low-level motion and/or grasping capabilities that constitute the basic building blocks from which tasks are built. One major challenge in robot programming is to be able to program skills in an easy and fast manner with high accuracy. *Programming by Demonstration (PbD)* (PbD) is one such approach that has the afore mentioned properties. In PbD a human who demonstrates the skills is equipped with data-capturing devices (e.g., data glove, cameras, haptic devices etc.). The demonstrator performs a skill while the robot captures the associated motion data, analyzes it and generates a robot-centered model of the demonstrated skill, that is a corresponding robot skill. Once the robot has acquired a number of robot skills from demonstrations it is able to recognize a demonstrated human skill as one of its already available robot skills. In a final step, when a task is demonstrated then the robot recognizes the robot skills that constitute it and thus creates a program consisting of these skills. This approach can be used not only for industrial robots but also in the fields of prosthetics, humanoid service robots, remote control and teleoperation in hasardous and dangerous environments, and last but not least in the entertainment industry. However, such applications are relatively few so far due to the lack of appropriate sensor systems and some unsolved problems with the man-robot interaction. Selected skills are

- contour following
- assembly (peg-in-hole insertion)
- handling of objects
- grasping of objects.

Different techniques for recognition of skills have been applied for PbD. For the manipulation domain Morrow and Khosla describe the construction of a library of robot capabilities by analysis and identification of tasks using a camera and a force-torque sensor [1]. Their approach is to develop a sensorimotor layer which integrates sensing into the robot programming primitives. In [2] Kaiser and Dillmann describe a neural net approach for the initial skill learning and reinforcement learning, skill refinement and adaptation. Experimental results were shown by an insertion example and a door-opening experiment. In the context of task learning Geib et. al. proposed an approach to integrating high-level artificial intelligence planning technology with low-level robotic control [3]. Kwun Han and M. Veloso describe an automated recognition of the behavior of robots using HMMs to represent and recognize strategic behaviors of robotic agents [4]. In the field of recognition of robot behaviors the following publications are important: Zoellner et. al. [5] use a data glove with integrated tactile sensors for behavior recognition which is based on support vector machines (SVM). Ekvall and Kragic [6] apply Hidden Markov Models (HMM) and address the PbD-problem using the arm trajectory as an additional feature for grasp classification. Li et. al. [7] use the singular value decomposition (SVD) for the generation of feature vectors of human grasps and support vector machines (SVM) which are applied to the classification problem. A fuzzy logic approach for gesture recognition was published by Bimber [8]. The method is applied to 6 d.o.f. (degrees of freedom) trajectories of a human arm but cannot cope with more than 6 dimensions. Despite the advanced state of the art, the cited methods on dynamic classification do not consider the evolution of a robot behavior in time and space. This appears to be a disadvantage because neither the estimation of the occurence of a specific skill nor the problem of segmentation can be solved in a general way. The method presented in this paper tries to overcome some of these drawbacks by modeling trajectories in time and space. Its advantage over other methods like HMM or Gaussian Mixture Models is discussed in [9] and [10]. To solve a 'fuzzy' task like skill recognition it turns out that fuzzy modeling is a suitable approach to deal with. The main focus of this method is *programming by demonstration* and *recognition of human skills* using fuzzy clustering and Takagi-Sugeno (TS) fuzzy modeling (see also [11] and [12]). The paper is organized as follows: In Section 2 a general approach to learning skills by partitioning them into phases is discussed. Section 3 deals with fuzzy time-modeling and the segmentation principle. Section 4 describes the recognition of phases and the decision process for the classification of skills. Section 5 presents simulations and experimental results. The final Section 6 draws some conclusions and directions for future work.

ISBN: 978-989-95079-6-8

2 Programming of robot skills by human demonstration

Programming of robot skills requires the building of a library of models of skills being taught or trained by a human demonstrator. In a next step newly demonstrated skills lead to test models which are then compared with the training models (skills) of the library. By such a comparison the robot is able to recognize these newly demonstrated skills. Finally a robot task including the recognized skills can automatically be generated (see Fig. 1). For the training phase two main tasks are needed to perform: *segmentation* of human demonstrations into skill phases and *phase modeling* of the skill phases. *Segmentation* means a partition of the data record into a sequence of episodes, where each one contains a single skill phase. For the test phase three main tasks are needed to perform: *segmentation* of the human test demonstrations, *phase recognition*, and *skill classification*. *Phase recognition* means to recognize the phases performed in each episode. The third task is to connect the recognized skill phases in such a way that a full skill can be identified.

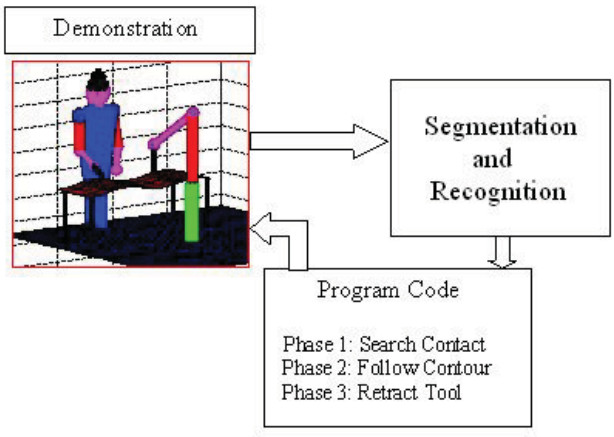


Figure 1: Learning skills from human demonstrations

3 Fuzzy time-modeling and the segmentation principle

Let a skill be partitioned into a sequence of phases as described above. Each phase starts and ends with a discrete event coming either from a discrete sensor or from some appropriate preprocessing of continuous sensor signals. The structure of a skill can be described most appropriately by a hybrid automaton in which nodes represent continuous phases and arcs the discrete transitions (switches) between them. Figure 2 shows an example of such a hybrid automaton. The hybrid process is event-controlled (see [13]) and is assumed to be stable both within the individual phases and with respect to the switching behavior between them. On the other hand the purpose of segmentation is to identify the discrete instants of time for the switches to occur in order to cut the whole skill into phases during demonstration.

The following subsection deals with fuzzy time-modeling in general that is used both for phase modeling and for segmentation. In the same context the training of time cluster models using new data is described. After that the segmentation procedure is presented.

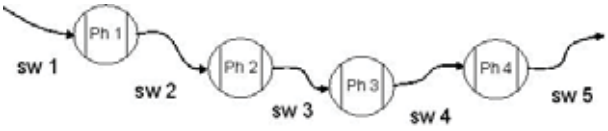


Figure 2: Hybrid automaton of a skill

3.1 Fuzzy time-modeling

Let us for the time being only concentrate on the modeling of a skill phase. The recognition of a skill phase is achieved by a model that reflects the *behavior of the robot end-effector in time* during the episode considered. Each demonstration is repeated several times to collect enough samples of every particular skill phase. From those data, models for each individual phase are developed using fuzzy clustering and Takagi-Sugeno fuzzy modeling ([14, 11]). We consider time instants as model inputs and end-effector coordinates as model outputs. Define the end-effector coordinate by

$$\mathbf{x}(t) = \mathbf{f}(t) \tag{1}$$

where $\mathbf{x}(t) \in R^3, \mathbf{f} \in R^3$, and $t \in R^+$. Further linearize (1) at selected time points t_i

$$\mathbf{x}(t) = \mathbf{x}(t_i) + \frac{\Delta \mathbf{f}(t)}{\Delta t}|_{t_i} \cdot (t - t_i) \tag{2}$$

which is a linear equation in t,

$$\mathbf{x}(t) = \mathbf{A}_i \cdot t + \mathbf{d}_i \tag{3}$$

where $\mathbf{A}_i = \frac{\Delta \mathbf{f}(t)}{\Delta t}|_{t_i} \in R^3$ and $\mathbf{d}_i = \mathbf{x}(t_i) - \frac{\Delta \mathbf{f}(t)}{\Delta t}|_{t_i} \cdot t_i \in R^3$. Using (3) as a local linear model one can express (1) in terms of a Takagi-Sugeno fuzzy model [15]

$$\mathbf{x}(t) = \sum_{i=1}^{c} w_i(t) \cdot (\mathbf{A}_i \cdot t + \mathbf{d}_i) \tag{4}$$

$w_i(t) \in [0, 1]$ is the degree of membership of a time point t to a cluster with the cluster center t_i, c is the number of clusters, and $\sum_{i=1}^{c} w_i(t) = 1$.

Let $\mathbf{x} = [x_1, x_2, x_3]^T$ be the 3 end-effector cordinates and t the time. The general clustering and modeling steps are described as follows

- Select an appropriate number of local linear models (data clusters) c
- Find c cluster centers $(t_i, x_{1i}, x_{2i}, x_{3i})$, $i = 1...c$, in the product space of the data quadruples (t, x_1, x_2, x_3) by Fuzzy-c-elliptotype clustering
- Find the corresponding fuzzy regions in the space of input data (t) by projection of the clusters of the product space into Gustafson-Kessel clusters (GK) within the input space [16]
- Calculate c local linear (affine) models (4) using the GK clusters from step 2.

ISBN: 978-989-95079-6-8

The membership degree $w_i(t)$ of an input data point t in an input cluster C_i is then calculated by

$$w_i(t) = \frac{1}{\sum_{j=1}^{c} \left(\frac{(t-t_i)^T M_{i_{pro}} (t-t_i)}{(t-t_j)^T M_{j_{pro}} (t-t_j)} \right)^{\frac{1}{\tilde{m}_{proj}-1}}} \quad (5)$$

The projected cluster centers t_i and induced matrices $M_{i_{pro}}$ define the input clusters C_i ($i = 1 \dots c$). The parameter $\tilde{m}_{pro} > 1$ determines the fuzziness of an individual cluster. The spheres in Fig. 3 covering the trajectories represent the local fuzzy models. The stripes along the time coordinate represent projections of the local fuzzy models onto the time line.

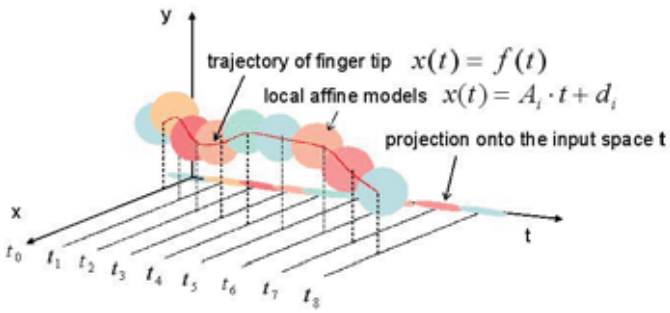


Figure 3: Time-clustering principle for the end-effector and its motion in (x,y)

3.1.1 *Training of time cluster models using new data*

A skill model can be built in several ways

- A single user trains the model by repeating the same skill n times
- m users train the model by repeating the same skill n times

The 1st model is generated by the time sequences $[(t_1, t_2, ..., t_N)_1 ... (t_1, t_2, ..., t_M)_n]$ and the output (end-effector position) sequences $[(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)_1 ... (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)_n]$.

The 2nd model is generated by the time sequences $[((t_1, t_2, ..., t_N)_1^1 ... (t_1, t_2, ..., t_M)_n^1) \; ... \; ((t_1, t_2, ..., t_N)_1^m ... (t_1, t_2, ..., t_M)_n^m)]$ and the output sequences $[((\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)_1^1 ... (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)_n^1) \; ... \; ((\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)_1^m ... (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)_n^m)]$ where m is the number of users in the training process, N, M are lengths of time sequences where $N \approx M$.

Once a particular skill model has been generated it might be necessary to take new data into account. These data may originate from different human operators to cover several ways of performing the same type of skill. Let for simplicity the old model be built by a time sequence $[t_1, t_2, ..., t_N]$ and a respective output sequence $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$. The old model is then represented by the input cluster centers t_i and the output cluster centers $\mathbf{x}_i$ ($i = 1...c$). It is also described by the parameters $\mathbf{A}_i$ and $\mathbf{d}_i$ of the local linear models. Let $[\tilde{t}_1, \tilde{t}_2, ..., \tilde{t}_M]$, $[\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, ..., \tilde{\mathbf{x}}_M]$ be new training data. A new model can be built by "chaining" old and new training data leading for the time sequences to $[t_1, t_2, ..., t_N, \tilde{t}_1, \tilde{t}_2, ..., \tilde{t}_M]$, and for the output sequences to $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N, \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, ..., \tilde{\mathbf{x}}_M]$. The result is a model that involves properties of the old model and the new data. If the old sequence of data is not available, a corresponding sequence can be generated by running the old model with the time instants $[t_1, t_2, ..., t_N]$ as inputs and the end-effector positions $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$ as outputs.

3.2 *Segmentation principle*

Let for simplicity the signals of a skill be the end-effector coordinates $\mathbf{x}(t) \in R^3$ and the forces $\mathbf{f}(t) \in R^3$ in the end-effector. In order to generate the 'events' determining the time bounds for the phases, $\mathbf{x}(t)$ is differentiated twice and $\mathbf{f}(t)$ only ones by time. The absolute values of the resulting vectors are collected in

$$\mathbf{X}(t) = [|\ddot{\mathbf{x}}(t)|^T, |\dot{\mathbf{f}}(t)|^T]^T. \quad (6)$$

For a segmentation we need the time-discrete case

$$\tilde{\mathbf{X}} = [\mathbf{X}(t_1) ... \mathbf{X}(t_n)] \in R^{6 \times n}. \quad (7)$$

Further define a vector of bounds $\mathbf{B} > 0 \in R^6$ above which $\mathbf{X}(t_i)$ are counted as 'events'. Then a vector $\mathbf{I} = [I_1 ... I_k ... I_m]^T$ is generated where I_k are discrete time stamps t_i for which at least one component of $\mathbf{X}(t_i)$ lies above the corresponding component of the vector of bounds $\mathbf{B}$.

$$I_k = t_i \quad if \quad \mathbf{X}(t_i) > \mathbf{B}. \quad (8)$$

The next step is to select the number of time clusters c and find the clusters by time clustering for the data $\mathbf{Y} = ([\mathbf{X}(I_1); I_1] ... [\mathbf{X}(I_k); I_k] ... [\mathbf{X}(I_m); I_m])$. $\mathbf{Y}$ is a combination of 'events' $\mathbf{X}(t_i)$ and their corresponding time instants I_k. Once the time clusters are found the skill could be cut at these time intants into phases. However, for complicated skills the number of phases c might not be known in advance. Therefore we choose a higher number c and merge those cluster centers who are located close to each other into one.

4 Recognition of robot skills

In this section the recognition of phases or sub-skills ist discussed first. In a second step on the basis of this knowledge - number and type of phases - the construction of the full skill is presented.

4.1 *Recognition of phases using the distance between fuzzy clusters*

Let the model of each phase have the same number of clusters $i = 1...c$ so that each duration T_l ($l = 1...L$) of the l-th phase is divided into $c - 1$ time intervals Δt_i, $i = 2...c$ of the same length. Let the phases be executed in an environment comparable with the modeled phase in order to avoid calibration and re-scaling procedures. Furthermore let

$$V_{model\,l} = [\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_c]_{model,l}$$
$$\mathbf{X}_i = [x, y, z, f_x, f_y, f_z]_i^T$$

where matrix $V_{model\,l}$ includes the output cluster centers $\mathbf{X}_i$ for the l-th phase model.

A model of the phase to be classified is built by the matrix

$$V_{test} = [\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_c]_{test,l} \quad (9)$$

ISBN: 978-989-95079-6-8

A decision about which phase is present is made by applying the Euclidean matrix norm

$$N_l = ||V_{model_l} - V_{test}|| \quad (10)$$

Once the unknown phase is classified to the phase model with the smallest norm $min(N_l), l = 1...L$ then the recognition of the phase is finished.

4.2 Recognition of skills using phase models

Once the phases of a test skill are recognized (identified) one should be able to recognize the skill as a whole and finally to reconstruct the hybrid automaton that represents the skill (see Fig. 2). For this purpose a list of possible skills and their phases should be produced. In the following we will discuss the following robot skills
- handling
- contour following
- assembly

The corresponding phases can be found in table 1. Figures 4,5, and 6 show the correspondence between the skills and their individual phases.

Table 1: Skills and phases

Phases Skills:	handling	contour	assembly
1. Grasp object	x		x
2. Free motion	x		
3. Search contact	x	x	x
4. Keep contact	x		
5. Follow with contact	x	x	x
6. Peg-in hole			x
7. release object/contact	x	x	x

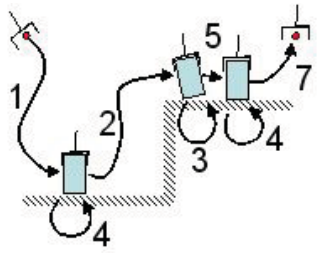


Figure 4: Handling skill

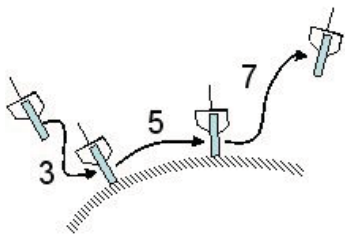


Figure 5: Contour following skill

5 Experiments and simulations

In this section an experimental evaluation of the recognition of phases will be presented. The experimental platform comprises a data glove with diodes mounted at the fingertips and links (see Fig. 7). A system of 4 stereo cameras takes records

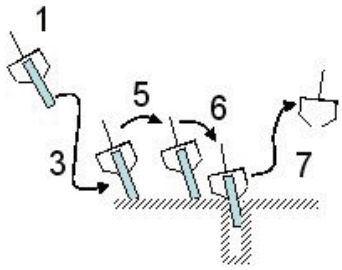


Figure 6: Assembly skill

of the positions of the diodes so that the position of the hand and its fingers can be tracked. In addition, tactile sensors are mounted at each finger tip in order to detect the contact between the fingertips and an object or a surface, respectively. In the experiment only the tip of the index finger is tracked in order to identify the contour performed by this finger. The experiments described here cover only contour following examples but using different contours running at different speeds. Each example consists of three phases: the approach phase, the contour following phase, and the retract phase. The experiment starts with the index fingertip being in contact with a defined start location at a distance from the contour. Then it follows the approach phase without any contact to an object ending at the begin of the contour to be followed. In the contour following phase the index finger follows the contour while the contact is preserved until the end of the contour is reached and the retract phase starts. During the retract phase there is no contact until the index finger reaches the start location again. The experiments can be divided into 3 groups:

A Straight lines
- 1: slow speed (see Fig. 8)
- 2: fast speed
- 3: ramp downhill slow
- 4: ramp downhill fast
- 5: ramp uphill slow
- 6: ramp uphill fast

B Meander
- 7: meander slow (see Fig. 9)
- 8: meander fast

C Loops
- 9: loop slow 1 (see Fig. 10)
- 10: loop slow 2
- 11: loop fast 1

Three modeling examples are shown in Figs. 8 - 10. The blue curves represent the modeled phases whereas the red curves represent the original data. Each phase of a skill is modeled by 15 cluster centers. The crosses depict the cluster centers. It can be observed that the modeling/approximation quality of the fuzzy time models is excellent. The partition of the skill into phases has been done by means of the forces applied to the tip of the index finger. Figure 11 shows the time plots for the meander experiment 7. By means of the force f applied to the fingertip and its derivative df a segmen-

ISBN: 978-989-95079-6-8

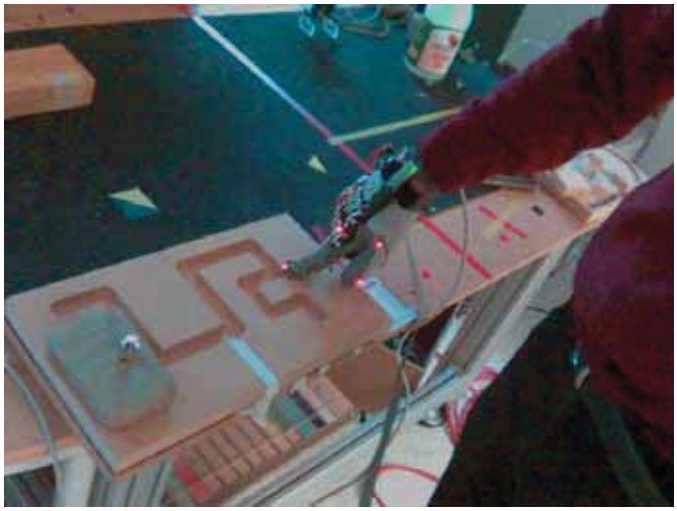


Figure 7: Experiment with a meader-like contour

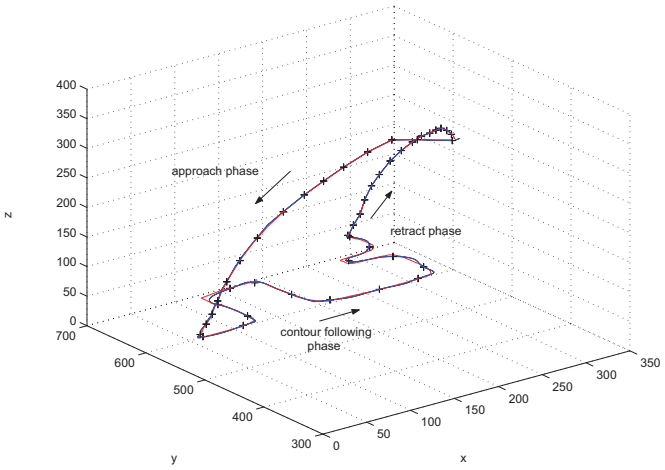


Figure 9: contour following, meander

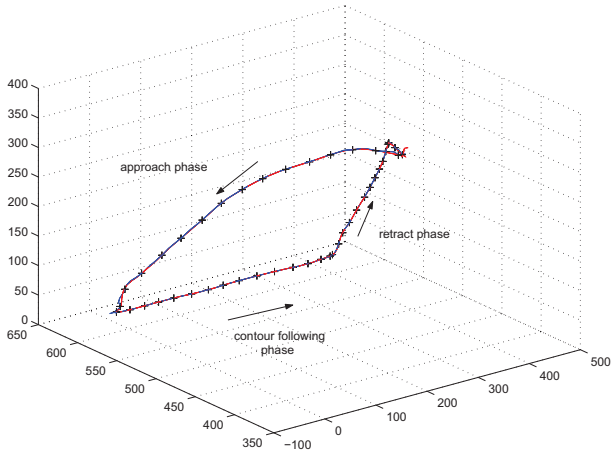


Figure 8: contour following, line

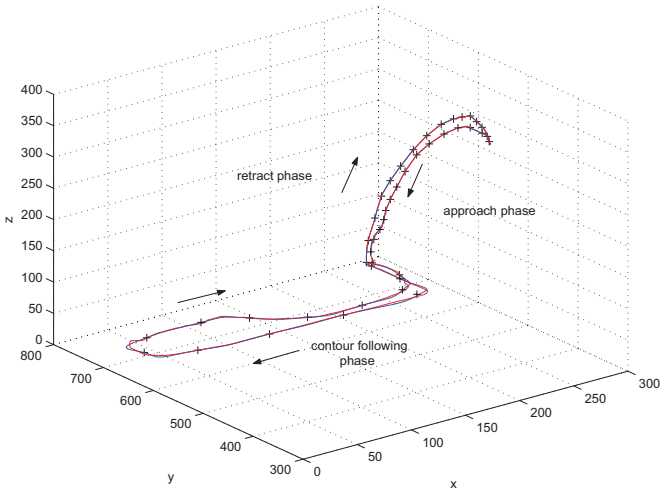


Figure 10: contour following, loop

tation can be done very easily because of the distinct derivatives of the force signals *df*. The recognition of skills has been done by comparing each skill with all other skills using the method described in the last section. Since we only deal with contour following experiments the norms over the whole skill have been taken into account instead of considering the phases seperately. The results are shown in Table 2. To explain the matrix $M(i,j)$ $(i,j = 1...11)$ in Table 2, let us consider experiment 3, *ramp downhill slow*, as an example (3rd row), $M(3,i)$ $(i = 1...11)$. Compared to itself the norm of the differences between model and test skill is zero, $M(3,3) = 0$. The next higher norm difference can be observed for experiment 4, *ramp downhill fast*, $M(3,4) = 0.3$. This corrresponds completely with the idea that a similarity of trajectories should lead to small norm differences. Going through all 11 experiments it turns out that almost all contour following skills can be indentified. One exception is experiment 6 where either skill 1 or 2 are identified instead of skill 5 as expected, $M(6,1) \wedge M(6,2) < M(6,5)$. It can also be noticed that the groups "A: Straight lines", $M(i,j)$ $(i,j = 1...6)$, "B: Meander" , $M(k,l)$ $(k,l = 7,8)$, and "C: Loops" , $M(m,n)$ $(m,n = 9...11)$, can be significantly distinguished from each other. Furthermore, one can see that, from the recognition point of view, the groups A and B are more related with each other than B and C or A and C, respectively.

6 Conclusions

In this paper modeling and recognition/identification of robot skills using fuzzy time modeling have been presented. Recognition of skills is a part of the teach-in process and the Programming-by-Demonstration (PbD) of robot tasks by a human operator. In this paper the focus is directed to the skills "handling", "contour following", and "assembly". In this context a partitioning of skills into phases by segmentation and their modeling using fuzzy time clustering is discussed. The recognition of phases and skills is done by comparing fuzzy time clusters of model skills and test skills. The experimental platform consists of a data glove with diodes mounted at the fingertips and links. 4 stereo cameras track the positions of the hand and its fingers. Tactile sensors are mounted at each fingertip to detect the contact between the fingertips and objects. In the experiments different groups of contour following skills are presented. In 11 experiments several contours like "straight line", "meander", or "loop", respectively,

ISBN: 978-989-95079-6-8

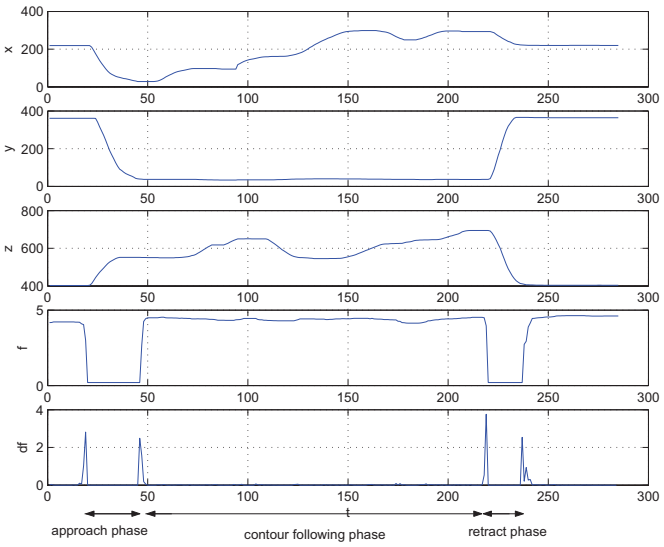


Figure 11: contour following, loop, time plots

Table 2: identification results

skill	1	2	3	4	5	6	7	8	9	10	11
1	0	0.2	0.6	0.5	0.6	0.4	0.7	0.9	1.8	1.8	1.8
2	0.2	0	0.7	0.6	0.6	0.5	0.7	0.9	1.8	1.9	1.8
3	0.5	0.7	0	0.3	0.8	0.6	0.9	1.0	1.9	1.9	1.8
4	0.5	0.6	0.3	0	0.8	0.6	0.8	0.9	1.8	1.9	1.8
5	0.6	0.6	0.8	0.8	0	0.5	0.9	1.0	1.9	2.0	1.9
6	0.4	0.5	0.6	0.6	0.5	0	0.8	1.0	1.8	2.0	1.9
7	0.7	0.7	0.9	0.8	0.9	0.8	0	0.4	1.4	1.4	1.4
8	0.9	0.9	1.0	0.9	1.0	1.0	0.4	0	1.5	1.4	1.3
9	1.8	1.8	1.9	1.8	1.9	1.8	1.4	1.5	0	0.6	0.4
10	1.8	1.9	1.9	1.9	2.0	1.9	1.4	1.4	0.6	0	0.5
11	1.8	1.8	1.8	1.8	1.9	1.9	1.4	1.4	0.4	0.5	0

are performed by the index fingertip. The segmentation has been done by differentiation of the tactile forces. The modeling and recognition results were very good to excellent. In the future more experiments will be done for handling and assembly tasks. The recognition of skills using fuzzy time modeling and hybrid automata will be further developed. A further focus will be the integration of high-level AI planning techniques with low-level robotic control.

References

[1] J. D. Morrow and P. K. Khosla. Manipulation task primitives for composing robot skills. *Proceedings of the 1997 IEEE Intemational Conference on Robotics and Automation, Albuquerque, New Mexico*, pages 3354–3359, 1997.

[2] M. Kaiser and R. Dillmann. Building elementary robot skills from human demonstration. *Proceedings of the 1996 IEEE Intemational Conference on Robotics and Automation, Minneapolis, Minnesota*, 1996.

[3] C. Geib, K. Mourao, R. Petrick, N. Pugeault, M. Steedman, N. Krueger, and F. Woergoetter. Object action complexes as an interface for planning and robot control. In *Proc. IEEE RAS Int Conf. Humanoid Robots(Genova)*, Genova, Dec. 4-6 2006. IEEE, IEEE.

[4] Kwun Han and M. Veloso. Automated robot behavior recognition. *Proceedings of the 1999 ijcai Conference*, 1999.

[5] R. Zoellner, O. Rogalla, R. Dillmann, and J.M. Zoellner. Dynamic grasp recognition within the framework of programming by demonstration. In *Proceedings Robot and Human Interactive Communication,10th IEEE International Workshop*, Bordeaux, Paris, France, September 18-21 2001. IEEE.

[6] S. Ekvall and D. Kragic. Grasp recognition for programming by demonstration. In *Proceedings of International conference on robotics and automation, ICRA 2005*, Barcelona, Spain, 2005.

[7] Ch. Li, L. Khan, and B. Prabhaharan. Real-time classification of variable length multi-attribute motions. *Knowledge and Information Systems*, DOI 10.1007/s10115-005-0223-8, 2005.

[8] O. Bimber. Continuous 6d gesture recognition: A fuzzy-logic approach. *Proceedings of 7-th International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*, WSCG'99:24–30, 1999.

[9] R. Palm and B. Iliev. Grasp recognition by time clustering, fuzzy modeling, and hidden markov models (hmm) - a comparative study. In *Proceedings FUZZ-IEEE 2008 - IEEE International Conference on Fuzzy Systems*, Hong Kong, July 1-5 2008. IEEE.

[10] Z. Ju, H. Liu, X.Y. Zhu, and Y.L. Xiong. Dynamic grasp recognition using time clustering, gaussian mixture models and hidden markov models. In *Proceedings Intern. Conf. on Intell. Robots and Appl.s, ICIRA-2008*, Wuhan,China, Oct. 15-17 2008. Springer.

[11] R. Palm and B. Iliev. Learning of grasp behaviors for an artificial hand by time clustering and takagi-sugeno modeling. In *Proceedings FUZZ-IEEE 2006 - IEEE International Conference on Fuzzy Systems*, Vancouver, BC, Canada, July 16-21 2006. IEEE.

[12] R. Palm and B. Iliev. Segmentation and recognition of human grasps for programming-by-demonstration using time clustering and takagi-sugeno modeling. In *Proceedings FUZZ-IEEE 2007 - IEEE International Conference on Fuzzy Systems*, London, UK, July 23-26 2007. IEEE.

[13] M.S.Branicky, V.S.Borkar, and S.K.Mitter. A unified framework for hybrid control: Background, model, and theory. *Proceedings of the 11th Intern. Conference on Analysis ond Optimization of Systems. INRIA. France*, June 15-17, 1994.

[14] R. Palm and Ch. Stutz. Open loop dynamic trajectory generator for a fuzzy gain scheduler. *Engineering Applications of Artificial Intelligence*, Vol. 16:213–225, 2003.

[15] T. Takagi and M. Sugeno. Identification of systems and its applications to modeling and control. *IEEE Trans. on Syst., Man, and Cyb.*, Vol. SMC-15. No.1:116–132, January/February 1985.

[16] D.E. Gustafson and W.C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. *Proceedings of the 1979 IEEE CDC*, pages 761–766, 1979.

ISBN: 978-989-95079-6-8

A Consensus Model for Group Decision Making with Incomplete Unbalanced Fuzzy Linguistic Preference Relations

F.J. Cabrerizo[1] J.M. Moreno[2] E. Herrera-Viedma[3]

1.Department of Software Engineering and Computer Systems, Distance Learning University of Spain
28040 Madrid, Spain
2.Department of Information and Communication Engineering, University of Murcia
30100 Murcia, Spain
3.Department of Computer Science and Artificial Intelligence, University of Granada
18071 Granada, Spain
Email: cabrerizo@issi.uned.es, jmmoreno@um.es, viedma@decsai.ugr.es

***Abstract*— *We present a consensus model for group decision making with unbalanced fuzzy linguistic preference relations, i.e., assuming that the preferences are assessed on linguistic term sets whose terms are not symmetrically and uniformly distributed. This consensus model can manage incomplete information situations, that is, situations where the experts do not give all the preference values that they are usually requested. In addition, both consistency and consensus measures are used and it allows to achieve consistent solutions with a great level of agreement.*

***Keywords*—** Group decision making, consensus, unbalanced fuzzy linguistic preference relations, incomplete information, consistency.

1 Introduction

A Group Decision Making (GDM) problem is usually understood as a decision problem which consists in finding the best alternative(s) from a set of feasible alternatives according to the preferences provided by a group of experts characterized by their experience and knowledge. To do this, experts have to express their preferences by means of a set of evaluations over the set of alternatives. In this contribution, we assume that experts use preference relations [1, 2, 3] in an unbalanced fuzzy linguistic context [4, 5] (see Fig. 1).



Figure 1: Example of an unbalanced fuzzy linguistic term set.

In these problems, a difficulty that has to be addressed is the lack of information. Since each expert has his/her own experience concerning the problem being studied, there may be cases where an expert would not be able to express the preference degree between two or more of the available alternatives. This may be due to an expert not possessing a precise or sufficient level of knowledge of part of the problem, or because that expert is unable to discriminate the degree to which some options are better than others.

To solve GDM problems, the experts are faced by applying two processes before obtaining a final solution [6, 7]: *the consensus process* and *the selection process*. The consensus process is defined as a dynamic and iterative group discussion process, coordinated by a moderator helping experts to bring their opinions closer. If the consensus level is lower than a specified threshold, the moderator would urge experts to discuss their opinions further in an effort to bring them closer. Otherwise, the moderator would apply the selection process which consists in obtaining the final solution to the problem from the opinions expressed by the experts. Clearly, it is preferable that the experts achieve a great agreement among their opinions before applying the selection process and, therefore, we focus on the consensus process.

The aim of this paper is to present a consensus model to deal with GDM problems in which experts use incomplete unbalanced fuzzy linguistic preference relations (FLPRs) to provide their preferences. We use two kinds of consensus measures to guide the consensus reaching process, *consensus degrees*, which evaluate the agreement of all the experts, and *proximity measures*, which evaluate the agreement between the experts' individual opinions and the group opinion. However, this consensus model will not only be based on consensus measures but also on consistency measures. To compute them, first, all missing values are estimated using an estimation procedure based on the Tanino's consistency principle [3]. Both consistency and consensus measures are used to design a feedback mechanism, and, in such a way, we substitute the actions of the moderator and give advice to the experts on how they should change and complete their opinions to obtain a solution with a high consensus degree (making experts' opinions closer).

The rest of the paper is set out as follows. Section 2 deals with the preliminaries necessary to develop our consensus model. In Section 3, the consensus model for GDM problems with incomplete unbalanced FLPRs is presented. Finally, some concluding remarks are pointed out in Section 4.

2 Preliminaries

2.1 Methodology to Manage Unbalanced Fuzzy Linguistic Information

To manage unbalanced fuzzy linguistic information, we propose a methodology similar to those proposed in [4, 5]. This methodology is based on the transformation of the unbalanced fuzzy linguistic information in a *Linguistic Hierarchy (LH)* [8], which is the linguistic representation domain that allows us to develop comparison and combination processes of unbalanced fuzzy linguistic information.

A LH is a set of levels, where each level represents a linguistic term set with different granularity from the remaining

ISBN: 978-989-95079-6-8

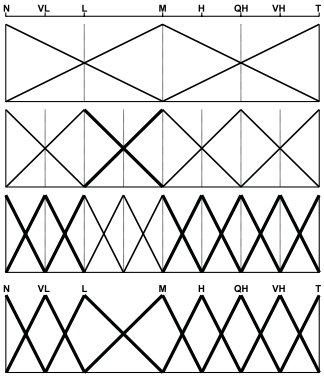


Figure 2: Representation for an unbalanced linguistic fuzzy term set.

levels of the hierarchy. Each level is denoted as $l(t, n(t))$, where t is a number indicating the level of the hierarchy, and $n(t)$ is the granularity of the linguistic term set of t. Then, a LH can be defined as the union of all levels t: $LH = \bigcup_t l(t, n(t))$. Given a LH, we denote as $\mathcal{S}^{n(t)}$ the linguistic term set of LH corresponding to the level t of LH characterized by a cardinality $n(t)$: $\mathcal{S}^{n(t)} = \{s_0^{n(t)}, \ldots, s_{n(t)-1}^{n(t)}\}$.

The procedure to represent unbalanced fuzzy linguistic information presents the following steps:

1. Find a level t^- of LH to represent the subset of linguistic terms $\mathcal{S}_{un}^L$ on the left of the mid linguistic term of unbalanced fuzzy linguistic term set $\mathcal{S}_{un}$.
2. Find a level t^+ of LH to represent the subset of linguistic terms $\mathcal{S}_{un}^R$ on the right of the mid linguistic term of $\mathcal{S}_{un}$.
3. Represent the mid term of $\mathcal{S}_{un}$ using the mid terms of the levels t^- and t^+.

If there does not exist a level t^- or t^+ in LH to represent $\mathcal{S}_{un}^L$ or $\mathcal{S}_{un}^R$, respectively, then the procedure applies the following recursive algorithm, which is defined, in this case, assuming that there does not exist t^-, as it happens with the unbalanced fuzzy linguistic term set given in Fig. 1:

1. Represent $\mathcal{S}_{un}^L$:
   (a) Identify the mid term of $\mathcal{S}_{un}^L$, called $\mathcal{S}_{mid}^L$.
   (b) Find a level t_2^- of the left sets of LH^L to represent the left term subset of $\mathcal{S}_{un}^L$, where LH^L represents the left part of LH.
   (c) Find a level t_2^+ of the right sets of LH^L to represent the right term subset of $\mathcal{S}_{un}^L$.
   (d) Represent the mid term $\mathcal{S}_{mid}^L$ using the levels t_2^- and t_2^+.
2. Find a level t^+ of LH to represent the subset of linguistic terms $\mathcal{S}_{un}^R$.
3. Represent the mid term of $\mathcal{S}_{un}$ using the levels t_2^+ and t^+.

For example, applying this algorithm, the representation of the unbalanced fuzzy linguistic term set $\mathcal{S}_{un} = \{N, VL, L, M, H, QH, VH, T\}$ shown in Fig. 1, using a linguistic hierarchy LH, would be as it is shown in Fig. 2.

To operate with the linguistic information in LH, the 2-tuple fuzzy linguistic model [9] is used.

Definition 2.1. Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value representing the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function $\Delta\colon [0, g] \longrightarrow S \times [-0.5, 0.5)$:

$$(\beta) = (s_i, \alpha), \text{ with } \begin{cases} s_i, & i = round(\beta) \\ \alpha = \beta - i, & \alpha \in [-0.5, 0.5), \end{cases} \quad (1)$$

where $round(\cdot)$ is the usual round operation, s_i has the closest index label to "β", and "α" is the value of the symbolic translation. In addition, for all Δ, there exists Δ^{-1}, defined as $\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$.

Finally, transformation functions between labels from different levels to make processes of computing with words in multigranular linguistic information contexts without loss of information were defined in [8].

Definition 2.2. [8] Let $LH = \bigcup_t l(t, n(t))$ be a linguistic hierarchy whose linguistic term sets are denoted as $\mathcal{S}^{n(t)} = \{s_0^{n(t)}, \ldots, s_{n(t)-1}^{n(t)}\}$, and let us consider the 2-tuple fuzzy linguistic representation. The transformation function from a linguistic label in level t to a label in level t' is defined as $TF_{t'}^t : l(t, n(t)) \longrightarrow l(t', n(t'))$ such that

$$TF_{t'}^t(s_i^{n(t)}, \alpha^{n(t)}) = \Delta_{t'}\left(\frac{\Delta_t^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \cdot (n(t') - 1)}{n(t) - 1}\right). \quad (2)$$

2.2 Incomplete Unbalanced FLPRs

In this paper, we deal with GDM problems where the experts e_h express their preferences relations $P^h = (p_{ik}^h)$ on the set of alternatives X using an unbalanced linguistic fuzzy term set, $\mathcal{S}_{un} = \{s_0, \ldots, s_{mid}, \ldots, s_g\}$, which has a minimum label, called s_0, a maximum label, called s_g, and the remaining labels are non-uniformly and non-symmetrically distributed around the central one, called s_{mid} (Fig. 1). Therefore, $p_{ik}^h \in \mathcal{S}_{un}$ represents the preference of alternative x_i over alternative x_k for the experts e_h assessed on the unbalanced fuzzy linguistic term set $\mathcal{S}_{un}$.

Definition 2.3. An unbalanced FLPR P^h on a set of alternatives X is characterized by a membership function $\mu_{P^h} : X \times X \longrightarrow \mathcal{S}_{un}$. If it is not possible to give the preference degree for every pair of alternatives, we have an incomplete unbalanced FLPR.

When cardinality of X is small, the preference relation may be conveniently represented by a $n \times n$ matrix $P^h = (p_{ik}^h)$, being $p_{ik}^h = \mu_{P^h}(x_i, x_k), \forall i, k \in \{1, \ldots, n\}$ and $p_{ik}^h \in \mathcal{S}_{un}$.

2.3 Consistency Measures

For GDM problems with preference relations, some properties are usually assumed desirable to avoid contradictions within the preferences expressed by the experts, that is, to avoid inconsistent opinions. One of them is the *additive transitivity*,

ISBN: 978-989-95079-6-8

$$cp_{ik}^h = TF_t^{t'}(\Delta_{t'}\left(\frac{\sum_{j=1;i\neq k\neq j}^{n}(\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j1})) + \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j2})) + \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j3})))}{3(n-2)}\right)). \quad (3)$$

$$cp_{ik}^h = TF_t^{t'}(\Delta_{t'}\left(\frac{\sum_{j\in H_{ik}^{h1}}\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j1})) + \sum_{j\in H_{ik}^{h2}}\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j2})) + \sum_{j\in H_{ik}^{h3}}\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j3}))}{(\#H_{ik}^{h1} + \#H_{ik}^{h2} + \#H_{ik}^{h3})}\right)). \quad (4)$$

which was defined for fuzzy preference relations [1, 3] as:

$$(p_{ij}^h - 0.5) + (p_{jk}^h - 0.5) = (p_{ik}^h - 0.5),\ \forall i,j,k \in \{1,\ldots,n\}. \quad (5)$$

In the case of an unbalanced fuzzy linguistic context, previously to carry out any computation task, we have to choose a level $t' \in \{t^-, t_2^-, t^+, t_2^+\}$, such that $n(t') = max\{n(t^-), n(t_2^-), n(t^+), n(t_2^+)\}$. Then, once a result is obtained, it is transformed to the correspondent level $t \in \{t^-, t_2^-, t^+, t_2^+\}$ by means of $TF_t^{t'}$ for expressing the result in the unbalanced fuzzy linguistic term set $\mathcal{S}_{un}$. In this way, the unbalanced fuzzy linguistic additive transitivity for unbalanced FLPRs is defined as:

$$\begin{array}{l} TF_t^{t'}(\Delta_{t'}[(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid})))+ \\ (\Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid})))]) = \\ TF_t^{t'}(\Delta_{t'}[(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid})))]), \end{array} \quad (6)$$

being $p_{ij}^h = (s_v^{n(t)}, \alpha_1)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, $p_{jk}^h = (s_w^{n(t)}, \alpha_2)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, $p_{ik}^h = (s_z^{n(t)}, \alpha_3)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, s_{mid} is the mid term of $\mathcal{S}_{un}$ and $t' \in \{t^-, t_2^-, t^+, t_2^+\}$.

Expression (6) can be rewritten as:

$$\begin{array}{l} p_{ik}^h = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \\ \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0)))),\ \ \forall i,j,k \in \{1,\ldots,n\}. \end{array} \quad (7)$$

Equation (7) can be used to calculate an estimated value of a preference degree p_{ik}^h $(i \neq k)$ using an intermediate alternative x_j in three different ways:

1. From $p_{ik}^h = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0))))$, we obtain:

$$\begin{array}{l} (cp_{ik}^h)^{j1} = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) + \\ \Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0)))). \end{array} \quad (8)$$

2. From $p_{jk}^h = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ji}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0))))$, we obtain:

$$\begin{array}{l} (cp_{ik}^h)^{j2} = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \\ \Delta_{t'}^{-1}(TF_{t'}^t(p_{ji}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0)))). \end{array} \quad (9)$$

3. From $p_{ij}^h = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(p_{kj}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0))))$, we obtain:

$$\begin{array}{l} (cp_{ik}^h)^{j3} = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) - \\ \Delta_{t'}^{-1}(TF_{t'}^t(p_{kj}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0)))). \end{array} \quad (10)$$

The overall estimated value cp_{ik}^h of p_{ik}^h is obtained as the average of all possible $(cp_{ik}^h)^{j1}$, $(cp_{ik}^h)^{j2}$ and $(cp_{ik}^h)^{j3}$ values as shown in (3).

When the information provided is completely consistent, then $(cp_{ik}^h)^{jl} = p_{ik}^h$, $\forall j,l$. The error between a preference value and its estimated one in $[0,1]$ is defined as follows:

$$\varepsilon p_{ik}^h = \frac{|\Delta_{t'}^{-1}(TF_{t'}^t(cp_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^h))|}{n(t') - 1}. \quad (11)$$

We should point out that some estimated values of an incomplete unbalanced FLPR could lie outside the $\mathcal{S}_{un}$, i.e., we may have $cp_{ik}^h < s_0$ or $cp_{ik}^h > s_g$. In order to normalize the expression domains, the following function is used:

$$f(cp_{ik}^h) = \begin{cases} s_0, & \text{if } cp_{ik}^h < s_0 \\ s_g, & \text{if } cp_{ik}^h > s_g \\ cp_{ik}^h, & \text{otherwise.} \end{cases} \quad (12)$$

Thus, it can be used to define the consistency level between the preference degree p_{ik}^h and the rest of the preference values of the unbalanced FLPR as follows:

$$cl_{ik}^h = 1 - \varepsilon p_{ik}^h. \quad (13)$$

Easily, we can define the consistency measures for particular alternatives and for the whole unbalanced FLPR.

Definition 2.4. The consistency measure, $cl_i^h \in [0,1]$, associated to a particular alternative x_i of an unbalanced FLPR P^h is defined as:

$$cl_i^h = \frac{\sum_{k=1;i\neq k}^{n}(cl_{ik}^h + cl_{ki}^h)}{2(n-1)}. \quad (14)$$

Definition 2.5. The consistency level, $cl^h \in [0,1]$, of an unbalanced FLPR P^h is defined as follows:

$$cl^h = \frac{\sum_{i=1}^{n} cl_i^h}{n}. \quad (15)$$

When working with an incomplete unbalanced FLPR, (3) cannot be used to obtain the estimate of a known preference value. In this case, the following sets can be defined [10]:

$$\begin{array}{l} A = \{(i,j) \mid i,j \in \{1,\ldots,n\} \wedge i \neq j\} \\ MV^h = \{(i,j) \in A \mid p_{ij}^h \text{ is unknown}\} \\ EV^h = A \setminus MV^h \\ H_{ik}^{h1} = \{j \neq i,k \mid (i,j),(j,k) \in EV^h\} \\ H_{ik}^{h2} = \{j \neq i,k \mid (j,i),(j,k) \in EV^h\} \\ H_{ik}^{h3} = \{j \neq i,k \mid (i,j),(k,j) \in EV^h\} \\ EV_i^h = \{(a,b) \mid (a,b) \in EV^h \wedge (a = i \vee b = i)\}, \end{array} \quad (16)$$

Then, the estimated value of a particular preference degree p_{ik}^h $((i,k) \in EV^h)$ can be calculated as shown in (4) assuming $(\#H_{ik}^{h1} + \#H_{ik}^{h2} + \#H_{ik}^{h3}) \neq 0$.

ISBN: 978-989-95079-6-8

function estimate_p(h,i,k)

1) $(cp_{ik}^h)^1 = (s_0, 0),\ (cp_{ik}^h)^2 = (s_0, 0),\ (cp_{ik}^h)^3 = (s_0, 0)$

2) if $\#H_{ik}^{h1} \neq 0$, then $(cp_{ik}^h)^1 = TF_t^{t'}(\Delta_{t'}(\sum_{j \in H_{ik}^{h1}} \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j1}))))$

3) if $\#H_{ik}^{h2} \neq 0$, then $(cp_{ik}^h)^2 = TF_t^{t'}(\Delta_{t'}(\sum_{j \in H_{ik}^{h2}} \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j2}))))$

4) if $\#H_{ik}^{h3} \neq 0$, then $(cp_{ik}^h)^3 = TF_t^{t'}(\Delta_{t'}(\sum_{j \in H_{ik}^{h3}} \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j3}))))$

5) Calculate $cp_{ik}^h = TF_t^{t'}(\Delta_{t'}\left(\frac{\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^1)) + \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^2)) + \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^3))}{(\#H_{ik}^{h1} + \#H_{ik}^{h2} + \#H_{ik}^{h3})}\right))$

end function

2.4 Estimation Procedure of Missing Values

The procedure estimates missing information in an expert's incomplete unbalanced FLPR using only the preference values provided by that particular expert. It is designed using (4) and estimates missing information values by means of two different tasks:

1. **Elements to be estimated in step *t* of the procedure:**

$$EMV_t^h = \{(i,k) \in MV^h \setminus \bigcup_{l=0}^{t-1} EMV_l^h \mid i \neq k \wedge \exists j \in \{H_{ik}^{h1} \cup H_{ik}^{h2} \cup H_{ik}^{h3}\}\}, \tag{17}$$

   and $EMV_0^h = \emptyset$ (by definition).

2. **Estimation of a particular missing value:** In order to estimate a particular value, the function $estimate_p(h,i,k)$ at the top of the page is used.

3 Consensus model

In this section, we present a consensus model for GDM problems where experts provide their preferences using incomplete unbalanced FLPRs. To solve GDM problems with this kind of preference relations, firstly, it is necessary to deal with the missing values [10]. The previous consistency based procedure of missing values allows us to measure the consistency levels of each expert. This consistency information is used in this section to propose a consensus model based not only on consensus criteria but also on consistency criteria. We consider that both criteria are important to guide the consensus process in an incomplete decision framework. In such a way, we get that experts change their opinions toward agreement positions in a consistent way, which is desirable to achieve a consistent and consensus solution. The main characteristics of the proposed consensus model are the following:

- It is designed to guide the consensus process of incomplete unbalanced fuzzy linguistic GDM problems.
- It uses a consistency based procedure to calculate the incomplete unbalanced fuzzy linguistic information.
- It is based both consensus criteria and consistency criteria. The proposed consensus model is designed with the aim of obtaining the maximum possible consensus level while trying to achieve a high level of consistency in experts' preferences.
- A feedback mechanism is defined using the above criteria. It substitutes the moderator's actions, avoiding the possible subjectivity that he/she can introduce, and gives advice to the experts to find out the changes they need to make in their opinions to obtain a solution with certain consensus and consistency degrees simultaneously.

In particular, the consensus model develops its activity in five phases that will be described in further detail in the following subsections: 1) computing missing information, 2) computing consistency measures, 3) computing consensus measures, 4) controlling the consistency/consensus state, and 5) feedback mechanism.

3.1 Computing Missing Information

In this first step, for each incomplete unbalanced FLPR P^h, we obtain its corresponding complete unbalanced FLPR $\bar{P}^h$ using the estimation procedure described in Section 2.4.

3.2 Computing Consistency Measures

To compute consistency measures, first, for each $\bar{P}^h$, we compute its corresponding unbalanced FLPR $CP^h = (cp_{ik}^h)$ according to expression (3). Second, we apply (13)-(15) to $(\bar{P}^h, CP^h)$ $(\forall h)$ to compute the consistency measures $CL^h = (cl_{ik}^h)$, cl_i^h, cl^h, $\forall i,k \in \{1,\ldots,n\}$. Finally, we define a global consistency measure among all experts to control the global consistency situation as follows:

$$CL = \frac{\sum_{h=1}^m cl^h}{m}. \tag{18}$$

3.3 Computing Consensus Measures

As in [6, 10], we compute two different kinds of measures: consensus degrees and proximity measures. Consensus degrees are used to measure the actual level of consensus in the process, while the proximity measures give information about how close to the collective solution every expert is. These measures are given on three different levels for a preference relation: pairs of alternatives, alternatives and relation. It will allow us to find out the consensus state of the process at different levels. For example, we will be able to identify which experts are close to the consensus solution, or in which alternatives the experts are having more trouble to reach consensus.

3.3.1 Consensus Degrees

For each pair of experts (e_h, e_l) $(h = 1,\ldots,m-1,\ l = h+1,\ldots,m)$, a similarity matrix, $SM^{hl} = (sm_{ik}^{hl})$, is defined, where

$$sm_{ik}^{hl} = 1 - \frac{|\Delta_{t'}^{-1}(TF_{t'}^t(\bar{p}_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(\bar{p}_{ik}^l))|}{n(t') - 1}, \tag{19}$$

being $\bar{p}_{ik}^h = (s_v^{n(t)}, \alpha_1),\ t \in \{t^-, t_2^-, t^+, t_2^+\},\ \bar{p}_{ik}^l = (s_w^{n(t)}, \alpha_2), t \in \{t^-, t_2^-, t^+, t_2^+\}$, and $t' \in \{t^-, t_2^-, t^+, t_2^+\}$.

ISBN: 978-989-95079-6-8

Then, a consensus matrix, $CM = (cm_{ik})$, is calculated by aggregating all the similarity matrices using the arithmetic mean as the aggregation function ϕ:

$$cm_{ik} = \phi(sm_{ik}^{hl},\ h = 1, \ldots, m-1,\ l = h+1, \ldots, m). \quad (20)$$

Once the consensus matrix, CM, is computed, we proceed to calculate the consensus degrees at the three different levels:

1. **Level 1.** *Consensus degree on pairs of alternatives,* cop_{ik}. It measures the consensus degree amongst all the experts on the pair of alternatives (x_i, x_k).

$$cop_{ik} = cm_{ik};\ \forall\, i, k = 1, \ldots, n \wedge i \neq k. \quad (21)$$

2. **Level 2.** *Consensus degree on alternatives,* ca_i. It measures the consensus degree amongst all the experts on the alternative x_i.

$$ca_i = \frac{\sum_{k=1;k\neq i}^{n} (cop_{ik} + cop_{ki})}{2(n-1)}. \quad (22)$$

3. **Level 3.** *Consensus degree on the relation,* cr. It measures the global consensus degree amongst all the experts and is used to control the consensus situation.

$$cr = \frac{\sum_{l=1}^{n} ca_i}{n}. \quad (23)$$

3.3.2 Proximity Measures

These measures evaluate the agreement between the individual experts' opinions and the group opinion. To compute them for each expert, we need to obtain the collective unbalanced FLPR, $P^c = (p_{ik}^c)$, which summarizes preferences given by all the experts and is calculated by means of the aggregation of the set of individual unbalanced FLPRs $\{\bar{P^1}, \ldots, \bar{P^m}\}$. In this way, to obtain P^c we use the unbalanced fuzzy linguistic version of an IOWA operator [11, 12], which uses both consensus and consistency criteria as inducing variable. Thus, we obtain each collective unbalanced fuzzy linguistic preference degree p_{ik}^c according to the most consistent and consensual individual unbalanced fuzzy linguistic preference degrees.

Thus, to obtain each p_{ik}^c according to the most consistent and consensual individual unbalanced fuzzy linguistic preference degrees, we propose to use an unbalanced fuzzy linguistic IOWA operator with the consistency/consensus values, $\{z_{ik}^1, z_{ik}^2, \ldots, z_{ik}^m\}$, as the values of the order inducing variable, i.e.,

$$\begin{aligned} p_{ik}^c = \Phi_W(\langle z_{ik}^1, \bar{p}_{ik}^1\rangle, \ldots, \langle z_{ik}^m, \bar{p}_{ik}^m\rangle) = \\ TF_t^{t'}(\Delta_{t'}(\textstyle\sum_{h=1}^{m} w_h \cdot \Delta_{t'}^{-1}(TF_{t'}^t(\bar{p}_{ik}^{\sigma(h)})))), \end{aligned} \quad (24)$$

where

- σ is a permutation of $\{1, \ldots, m\}$ such that $z_{ik}^{\sigma(h)} \geq z_{ik}^{\sigma(h+1)}$, $\forall h = 1, \ldots, m-1$, i.e., $\langle z_{ik}^{\sigma(h)}, \bar{p}_{ik}^{\sigma(h)}\rangle$ is the 2-tuple with $z_{ik}^{\sigma(h)}$ the h-th highest value in the set $\{z_{ik}^1 \ldots, z_{ik}^m\}$;
- the weighting vector is computed according to the following expression:

$$w_h = Q\left(\frac{\sum_{j=1}^{h} z_{ik}^{\sigma(j)}}{T}\right) - Q\left(\frac{\sum_{j=1}^{h-1} z_{ik}^{\sigma(j)}}{T}\right), \quad (25)$$

  with $T = \sum_{j=1}^{m} z_{ik}^j$;
- and the set of values of the inducing variable $\{z_{ik}^1, \ldots, z_{ik}^m\}$ are computed as follows:

$$z_{ik}^h = (1-\delta)\cdot cl_{ik}^h + \delta \cdot co_{ik}^h, \quad (26)$$

  being co_{ik}^h the consensus measure for the preference value $\bar{p}_{ik}^h$ and $\delta \in [0,1]$ a parameter to control the weight of both consistency and consensus criteria in the inducing variable. Usually $\delta > 0.5$ will be used to give more importance to the consensus criterion. We should note that in our framework, each value co_{ik}^h used to calculate $\{z_{ik}^1, \ldots, z_{ik}^m\}$ is defined as follows:

$$co_{ik}^h = \frac{\sum_{l=h+1}^{n} sm_{ik}^{hl} + \sum_{l=1}^{h-1} sm_{ik}^{lh}}{n-1}. \quad (27)$$

Once we have computed P^c, we can compute the proximity measures in each level of an unbalanced FLPR.

1. **Level 1.** *Proximity measure on pairs of alternatives,* pp_{ik}^h. The proximity measure of an expert e_h on a pair of alternatives (x_i, x_k) to the group's one is calculated as:

$$pp_{ik}^h = 1 - \frac{\left|\Delta_{t'}^{-1}(TF_{t'}^t(\bar{p}_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^c))\right|}{n(t')-1}. \quad (28)$$

   being $\bar{p}_{ik}^h = (s_v^{n(t)}, \alpha_1)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, $p_{ik}^c = (s_w^{n(t)}, \alpha_2)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, and $t' \in \{t^-, t_2^-, t^+, t_2^+\}$.

2. **Level 2.** *Proximity measure on alternatives,* pa_i^h. The proximity measure of an expert e_h on an alternative x_i to the group's one is calculated as follows:

$$pa_i^h = \frac{\sum_{k=1;k\neq i}^{n} (pp_{ik}^h + pp_{ki}^h)}{2(n-1)}. \quad (29)$$

3. **Level 3.** *Proximity measure on the relation,* pr^h. The proximity measure of an expert e_h on his/her unbalanced FLPR to the group's one is calculated as follows:

$$pr^h = \frac{\sum_{l=1}^{n} pa_i^h}{n}. \quad (30)$$

3.4 Controlling Consistency/Consensus State

The consistency/consensus state control process will be used to decide when the feedback mechanism should be applied to give advice to the experts or when the consensus reaching process has to come to an end. It should take into account both the consensus and consistency measures. To do that, we use a measure or level of satisfaction, called *consistenty/consensus level* (CCL) [10], which is used as a control parameter:

$$CCL = (1-\delta)\cdot CL + \delta \cdot cr, \quad (31)$$

with δ the same value used in [7]. When CCL satisfies a minimum threshold value $\gamma \in [0,1]$, the consensus reaching process finishes and the selection process can be applied. To avoid that the consensus process does not converge, a maximum number of consensus rounds is incorporated.

ISBN: 978-989-95079-6-8

3.5 Feedback Mechanism

The feedback mechanism generates personalized advice to the experts according to the consistency and consensus criteria. It helps experts to change their preferences and to complete their missing values. This activity is carried out in two steps:

1. **Identification of the preference values.** We must identify preference values that are contributing less to reach a high consensus/consistency state. To do that, we define set APS that contains 3-tuples (h, i, k) symbolizing preference degrees p_{ik}^h that should be changed because they affect badly to that consistency/consensus state.

   (a) *Identification of experts.* We identify the set of experts EXPCH that should receive advice on how to change some of their preference values.

   $$EXCPCH = \{h \mid (1-\delta) \cdot cl^h + \delta \cdot pr^h < \gamma\}. \quad (32)$$

   (b) *Identification of alternatives.* We identify the alternatives ALT that the above experts should consider to change.

   $$ALT = \{(h,i) \mid h \in EXPCH \wedge (1-\delta) \cdot cl_i^h + \delta \cdot pa_i^h < \gamma\}. \quad (33)$$

   (c) *Identification of pairs of alternatives.* Finally, we identify preference values for every alternative and expert $(x_i; e_h \mid (h,i) \in ALT)$ that should be changed according to their proximity and consistency measures on the pairs of alternatives, i.e.,

   $$APS = \{(h,i,k) \mid (h,i) \in ALT \wedge (1-\delta) \cdot cl_{ik}^h + \delta \cdot pp_{ik}^h < \gamma\}. \quad (34)$$

   Additionally, the feedback process must provide rules for missing preference values. To do so, it has to take into account in APS all missing values that were not provided by the experts, i.e.,

   $$APS' = APS \cup \{(h,i,k) \mid p_{ik}^h \in MV_h\}. \quad (35)$$

2. **Generation of advice.** In this step, the feedback mechanism generates personalized recommendations to help the experts to change their preferences. These recommendations are based on easy recommendation rules that will not only tell the experts which preference values they should change, but will also provide them with particular values for each preference to reach a higher consistency/consensus state.

   The new preference degree of alternatives x_i over alternative x_k to recommend to the expert e_h, rp_{ik}^h, is calculated as the following weighted average of the preference value cp_{ik}^h and the collective preference value p_{ik}^c:

   $$rp_{ik}^h = TF_t^{t'}(\Delta_{t'}((1-\delta) \cdot \Delta_{t'}^{-1}(TF_{t'}^t(cp_{ik}^h)) + \delta \cdot \Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^c)))). \quad (36)$$

   As previously mentioned, with $\delta > 0.5$, the consensus model leads the experts towards a consensus solution rather than towards an increase on their own consistency levels.

   Finally, we should distinguish two cases:

   (a) $\forall (h,i,k) \in APS'$, if $p_{ik}^h \in EV_h$, the recommendation generated for the expert e_h is: "You should change your preference value (i,k) to a value close to rp_{ik}^h".

   (b) $\forall (h,i,k) \in APS'$, if $p_{ik}^h \in MV_h$, the recommendation generated for the expert e_h is: "You should provide a value for (i,k) close to rp_{ik}^h".

4 Concluding Remarks

In this paper, we have proposed a model of consensus for GDM with incomplete unbalanced fuzzy linguistic information. It uses two different kinds of measures to guide the consensus reaching process, consistency and consensus measures, and generates advice to experts in a discriminate way. As a consequence, this model allows us to achieve consistent and consensus solutions. In addition, it supports the consensus process automatically, without moderator.

Acknowledgment

This paper has been developed with the financing of SAINFOWEB project (TIC00602), Feder Funds in FUZZYLING project (TIN2007-61079) and PETRI project (PET2007-0460).

References

[1] E. Herrera-Viedma et al. Some issues on consistency of fuzzy preference relations. *European Journal of Operational Research*, 154(1):98–109, 2004.

[2] Th.L. Saaty. *The Analytic Hierarchy Process*. MacGraw-Hill, New York, 1980.

[3] T. Tanino. Fuzzy preference orderings in group decision making. *Fuzzy Sets and Systems*, 12:117–131, 1984.

[4] F. Herrera et al. A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems*, 16(2):354–370, 2008.

[5] E. Herrera-Viedma and A.G. López-Herrera. A model of information retrieval system with unbalanced fuzzy linguistic information. *International Journal of Intelligent Systems*, 22(11):1197–1214, 2007.

[6] F. Herrera et al. A model of consensus in group decision making under linguistic assessments. *Fuzzy Sets and Systems*, 78(1):73–87, 1996.

[7] J. Kacprzyk et al. Group decision making and consensus under fuzzy preferences and fuzzy majority. *Fuzzy Sets and Systems*, 49(1):21–31, 1992.

[8] F. Herrera and L. Martínez. A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 31(2):227–234, 2001.

[9] F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.

[10] E. Herrera-Viedma et al. A consensus model for group decision making with incomplete fuzzy preference relations. *IEEE Transactions on Fuzzy Systems*, 15(5):863–877, 2007.

[11] R.R. Yager. Induced aggregation operators. *Fuzzy Sets and Systems*, 137(1):59–69, 2003.

[12] R.R. Yager and D.P. Filev. Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(2):141–150, 1999.

ISBN: 978-989-95079-6-8

Real-time segmentation of moving objects in H.264 compressed domain with dynamic design of fuzzy sets

C. Solana-Cipres, L. Rodriguez-Benitez[1] J. Moreno-Garcia[2] L. Jimenez[1] G. Fernandez-Escribano[3]

1.Escuela Superior de Informatica, Universidad de Castilla-La Mancha
Ciudad Real, Spain
2.Escuela Universitaria de Ingenieria Tecnica Industrial, Universidad de Castilla-La Mancha
Toledo, Spain
3.Instituto de Investigacion en Informatica, Universidad de Castilla-La Mancha
Albacete, Spain
Email: {cayetanoj.solana, luis.rodriguez, luis.jimenez, juan.moreno, gerardo.fernandez}@uclm.es

Abstract— *This paper presents a real-time segmentation algorithm to obtain moving objects from the H.264 compressed domain. The proposed segmentation works with very little information and is based on two features of the H.264 compressed video: motion vectors associated to the macroblocks and decision modes. The algorithm uses fuzzy logic and allows to describe position, velocity and size of the detected regions in a comprehensive way, so the proposed approach works with low level information but manages highly comprehensive linguistic concepts. The performance of the algorithm is improved using dynamic design of fuzzy sets that avoids merge and split problems. Experimental results for several traffic scenes demonstrate the real-time performance and the encouraging results in diverse situations.*

Keywords— Moving object detection, image segmentation, H.264 advanced video coding, dynamic fuzzy sets.

1 Introduction

In Computer Vision, moving object segmentation is referred to the process of identifying and partitioning of the meaningful regions present in video sequences. Classical techniques like *background subtraction* or *temporal differencing* use low level information about each pixel. These algorithms have a good performance and can identify with accuracy the border of the regions, but the most of them can not work in real-time because they have to decode and process each pixel of each frame. So fast algorithms to segment moving objects usually work directly on compressed video. The proposed algorithm uses only the information related to the motion and works with groups of pixels called *macroblocks*.

The proposed segmentation algorithm is designed for the H.264 advanced video coding standard. This standard has a better compression ratio than the MPEG compressed video family standards and provides the mechanisms for coding video that are optimized for compression efficiency and aims to meet the needs of practical multimedia communication applications. This new standard has more choices of coding parameters and strategies than MPEG-4. This work is based on previous algorithms for MPEG [1, 2], but the new proposal is faster and can fulfill the requeriment of real-time applications. Another contribution of this work is the improvement of the algorithm's robustness. This improvement can be obtained because the proposed technique can adapt itself to the scenario and the objects present in it in a dynamic way. More concretely, the fuzzy sets describing the linguistic variables are updated dynamically and this fact allows to reduce the merge and split ratios.

The paper is organized as follows. Section 2 describes the approaches to the H.264 segmentation algorithms. In the next section, some features of the H.264 video standard are presented. Fuzzy linguistic concepts are shown in Section 4. Later, in Section 5 it is analyzed the proposed method of moving object segmentation. The experimental results in several video traffic scenes are presented in Section 6. Finally, conclusions and future works are described in Section 7.

2 Related Work

The main problem related to the compressed video processing is the shortage of information and the uncertainty that it implies. In this segmentation method luminance and chrominance values are not used, but information related to the motion compensation between frames is needed. This fact implies the increase of the uncertainty related to the input information. In this noisy environment, fuzzy logic presents itself as the adequate theoretical medium to reduce the negative influence of the uncertainty.

Fuzzy logic has been applied in many tasks in computer vision [3, 4] because of its potential and its link with the natural language. It allows to translate video information in linguistic representations capable to be processed through fuzzy techniques. Approximate reasoning is used in this work to carry out a segmentation algorithm of moving objects that obtains conceptual representations which describe position, velocity and size of the regions detected in a comprehensive way.

The segmentation algorithms works directly over H.264 compressed domain video and exploit two features of macroblock: motion vectors and DCT coefficients. Zeng *et al.* [5] present an algorithm that employs a block-based Markov Random Field (MRF) model to segment moving objects from the sparse motion vector field obtained directly from the H.264 bitstream. A later approximation has been proposed in Liu *et al.* [6], where a real-time spatiotemporal segmentation is presented. In this case, spatial segmentation only exploits the motion vector field extracted from the H.264 compressed video. Regions are classified using the block residuals of global motion compensation and the projection is exploited for interframe tracking of video objects.

ISBN: 978-989-95079-6-8

3 Features of the H.264 Advanced Video Coding

H.264 [7], also known as MPEG-4 Part 10, is a standard for video compression developed jointly between the *Motion Picture Expert Group* (MPEG) and the *Video Coding Experts Group* (VCEG). Richardson explains deeply in [8] the features of the H.264 compressed domain, but in this section only the H.264 concepts used into the segmentation algorithm are described. This standard provides mechanisms for video coding that are optimized for a better compression efficiency and aims to meet the multimedia communication applications.

The pictures in a video sequence may be encoded in several ways. The goal of the different kinds of frames is to exploit the diverse class of redundant information that it is possible to find in a video sequence. Three of the most common kinds of pictures available in H.264, as well as in other multimedia standards are **Intra pictures** (*I Type*), **Predicted pictures** (*P Type*) and **Bi-directional or Interpolated pictures** (*B Type*).

The most relevant concept to this work is the **macroblock** (MB), which is the basic unit in which an image is divided into. It contains the information of a 16x16 pixels region and there are two types depending on the encoding: **Intra macroblock**, in which Intra-prediction algorithms are applied directly to exploit the spatial redundancy according to the H.264 standard, and **Inter macroblock**, in which motion compensation is used to exploit temporal redundancy from a reference macroblock (earlier, later or a combination of both).

H.264 uses *block-based motion compensation*, the same principle adopted by every major coding standard since H.261. This motion compensation is done through the redundant information between consecutive frames looking for a pattern that captures the kind of movement between pictures. This pattern is represented as a **motion vector**, which defines a distance and a direction and has two dimensions: *right_x* and *down_x*. Important differences from earlier standards include the support for a range of block sizes and the use of multiple reference frames to improve the performance of the coding.

H.264 supports motion compensation block sizes ranging from 16x16 to 4x4 samples. Each macroblock may be split up in 4 ways: 16x16, 16x8, 8x16 or 8x8. Each of the subdivided regions is a **macroblock partition**. If the 8x8 mode is chosen, each of the four 8x8 macroblock partitions within the macroblock may be further split in 4 ways: 8x8, 8x4, 4x8 or 4x4 (known as **sub-macroblock partitions**). These partitions and sub-partitions give rise to a large number of possible combinations within each macroblock. This method of partitioning macroblocks into motion compensated sub-blocks of varying size is known as *tree structured motion compensation*.

Since each macroblock and sub-macroblock partition has a motion vector associated, a macroblock has from 0 to 16 pairs of *motion vectors*. For example, an Intra macroblock has not any motion vector, an Inter macroblock partitioned into two 16x8 blocks has two pairs of motion vectors and an Inter macroblock partitioned into four 8x8 blocks -each of them partitioned into 4x4 sub-macroblocks- has 16 pairs of motion vectors ($4 * 4 = 16$).

Since H.264 allows block-based motion compensation, the *decision mode* and the *block size* give additional information about the motion degree on a specific region of a frame (Fig. 1). Thus the more partitions into a macroblock, the more implicit motion ratio over that region. The H.264 standard supports the **decision modes** illustrated in Table 1, where can also be seen the average frequency. This information is relevant to the segmentation algorithm.

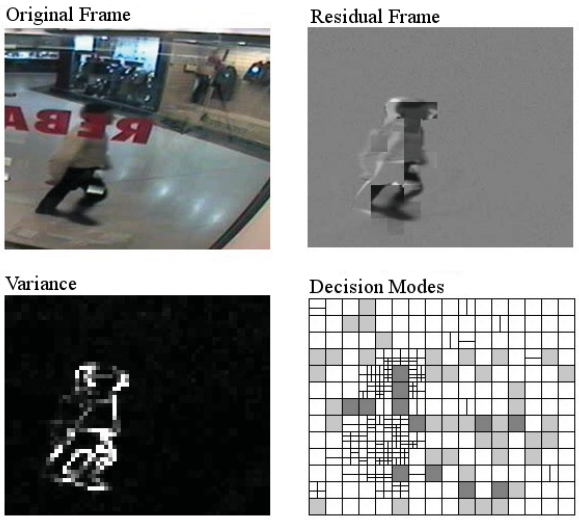


Figure 1: Processing a frame.

Table 1: Adopted nomenclature for decision modes.

Id	Macroblock type	Frequency
0	Skip, without added information	19,6 %
1	One 16x16 partition	50,6 %
2	Two 16x8 partitions	3,3 %
3	Two 8x16 partitions	2,2 %
4	Four 8x8 partitions	10,7 %
5	Two 8x4 sub-partitions in an 8x8 partition	1,9 %
6	Two 4x8 sub-partitions in an 8x8 partition	1,5 %
7	Four 4x4 sub-partitions in an 8x8 partition	0,5 %
8	Four 8x8 partitions without sub-partitions	0,4 %
9	4x4 Intra-frame macroblock	4,8 %
10	16x16 Intra-frame macroblock	0,9 %
11	Without using in an standard encoding	0,0 %
12	Without using in an standard encoding	0,0 %
13	8x8 Intra-frame macroblock	3,7 %

4 Fuzzy Linguistic Concepts

The motion vectors of an H.264 video flow are imprecise approximations to the real *Optical Flow Fields* [9]. In H.264 frames, the areas more sensitive to contain wrong motion vectors [10] are those with imperceptible variation of luminance values and those corresponding to the boundary objects. In cases like that, it is complex to the motion compensation algorithms to find correspondences between macroblocks in consecutive frames. The most compressed-domain techniques for image manipulation apply any filter to the motion vectors field to avoid the anomalous values as consequence of the mistakes introduced by the encoders. The use of the filters implies the growth of the information used in the algorithm and this kind of operations use to be expensive from a computational point of view; thus, these procedures are inappropriate for real-time algorithms.

ISBN: 978-989-95079-6-8

The related uncertainty justifies the use of the fuzzy logic in the H.264 compressed domain. The algorithm uses approximate reasoning to compact the information related to the video flow, so values not much affected by the encoding error will have similar fuzzy values. Besides, given that the goal of this work is to obtain the linguistic description of the moving regions present in a video scene, it is necessary to define the set of linguistic elements used in the description of the regions.

The proposed algorithm uses four **linguistic variables** [11] to represent the horizontal position (HP), vertical position (VP), horizontal velocity (HV) and vertical velocity (VV). One of the critical points of the system is the definition of the membership function for each variable because it depends on the scenario under observation. For example, Fig. 2 illustrates the linguistic labels belonging to the linguistic variable HP used as initial values in this study. These values can be adapted to obtain the optimal design for the specific scenario.

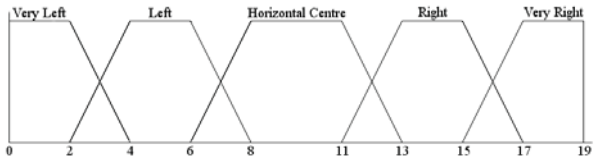


Figure 2: *Horizontal Position* linguistic labels.

A *linguistic motion vector* is the fuzzy representation of a motion vector belonging to a macroblock. It represents a linguistic description of the motion of a macroblock between consecutive frames and is defined as:

$$LMV =< FN, I_{HP}(mb_row), I_{VP}(mb_col), \\ , I_{HV}(right_x), I_{VV}(down_x) > \quad (1)$$

where the first element identifies the frame number where the motion vector is and the next four elements are four linguistic intervals. A **linguistic interval** is an ordered set of consecutive pairs of linguistic labels ($\{Left : 0,5 ; Horizontal\ Centre : 0,5\}$). The I_{PH} and I_{PV} intervals represent the column and row of the macroblock. The I_{HV} e I_{VV} intervals are obtained from the *right_x* and *down_x* values of the motion vector.

A **valid linguistic motion vector** is a linguistic motion vector that contains relevant information about the direction and velocity of an object, i.e., at least one of the two magnitudes concerning the velocity (I_{HV} and I_{VV}) of the vector is distinct of the label *No Motion*.

A *linguistic blob* is an 7-tuple composed by one or more conceptually similar motion vectors that represents in a linguistic way a region (*blob*) in a frame:

$$LB =< FN, Size, MBs, I_{HP}, I_{VP}, I_{HV}, I_{VV} > \quad (2)$$

where FN is the frame number in which is located the blob, $Size$ is the number of linguistic motion vectors grouped in the blob, MBs is a list of macroblocks belongs to the blob and the last four elements are the linguistic intervals that represent the position and velocity of the blob.

5 Segmentation Algorithm

In this work, the real displacement between frames is computed by using a fuzzification procedure working on motion vectors, neither DCT coefficients information nor statistical filtering is needed; so the algorithm requires very little of data; from a frame size of 320x240 pixels and about 35-40 KB, the algorithm only needs about 2-4 KB and around 18 KB per frame for a 640x480 video sequence with a frame size of 120-140 KB.

Fig. 3 shows the overview of the proposed system. The H.264 compressed video streams are the input of the algorithm. The first stage is the decoding of the video streaming with the goal of extracting the motion vectors and the decision modes. In the next stage, a k-neighbor algorithm is applied to the mode decision matrix to obtain the motion vectors to take into account. After that, in the fuzzification stage, motion vectors are converted into linguistic motion vectors in order to obtain the fuzzy representation of them. Then the noisy information is removed and the valid linguistic motion vectors are obtained. The algorithm rules out the motion vectors with low values of *right_x* and *down_x* because they could be noisy and they do not provide information about moving objects. Next stage is the segmentation, i.e., the valid vectors are grouped into linguistic blobs, each of them could be identified as a moving object of the video scene. Finally, the linguistic blobs are filtered to delete noisy ones. Note that the result of the algorithm is not a set of objects detected in a video sequence, but a set of linguistic blobs or regions which form the previous step to the identification of the objects. In a parallel process to the segmentation, the fuzzy sets defined over each linguistic variable are modified in a dynamic way to adapt the algorithm to a specific scenario, as it is explained in Section 5.3. Thus, the linguistic labels are modified in accordance with the size of the blobs identified in the scene and this fact has influence on the performance of the algorithm.

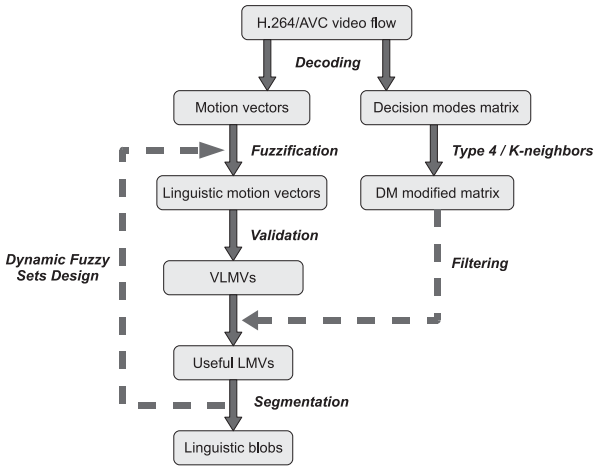


Figure 3: Overview of the proposed system.

5.1 Decision Modes Selection

The segmentation algorithm has to choose the valid motion vectors to obtain the right blobs. It can be selected between two ways of decision modes selection. In the first one, the algorithm takes into account only motion vectors belonging to a **macroblock of type 4** (four 8x8 partitions) because it has

ISBN: 978-989-95079-6-8

been observed experimentally that the moving objects are usually coded with this decision mode (Fig. 4(a)). In the second one, it is obtained the decision modes matrix corresponding to a frame and it is applied a **k-neighbor algorithm** to obtain the implicit motion degree in a region of the frame. The decision modes has been named in increasing motion degree, thus the higher is the k-neighbor macroblock value, the more is the movement inside the macroblock (Fig. 4(b)). In this case, the segmentation procedure works only with motion vector values greater than a predefined threshold U_{kn} in the k-neighbor algorithm. In Section 6 the results using both decision modes selection are analyzed.

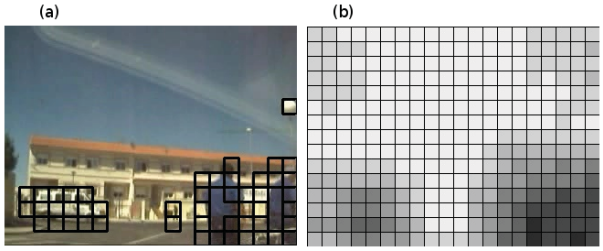


Figure 4: Decision modes selection: (a) type 4 macroblocks in a frame, (b) decision modes matrix with k-neighbor algorithm.

5.2 Total distance and weighted aggregation

After fuzzification of motion vectors, the proposed system clasifies the valid linguistic motion vectors into linguistic blobs through a *clustering* algorithm. If there is a linguistic motion vector suitable to belong to a blob, it is included into the linguistic blob and is modified the representation of this one according to a **weighted aggregation**. The similarity between a vector and a blob is calculated with a distance measure based in the numbering order of the defined linguistic labels over a variable. The distance between two linguistic intervals, the **total distance** between a linguistic motion vector and a blob and the weighted aggregation are explained deeply in [1, 2].

5.3 Dynamic Design

The segmentation algorithm proposed presents the typical problems of any segmentation system: **merge** (two or more objects are identified as one; it could happen with little objects) and **split** (one object is identified as two or more; it could happen with big objects). One of the reasons of these problems come from the static design of the linguistic variables used in the algorithm, i.e., their values remain unaltered for different kinds of objects. Since the definition of the total distance and the weighted aggregation, this segmentation system has a rigid behaviour and cannot adapt itself to the nature of the objects. Besides, the initial version of the algorithm is strongly dependent of the expert knowledge because the fuzzy sets defined over each linguistic variable are manually established and their values are invariables. Since the application is designed for real traffic scenarios, the linguistic variables are initially configured for the most frequently appeared object in that domain, i.e., the car. However, there could be a merge problem if an object little than a car (for example, a person) appears and there could be an split problem if a big object like a truck appears on the video. Then, if the number and the values of the fuzzy sets are selected in a wrong way, the result of the algorithm will be poor.

For this reason, this algorithm introduces the dynamic design of the linguistic labels to adapt them to each specific scenario, where the *dynamic design of a linguistic variable* X is understood as the adaptation and modification of the values of the fuzzy sets belonging to X during the execution of a fuzzy procedure. Thus, the segmentation procedure allows to give itself some feedback to improve its results. The algorithm has an initial configuration, but it could change the values of the fuzzy sets in function of the obtained results taking into consideration the size and shape of the obtained blobs in previous frames. So the automatic changes allow to improve the quality of the algorithm because they have positive influence on the merge and split rates, as shown in Fig 5.

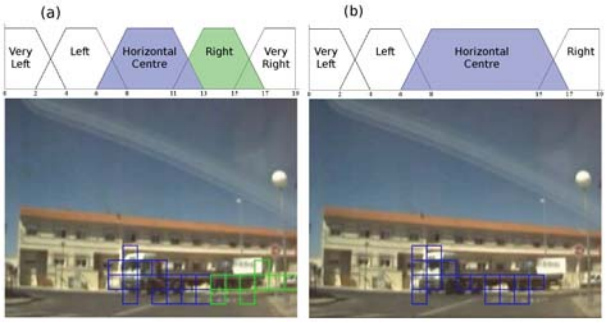


Figure 5: Dynamic fuzzy sets avoid the merge problem.

The dynamic design of the fuzzy sets is based on some common characteristics of the fuzzy systems like the **support set**, the **height** of a fuzzy set and the **alpha cut threshold** [12]. Being X a linguistic variable, A a fuzzy set of X, D the domain of X and α the threshold of the alpha cut set:

$$support(A) = \{x \in D : \mu_A(x) > 0\} \quad (3)$$

$$height(A) = max(\mu_A(x)) \quad (4)$$

$$alpha_cut_\alpha(A) = \{x \in D : \mu_A(x) \geq \alpha\} \quad (5)$$

The linguistic variables whose values are updated dynamically are HP and VP. Each variable X has a set of fuzzy sets $\{A_1, A_2, ..., A_{size}\}$ over his domain D_{HP} or D_{VP}. Only normal fuzzy sets are used, so the height of them is always equal to one. The algorithm is ready to work with both triangular and trapezoidal fuzzy sets. The support of each fuzzy set A is $support(A) = alpha_cut_1(A) + 4$, except for the first and the last sets (A_1 and A_{size}) of a linguistic variable, where $support(A) = alpha_cut_1(A) + 2$.

The dynamic design of the fuzzy sets is executed at the end of the segmentation algorithm, as shown in Fig. 3. It is only processed to the biggest blob detected in the frame, if there is at least one blob detected in said frame. Being the macroblocks $MBs = \{MB_1, MB_2, ..., MB_m\}$ belonging to the biggest blob LB, it can be calculated the **rectangular size** of the blob by:

$$height(LB) = row(MB_m) - row(MB_1) + 1 \quad (6)$$

ISBN: 978-989-95079-6-8

$$width(LB) = max_i(col(MB_i)) - min_i(col(MB_i)) + 1 \quad (7)$$

where col is the column of the macroblock, max is he maximum of the values and min is the minimum value.

The system calculates the new **number** of the new fuzzy sets defined to the linguistic variables HP and VP in function of the size of the biggest blob in the frame. Being W and H the width and the height in macroblocks of each frame, the number of the redisegned linguistic labels can be calculated as $size(HP) = W/width(LB)$ and $size(VP) = H/height(LB)$. The **names** of the linguistic labels of HP and VP are got from two configuration files that have a correspondence between the number of labels and the names of them. After that, if trapezoidal fuzzy sets are used, it is necessary to calculate the support and the alpha-level set with threshold 1 to define the **shape** of the corresponding fuzzy sets. Since $size(HP)$ and $size(VP)$, it can be calculated the size of the alpha-level set with threshold value 1 with the same size to all the fuzzy sets for each linguistic variable. Being A_{HP} a fuzzy set of HP and A_{VP} a fuzzy set of VP, the size of the alpha cut set is calculated as:

$$size(alpha_cut_1(A_{HP})) = \frac{W - 1 - 2 * (size(HP) - 1)}{size(HP)} \quad (8)$$

$$size(alpha_cut_1(A_{VP})) = \frac{H - 1 - 2 * (size(VP) - 1)}{size(VP)} \quad (9)$$

For example, being the fuzzy sets of the linguistic variable HP shown in Fig. 2, if the algorithm finds a full region in a frame which a width of 6 macroblocks ($width(LB) = 6$), the module which updates the values of the linguistic variable HP will change its values. First, it calculates the number of the labels as $size(HP) = W/width(LB) = 20/7 \backsimeq 3$ and reads the names of them in the configuration file (LE, HC and RI). Then, the size of the alpha cut set and the support of each set are calculated as:

$$size(alpha_cut_1(A_{HP})) = \tfrac{W-1-2*(size(HP)-1)}{size(HP)} =$$

$$= \frac{20 - 1 - 2 * (3 - 1)}{3} = 5 \quad (10)$$

$$support(A_1) = alpha_cut_1(A) + 2 = 5 + 2 = 7 \quad (11)$$

$$support(A_2) = alpha_cut_1(A) + 4 = 5 + 4 = 9 \quad (12)$$

The module can redefine the values of the linguistic variable HP with this information and obtains the new values, which are shown in Fig. 6.



Figure 6: New *Horizontal Position* linguistic labels.

6 Experimental Results

In this work a supervised method to evaluate the segmentation quality is used. The segmentation approach has been evaluated on several video sequences compressed using the H.264 encoder of JM 14.0. The encoder configuration has been set with *baseline profile* and low complexity mode on RD Optimization. The video tested resolution varies from 320x240 to 640x480 pixels. The specific configuration parameters of the algorithm (U_{mv}, U_{kn} and the minimum blob size) change depending on the application scenario. Fig. 7 shows three snapshots of three different video sequences tested where the detected blobs are drawn over the frames.

As it has been explained in Section 5.1, the algorithm can be executed in two modes depending of the decision modes selection: only *type 4* macroblocks or *k-neighbor algorithm* with a threshold. Table 2 shows the differences between both modes in several experiments (320x240 pixel resolution) using the same configuration parameters. It can be seen that one of them is faster but its performance is worse. Thus the choice will depend on the application field.

Table 2: Decision modes selection.

Measurement	**Type 4**	**K-neighbor**
Detection possibility	73 %	81 %
Precision	78 %	89 %
Processing time	16,2 ms.	18,6 ms.
Segmentation size	148 bits/frame	103 bits/frame

The hardware used in the experiments is an Intel Pentium IV Core 2 Duo T7100 1.8 GHz with 2 GB RAM memory. Table 3 shows average processing time depending on the frame resolution. Since a 320x240 frame is processed in around 20 ms. and a 640x480 one in around 38 ms., it can be ensured that the segmentation works in real-time, i.e., at least 25 fps.

Table 3: Average processing time.

Frame resolution (pixels)	**Processing time**
320x240	20 ms.
640x240	26 ms.
640x480	38 ms.

Table 4 shows the improvement of the results using dynamic design of the fuzzy sets. Two measurements are used to describe the segmentation performance. The first measurement is *detection possibility* that is defined as the percentage of the real objects detected. The second one is *precision* that measures the percentage of the detected objects corresponding to real objects. Then, high detection possibility means less miss-segmentation and high precision means less false segmentation. Besides, two measurements are defined to analyze the *processing time* and the *segmentation size*, i.e., the temporal and spatial requirements of the algorithm.

7 Conclusions

This work presents a real-time segmentation approach of moving objects with the next features. First of all, the algorithm is designed for H.264/AVC compressed domain due to its excellent compression efficiency and its widespread field of multimedia applications. The algorithm can work in real-time and

ISBN: 978-989-95079-6-8

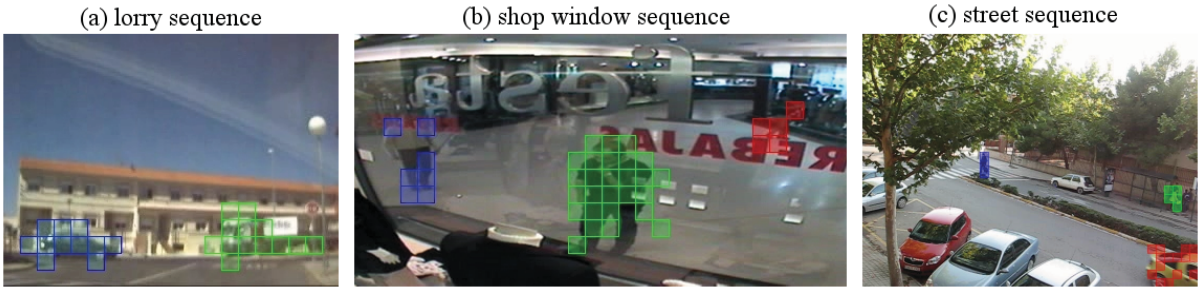


Figure 7: Snapshots of three video sequences with the detected blobs drawn.

Table 4: Improvement using dynamic design.

Measurement	Lorry sequence		Shop window sequence		Street sequence	
	Static	Dynamic	Static	Dynamic	Static	Dynamic
Resolution	320x240		640x240		640x480	
Ground truth	353	353	808	808	1355	1355
Detected blobs	378	362	649	684	1156	1138
True positives	293	302	631	665	996	1077
False positives	85	60	18	19	160	61
False negatives	60	51	177	143	359	278
Detection possibility	82 %	85 %	78 %	82 %	74 %	80 %
Precision	78 %	83 %	97 %	97 %	86 %	95 %
Processing time	17,4 ms.	19,7 ms.	25,5 ms.	32,7 ms.	37,2 ms.	41,1 ms.
Segmentation size	184 KB.	173 KB.	101 KB.	110 KB.	173 KB.	164 KB.

requires very little of data because it uses only the motion vectors field and the decision modes to carry out the segmentation. The proposed approach is based on fuzzy logic to detect the moving objects; fuzzy logic allows to avoid the noise inherent to the encoding process and to obtain conceptual representations that describe the regions detected in a comprehensive way. Finally, the segmentation performance is robust because it can adapt itself to the application scenario in a dynamic way updating in real-time the values of the fuzzy sets. This feature allows to improve the merge and split rates.

Acknowledgment

This work was supported by the Ministry of Science and Technology of Spain under CONSOLIDER Project CSD2006-46 and CICYT Project TIN2006-15516-C04-02 and by the Council of Science and Technology of Castilla-La Mancha under Projects PAI06-0106, PII2I09-0052-3440 and PII1C09-0137-6488.

References

[1] L. Rodriguez-Benitez, J. Moreno, J.J. Castro-Schez, J. Albusac, and L. Jimenez. An approximate reasoning technique for segmentation on compressed MPEG video. In *2nd International Conference on Computer Vision Theory and Applications, VISAPP'07*, pages 184–191, 2007.

[2] L. Rodriguez-Benitez, J. Moreno, J.J. Castro-Schez, J. Albusac, and L. Jimenez. Automatic objects behaviour recognition from compressed video domain. *Image and Vision Computing*, 2008. In Press, Accepted Manuscript, Available online 15 July 2008.

[3] P. Matsakis, J.M. Keller, L. Wendling, J. Marjamaa, and O.Sajhputera. Linguistic description of relative positions in images. *IEEE Transactions on systems, man and cybernetics*, 31(4):573–588, 2001.

[4] O. Sajhputera, P. Matsakis, J.M. Keller, and R. Bondougula. Linguistic descriptions for an object in motion. *Proceedings of the NAFIPS, New Orleans, Louisiana, USA*, pages 243–248, 2002.

[5] W. Zeng, J. Du, W. Gao, and Q. Huang. Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model. *Real-Time Imaging*, (11):290–299, 2005.

[6] Z. Liu, Y. Luy, and Z. Zhang. Real-time spatiotemporal segmentation of video objects in the H.264 compressed domain. *Journal of Visual Communication and Image Representation*, (18):275–290, 2007.

[7] ITU-T Recommendation H.264. Advanced video coding for generic audiovisual services, May 2003.

[8] I. E. G. Richardson. *H.264 and MPEG-4 Video Compression*. John Wiley & Sons Ltd, New Jersey, 2003.

[9] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[10] T. Yoshida, H. Katoh, and Y. Sakai. Block matching motion estimation using block integration based on reliability metric. *Proceedings of the International Conference on Image Processing (ICIP '97)*, 2:152–155, 1997.

[11] L.A. Zadeh. The concept of linguistic variable and its application to approximate reasoning. *Information Sciences*, 8:199–249, 1975.

[12] E. Cox. *The fuzzy systems handbook*. AP Professional, 2nd edition, 1998.

ISBN: 978-989-95079-6-8

On the Benefits of Representing Music Objects as Fuzzy Sets

Klaas Bosteels[1,2] Elias Pampalk[2] Etienne E. Kerre[1]

1. Fuzziness and Uncertainty Modelling Research Group
Dept. of Applied Mathematics and Computer Science, Ghent University, Ghent, Belgium
2. Last.fm Ltd., London, UK
Email: klaas.bosteels@ugent.be, elias@last.fm, etienne.kerre@ugent.be

***Abstract*— *Over the past few years, fuzzy set theory has become a well-established mathematical theory that has been applied in many practical applications. In this paper, we elaborate on the use of fuzzy sets in the context of music information retrieval. First, we demonstrate that music objects like artists and songs can straightforwardly be represented as fuzzy sets, and then we explain why it is beneficial to represent them in this way. In particular, we illustrate the potential of this fuzzy approach by describing how we relied on it to build the popular "Multi Tag Search" demonstration on* `http://playground.last.fm`.

***Keywords*—** Music information retrieval, fuzzy sets, fuzzy comparison measures.

1 Introduction

The way in which the objects are represented is an important aspect of most music information retrieval applications. Simple vectors are used in many cases, but certain more advanced representations like statistical distributions are popular too. In this paper, we advocate for the usage of fuzzy sets for this purpose. As we will explain in detail, representing a music object as a fuzzy set in some universe U, i.e., as a $U \to [0, 1]$ mapping that associates a degree of membership with each u from U, is a natural and mathematically sound approach that offers many possibilities and advantages. To enforce our claims, we will describe how we relied on fuzzy set theory for building the "Multi Tag Search" demonstration on `http://playground.last.fm`, the section of the CBS-owned music community website Last.fm[1] where new ideas and experimental technologies are showcased.

2 Representing music objects as fuzzy sets

There are three main ways to obtain descriptions for music objects:

1. Hiring music experts to manually annotate the objects.
2. Assembling a large community of music enthusiasts and then tapping into the wisdom of crowds.
3. Applying signal processing techniques to generate descriptions automatically.

Hence, song descriptions can come from experts, communities, or computers. In each case, it is often possible to convert them to fuzzy sets.

2.1 From experts

When music experts are hired to generate descriptions, they are usually asked to provide values for a predefined list of attributes. For instance, an expert could be used to assign values to the attributes "genre" and "complexity" for a particular song. Since the latter attribute is numerical, it can straightforwardly be converted to a gradual element of a fuzzy set. For example, if a song is 80% complex, then we would give "complexity" a membership degree of 0.8 in the fuzzy set that corresponds with this song. In case of a categorical attribute like "genre", however, we cannot interpret its values as membership degrees. Nevertheless, it is still possible to convert such attributes, namely, by regarding the values as binary attributes. For example, we can express that a song is an alternative rock song by assigning the membership degree 1 to "alternative rock" and setting the membership degrees of all other possible genres to 0. In fact, it might even be more appropriate to directly present the different genre values as separate numerical attributes to the expert, since a song often belongs to several genres to some extent. The expert can then express that, e.g., a particular song is mainly alternative rock, but also incorporates some notable blues and jazz influences. The well-known online radio station Pandora[2] uses 150-500 so called "genes" to describe songs [1, 2], where each gene is a numerical attribute that takes a value from the interval $[0, 5]$. Hence, the descriptions generated by Pandora's music experts could easily be interpreted as fuzzy sets.

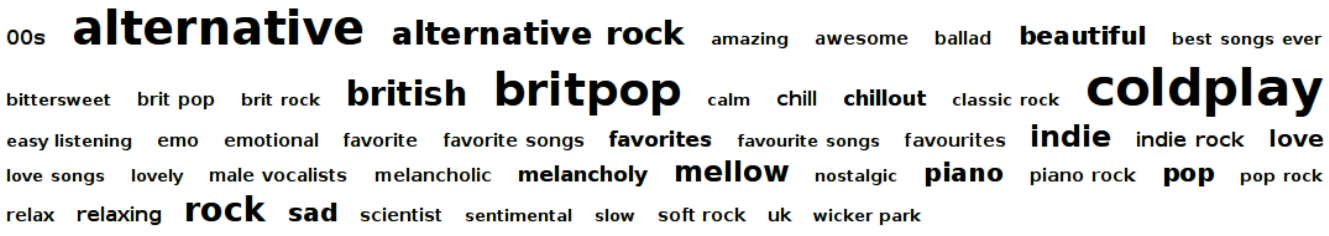
2.2 From communities

The data that powers Last.fm's recommendation engine is mainly generated by a technique called "scrobbling". When a user listens to a song, the name of this song is sent to Last.fm and added to his or her profile. By counting how many times each user listens to each song, implicit user ratings can be obtained from these "scrobbles". Since such ratings can easily be interpreted as membership degrees, they can be used to associate a fuzzy set of users with each song. For instance, the users that listened the most to the song in question can be given membership degree 1, and the membership degrees of the remaining users can be determined by dividing their play counts by the greatest play count, e.g., if there are only three users and their play counts for a particular song are 10, 5, and 1, then the membership degrees of these users in the fuzzy set for this song would be 1, 0.5, and 0.1, respectively. Comparing such fuzzy sets by means of the cosine similarity measure

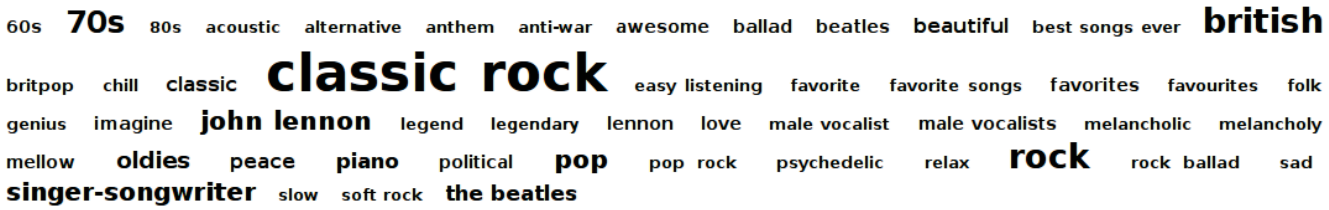
[1] http://www.last.fm

[2] http://www.pandora.com

ISBN: 978-989-95079-6-8



(a) Tag cloud for "The Scientist" by "Coldplay".



(b) Tag cloud for "Imagine" by "John Lennon".

Figure 1: Two examples of tag clouds.

is equivalent to the basic "item-to-item collaborative filtering" recommendation algorithm [3].

Last.fm also uses social tagging [4] as an additional community-driven way to generate descriptions. Social tags are individual keywords or phrases that people associated with objects, usually for organizational purposes. When the tags applied by a large community are combined, rich and complex descriptions emerge. A popular way to visualize such a description is by means of a tag cloud, i.e., an alphabetically ordered list of tags in which the font size is proportional to the number of people that associated the tag in question with the object. Fig. 1 shows two examples. By interpreting the font sizes in a tag cloud as membership degrees, we obtain a fuzzy set of tags that describes the corresponding object.

2.3 From computers

Using algorithms that analyze audio signals, song descriptions can be obtained in a fully automatic way [5]. However, these descriptions are usually very low-level and difficult to interpret. A possible solution for this problem is to teach computers how to add high-level annotations to songs by feeding the automatically generated low-level descriptions into machine learning algorithms that were trained on descriptions from experts or communities [6, 7], or by developing algorithms that propagate annotations by analysing the low-level descriptions [8]. Although it is certainly not impossible to convert the low-level descriptions to fuzzy sets [9], it makes more sense to apply such a conversion to the high-level ones obtained by incorporating machine learning techniques. Since these automatically generated high-level descriptions are intended to resemble the descriptions generated by humans, they can be converted to fuzzy sets as explained above.

3 The Benefits

In the 40 years since Zadeh introduced the concept of a fuzzy set [10], tens of thousands of papers and plenty of books have been written on this seminal topic, and more than 20 international journals were established in the corresponding field. The general advantage of representing songs as fuzzy sets is that we can rely on this vast amount of research for dealing with such representations. In this section, we describe a few more concrete benefits that ensue from this general one.

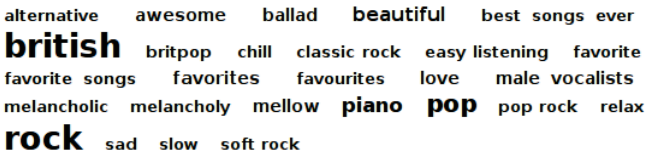


Figure 2: Intersection of the tag clouds shown in Fig. 1.

3.1 Fuzzy aggregation operators

Fuzzy aggregation operators are mathematical objects that aggregate multiple membership degrees into one. Formally, a fuzzy aggregation operator $\mathcal{H}$ of arity n, with $n \in \mathbb{N} \setminus \{0\}$, is an increasing $[0,1]^n \to [0,1]$ mapping that satisfies $\mathcal{H}(0,0,\ldots,0) = 0$ and $\mathcal{H}(1,1,\ldots,1) = 1$. Triangular norms [11], Quasi-arithmetic means [12], Sugeno and Choquet integrals [13, 14], and ordered weighted averaging (OWA) operators [15] are just a few examples of fuzzy aggregation operators that could be valuable for combining fuzzy descriptions of music objects. We refer to [16] for an introductionary overview of the extensively studied field of fuzzy aggregation operators. Fig. 2 provides a very simple example. It shows the result of applying the minimum operator to compute the intersection of the tag clouds in Fig. 1, which could be used as an intuitive visualization of the similarities between the songs that correspond with these tag clouds.

3.2 Fuzzy comparison measures

Expressions like, e.g.,

$$\frac{|A \cap B|}{|A \cup B|}, \qquad \frac{2\,|A \cap B|}{|A| + |B|}, \qquad \text{and} \qquad \frac{|A \cap B|}{\sqrt{|A|\,|B|}}$$

are commonly used to determine the similarity between two (ordinary) sets A and B. By defining the cardinality of a fuzzy set A in U as $|A| = \sum_{u \in U} A(u)$ and generalizing the classical set-theoretic operations to fuzzy sets, such expressions can

ISBN: 978-989-95079-6-8

be extended to similarity measures for fuzzy sets [17]. More generally, a measure that compares crisp sets can often be extended to fuzzy sets. We call such a measure a fuzzy comparison measure. A wide plethora of fuzzy comparison measures can be found in the literature, and the properties of many of these measures have been studied extensively. For instance, the measures M for which $1 - M$ satisfies the triangle inequality have been characterized for a broad family of fuzzy similarity measures [18], which can be very useful in practice since the triangle inequality allows to reduce the required number of comparisons [19]. Another family of fuzzy comparison measures that has practical value is the one developed in [20, 21, 22], since it can be used to systematically construct measures that satisfy certain properties. We rely on this family in the practical example described in Section 4.

3.3 Fuzzy relations

A (binary) fuzzy relation on U is a $U \times U \rightarrow [0, 1]$ mapping that associates a membership degree with each pair of elements from U, i.e., a fuzzy set in $U \times U$. Since fuzzy comparison measures are usually fuzzy relations on the (ordinary) set of all possible fuzzy sets, the approach described in the previous subsection has the additional advantage that the obtained similarities are also membership degrees, which implies that fuzzy aggregation operators can be used to combine them. For example, a fuzzy relation that models the similarities between the songs in the considered collection is a prerequisite of the general framework for defining dynamic playlist generation heuristics introduced in [23], because it relies on fuzzy set theory to make the definitions systematic, formal, and intuitively clear. Since music similarity measures based on a fuzzy similarity measure are fuzzy relations already, we do not have to apply any normalization procedures on them in order to be able to use this framework.

4 Practical example

In this section, we describe a practical example that illustrates the potential of representing songs as fuzzy sets. More precisely, we explain how we relied on fuzzy set theory to build the "Multi Tag Search", a web application that allows users to find music by supplying one or more tags. Fig. 3 shows the first results returned by this application when popular artists that match with the tags "british" and "rock" are requested. Any set of keywords or phrases can be used to search for either songs or artists, and the options "up-and-coming" and "free downloads" are also available, in addition to "popular".

4.1 Basic method

Putting it in one sentence, the "Multi Tag Search" ranks all relevant music objects from Last.fm's database by representing the query as a fuzzy set of tags and comparing this fuzzy set with the fuzzy tag set associated with each object. More formally, the ranking score for object i is given by:

$$score(\widetilde{Q}, \widetilde{O}_i) = w(i)\, M(\widetilde{Q}, \widetilde{O}_i) \tag{1}$$

with $\widetilde{Q}$ and $\widetilde{O}_i$ the fuzzy sets corresponding to the query and object i, respectively, M a fuzzy comparison measure, and w a function that associates a real number with each object index. The weighting function w allows us to push certain objects up



Figure 3: Screenshot of the "Multi Tag Search".

or down in the generated ranking. It is mainly used to implement the "popular", "up-and-coming", and "free downloads" search options, but we also utilize it for some small optimizations like penalizing objects to which only very few tags have been applied.

We used the following general definitions of the fuzzy tag sets $\widetilde{Q}$ and $\widetilde{O}_i$:

$$\widetilde{Q}(t) = \begin{cases} \dfrac{q(t)}{\max_{u \in T}\, q(u)} & \text{when } t \in Q \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\widetilde{O}_i(t) = \begin{cases} \dfrac{o_i(t)}{\max_{u \in T, j \in I}\, o_j(u)} & \text{when } t \in O_i \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

for all $t \in T$ and $i \in I$, with T the universe of all possible tags, I the set of all object indexes, Q the set of query tags, O_i the set consisting of all tags that have been applied to object i, and q and o_i, for each $i \in I$, a $T \rightarrow \mathbb{R}$ mapping. For the values of the parameters q and o_i in these definitions, we drew some inspiration from text retrieval, namely, we put $q = idf$ and $o_i = ntf_i$, with idf the so called "inverse document frequency", and ntf_i the "normalized term frequency" [24]:

$$idf(t) = \log\left(\frac{\text{total number of objects}}{\text{number of objects tagged with } t}\right) \tag{4}$$

$$ntf_i(t) = \frac{\text{number of taggings with } t \text{ for object } i}{\text{total number of taggings for object } i} \tag{5}$$

for each $t \in T$ and $i \in I$. Hence, tags that are not frequently used get a high membership degree in $\widetilde{Q}$, and the membership degree of the tags in $\widetilde{O}_i$ is proportional to the number of times they have been applied to object i. Many other term-weighting schemes can be found in the text retrieval literature [25], but using the inverse document frequency for the query and the normalized term frequency for the objects appeared to work very well for Last.fm's tag data.

ISBN: 978-989-95079-6-8

4.2 Constructing a suitable comparison measure

High-quality tag data is by far the most important requirement for a good tag-based search engine. Improving the quality of the underlying data is generally more worthwhile than trying to find a better comparison measure, and which comparison measure is used exacly becomes less important when the tag data is of high quality. "Worry about the data first before you worry about the algorithm," as Google's director of research Peter Norvig puts it [26]. Nevertheless, it is of course important to use a suitable comparison measure that has the right properties, but it makes sense to choose the simplest possible measure that satisfies these properties, especially when you have to deal with millions of objects and users.

For the "Multi Tag Search", we wanted the fuzzy comparison measure M to satisfy the following properties:

$$M(A,B) = 0 \Leftarrow \operatorname{supp} A \cap \operatorname{supp} B = \emptyset \qquad \text{(exclusive)}$$
$$M(A,B) = 0 \Rightarrow \operatorname{supp} A \cap \operatorname{supp} B = \emptyset \qquad \text{(coexclusive)}$$
$$M(A,B) = M(A/_{\operatorname{supp} A}, B/_{\operatorname{supp} A}) \qquad \text{(left-restrictable)}$$

for all A and B from the class $\mathcal{F}(U)$ of all fuzzy sets in U, with $\operatorname{supp} A$ the support of $A \in \mathcal{F}(U)$, i.e., $\operatorname{supp} A = \{u \in U \mid A(u) > 0\}$, and $A/_V$, for $A \in \mathcal{F}(U)$ and $V \subseteq U$, the restriction of A to V, i.e., the $V \to [0,1]$ mapping that associates $A(v)$ with each $v \in V$. Putting it in words, we want the value returned by M to be 0 if and only if none of the query tags have been applied to the object in question, and we do not want the returned value to depend on the membership degrees of the tags that are not part of the query. The latter is particularly important, mainly because it is advantageous from a computational point of view, but also because it ensures that we cannot obtain different values for objects with fuzzy tags sets in which each query tag always has the same membership degree. For instance, if we would use the well-known cosine similarity measure, which is not left-restrictable, then the query $Q = \{$"british", "rock"$\}$ would lead to a different value for the fuzzy sets obtained by directly interpreting the font sizes in the tag clouds shown in Fig.1 as membership degrees, while there really is no reason to prefer one over the other in this case since the font sizes for "british" and "rock" are exactly the same in both clouds. In the final ranking, one of the corresponding songs would have to be put before the other of course, but we want that decision to only depend on the value returned by w.

The family of fuzzy comparison measures introduced in [21], which includes the popular cosine similarity measure, makes it very easy to systematically construct fuzzy comparison measures that satisfy given properties. All members of this family are instances of a triparametric general form, and several properties can be ensured by imposing constraints on the parameters $\mathcal{H}$, φ_1, and φ_2 of this general form, where $\mathcal{H}$ is a binary fuzzy aggregation operator, i.e., an increasing $[0,1]^2 \to [0,1]$ mapping such that $\mathcal{H}(0,0) = 0$ and $\mathcal{H}(1,1) = 1$, and φ_1 and φ_2 are two $[0,1]^3 \to \mathbb{R}$ mappings.[3] As value of the first parameter $\mathcal{H}$, we chose the product t-norm T_{P} given by $T_{\mathrm{P}}(x,y) = x \cdot y$, for all $x, y \in [0,1]$, because it is a very simple operator that is interactive [27], i.e., a modification of x or y always implies an alteration of $T_{\mathrm{P}}(x,y)$ if $x \neq 0$ and $y \neq 0$. The general form from [21] then reduces to:

$$M(A,B) = \frac{\varphi_1(||A \cap_{T_{\mathrm{P}}} A||, ||B \cap_{T_{\mathrm{P}}} B||, ||A \cap_{T_{\mathrm{P}}} B||)}{\varphi_2(||A \cap_{T_{\mathrm{P}}} A||, ||B \cap_{T_{\mathrm{P}}} B||, ||A \cap_{T_{\mathrm{P}}} B||)} \quad (6)$$

for all $A, B \in \mathcal{F}(U)$, with $||.||$ the $\mathcal{F}(U) \to [0,1]$ mapping given by $||A|| = |A|/|U|$ for all $A \in \mathcal{F}(U)$, and $\cap_{T_{\mathrm{P}}}$ an infix notation for the pointwise extension of T_{P}, i.e., $(A \cap_{T_{\mathrm{P}}} B)(u) = T_{\mathrm{P}}(A(u), B(u))$ for all $u \in U$, with A and B fuzzy sets in U. Using the proofs provided in [21], it can easily be verified that this $\mathcal{F}(U) \times \mathcal{F}(U) \to \mathbb{R}$ mapping is an exclusive, coexclusive, and left-restrictable fuzzy comparison measure when the formulas

$$z \leq \sqrt{x \cdot y} \implies 0 \leq \varphi_1(x,y,z) \leq \varphi_2(x,y,z)$$
$$\varphi_1(x,y,0) = 0 \qquad \varphi_1(x,u,z) = \varphi_1(x,v,z)$$
$$\varphi_1(\dot{x},\dot{y},\dot{z}) > 0 \qquad \varphi_2(x,u,z) = \varphi_2(x,v,z)$$

are all satisfied for every $x, y, z, u, v \in [0,1]$ and $\dot{x}, \dot{y}, \dot{z} \in]0,1]$. The simplest solution that we could find for this is $\varphi_1(x,y,z) = z$ and $\varphi_2(x,y,z) = 1$, for all $x, y, z \in [0,1]$, which leads to:

$$M(A,B) = \frac{||A \cap_{T_{\mathrm{P}}} B||}{1} = \frac{|A \cap_{T_{\mathrm{P}}} B|}{|U|} \quad (7)$$

for all $A, B \in \mathcal{F}(U)$, i.e., a fuzzification of the Russell and Rao coefficient [28]. We then have

$$\begin{aligned} M(\widetilde{Q}, \widetilde{O}_i) &= M(\widetilde{Q}/_{\operatorname{supp}\widetilde{Q}}, \widetilde{O}_i/_{\operatorname{supp}\widetilde{Q}}) && (M \text{ left-restr.}) \\ &= M(\widetilde{Q}/_Q, \widetilde{O}_i/_Q) && (\operatorname{supp}\widetilde{Q} = Q) \\ &= \frac{1}{|Q|} \sum_{t \in Q} \widetilde{Q}(t)\, \widetilde{O}_i(t) && (\text{definition } M) \\ &\propto \sum_{t \in Q} idf(t)\, ntf_i(t) && (\text{def. } \widetilde{Q} \text{ and } \widetilde{O}_i) \end{aligned}$$

and thus:

$$score(\widetilde{Q}, \widetilde{O}_i) \propto w(i) \sum_{t \in Q} idf(t)\, ntf_i(t) \quad (8)$$

Hence, we merely need to compute $w(i) \sum_{t \in Q} idf(t)\, ntf_i(t)$ to determine the ranking of object i for query Q.

4.3 User feedback

Last.fm's Playground was launched in May 2008, with the "Multi Tag Search" as one of the initial demonstrations. Since then, this demonstration has rather consistently been the most popular one over a period of several months. It also seemed to be more "sticky" than the other initial demonstrations, since it was the only one for which the number of page views increased after launch. Moreover, we received substantially more positive direct user feedback for it than any of the other initial demonstrations. Hence, the users seemed to like the "Multi Tag Search" quite a lot, which suggests that it must at least work reasonably well. The great variety in the supplied queries confirms this observation. We received about 25 thousand unique queries so far, and, as illustrated by Table 1, several different types of tags occur in these queries.

[3] We required φ_1 and φ_2 to be increasing in their first and second argument in [21], but since the proofs for the constraints considered in this paper do not rely on this prerequisite, we can omit it here.

ISBN: 978-989-95079-6-8

Table 1: Results of a manual classification of the 250 most frequently occurring query tags.

Type	Frequency	Examples
Genre	51%	rock, pop
Style	26%	experimental
Locale	7%	british, detroit
Instrument	5%	piano, guitar
Mood	4%	sad, romantic
Time	3%	90s, oldies
Opinion	2%	beautiful, love
Other	2%	free, cover

5 Conclusion

We have shown that fuzzy set theory has a lot to offer to the music information retrieval community. By representing music objects as fuzzy sets, the possibility arises to exploit the rich mathematical toolset provided by this theory, including fuzzy aggregation operators, fuzzy comparison measures, and fuzzy relations. As a practical example, we explained how we employed fuzzy set theory to build the popular "Multi Tag Search" demonstration on Last.fm's Playground. Using previous work on fuzzy comparison measures, we were able to systematically construct a suitable comparison measure for this tag-based search engine, instead of merely relying on intuition and trial-and-error. Rather than arbitrarily choosing one of the many comparison measures available in the literature and checking if it works as expected, we started from the desired properties and systematically constructed a simple and computationally efficient measure that satisfies these properties. Since different applications have different requirements, such a systematic way of constructing comparison measures can be very valuable for building applications in which objects need to be compared, as is commonly the case in music information retrieval as well as in certain other domains.

Acknowledgement

The first author thanks the Fund for Scientific Research–Flanders (FWO) for funding this research.

References

[1] W. Glaser, T. Westergren, J. Stearns, and J. Kraft. Consumer item matching method and system. US Patent Number 7003515, 2006.

[2] D. Turnbull, L. Barrington, and G. Lanckriet. Five approaches to collecting tags for music. In *Proceedings of the ISMIR International Conference on Music Information Retrieval*, 2008.

[3] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7:76–80, 2003.

[4] P. Lamere. Social tagging and music information retrieval. *Journal of New Music Research*. To appear.

[5] E. Pampalk. *Computational models of music similarity and their application in music information retrieval*. PhD thesis, Vienna University of Technology, 2006.

[6] D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green. Automatic generation of social tags for music recommendation. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in neural information processing systems 20*. MIT Press, 2008.

[7] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech, and Language Processing*, 16:467–476, 2008.

[8] M. Sordo, C. Lauier, and O. Celma. Annotating music collections: How content-based similarity helps to propagate labels. In *Proceedings of the ISMIR International Conference of Music Information Retrieval*, 2007.

[9] K. Bosteels and E. Kerre. Fuzzy Audio Similarity Measures Based on Spectrum Histograms and Fluctuation Patterns. In A.-E. Hassanien, A. Abraham, and J. Kacprzyk, editors, *Computational Intelligence in Multimedia Processing: Recent Advances*. Springer, 2008.

[10] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[11] E. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, 2000.

[12] J. Aczel. On mean values. *Bulletin of the American Mathematical Society*, 54:392–400, 1948.

[13] M. Sugeno. *Theory of fuzzy integrals and its application*. PhD thesis, Tokyo Institute of Technology, 1974.

[14] G. Choquet. Theory of capacities. *Annales de L'Institut Fourier*, 5:131–295, 1953.

[15] R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190, 1988.

[16] M. Detyniecki. Numerical aggregation operators: State of the art. In *International Summer School on Aggregation Operators and their Applications*, 2001.

[17] B. De Baets and H. De Meyer. Transitivity-preserving fuzzification schemes for cardinality-based similarity measures. *European Journal of Operational Research*, 160(3):726–740, 2005.

[18] S. Janssens. *Bell inequalities in cardinality-based similarity measurement*. PhD thesis, Ghent University, 2006.

[19] J. Barros, J. French, W. Martin, P. Kelly, and T. Cannon. Using the triangle inequality to reduce the number of comparisons required for similarity-based retrieval. In *Storage and Retrieval for Image and Video Databases*, 1996.

[20] K. Bosteels and E. Kerre. A biparametric family of cardinality-based fuzzy similarity measures. *New Mathematics and Natural Computation*, 3:307–319, 2007.

[21] K. Bosteels and E. Kerre. A triparametric family of cardinality-based fuzzy similarity measures. *Fuzzy Sets and Systems*, 158:2466–2479, 2007.

[22] K. Bosteels and E. Kerre. On a reflexivity-preserving family of cardinality-based fuzzy comparison measures. *Information Sciences*. To appear.

[23] K. Bosteels and E. Kerre. Fuzzy heuristics for dynamic playlist generation. Submitted.

[24] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24:513–525, 1988.

[25] H. Blanken, A. de Vries, H. Blok, and L. Feng, editors. *Multimedia retrieval*. Springer, 2007.

[26] P. Norvig. Theorizing from data: Avoiding the capital mistake. CITRIS Distinguished Speaker Series, UC Berkeley, 2006.

[27] D. Dubois and H. Prade. *Fuzzy sets and systems: Theory and applications*. Academic Press, 1980.

[28] B. Sarker. The resemblance coefficients in group technology: A survey and comparative study of relational metrics. *Computers and Industrial Engineering*, 30:103–116, 1996.

ISBN: 978-989-95079-6-8

Type–II Fuzzy Possibilistic C-Mean Clustering

M.H. Fazel Zarandi[1], M. Zarinbal[1], I.B. Turksen[2,3]

1 Department of Industrial Engineering, Amirkabir University of Technology, P.O. Box 15875-4413, Tehran, Iran
2 Department of Industrial Engineering, TOBB Economy and Technology University, Sogutozo Cad. No:43, Sogutozo, Ankara 06560, Turkey
3 Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ont., Canada M5S 3G8
Email: zarandi@aut.ac.ir, mzarinbal@aut.ac.ir, turksen@mie.utoronto.ca

Abstract—*Fuzzy clustering is well known as a robust and efficient way to reduce computation cost to obtain the better results. In the literature, many robust fuzzy clustering models have been presented such as Fuzzy C-Mean (FCM) and Possibilistic C-Mean (PCM), where these methods are Type-I Fuzzy clustering. Type-II Fuzzy sets, on the other hand, can provide better performance than Type-I Fuzzy sets, especially when many uncertainties are presented in real data. The focus of this paper is to design a new Type-II Fuzzy clustering method based on Krishnapuram and Keller PCM. The proposed method is capable to cluster Type-II fuzzy data and can obtain the better number of clusters (c) and degree of fuzziness (m) by using Type-II Kwon validity index. In the proposed method, two kind of distance measurements, Euclidean and Mahalanobis are examined. The results show that the proposed model, which uses Mahalanobis distance based on Gustafson and Kessel approach is more accurate and can efficiently handle uncertainties.*

Keywords— *Type-II Fuzzy Logic; Possibilistic C-Mean (PCM); Mahalanobis Distance; Cluster Validity Index;*

1. Introduction

Clustering is an important method in data mining, decision-making, image segmentation, pattern classification, and etc. Fuzzy clustering can obtain not only the belonging status of objects but also how much the objects belong to the clusters. In the last 30 years, many fuzzy clustering models for crisp data have been presented such as Fuzzy K-Means and Fuzzy C-Mean (FCM) [1]. FCM is a popular clustering method, but its memberships do not always correspond well to the degrees of belonging and it may be inaccurate in a noisy environment [2]. To relieve these weaknesses, Krishnapuram and Keller presented a Possibilistic C-Mean (PCM) approach [3]. In addition, in real data, there exist many uncertainties and vaguenesses, which Type-I Fuzzy sets are not able to directly model as their membership functions are crisp. On the other hand, as Type-II membership functions are fuzzy, they are able to model uncertainties more appropriately [2, 4]. Therefore, Type-II Fuzzy Logic Systems have the potential to provide better performance than Type-I [5].

In the case of combining Type-II Fuzzy logic with clustering methods, the data can be clustered more appropriately and more accurately. The focus of this paper is to design a new Type-II Fuzzy clustering method based on Krishnapuram and Keller PCM. The proposed method is capable to cluster Type-II fuzzy data and can obtain the better number of clusters (c) and degree of fuzziness (m) by using Type-II Kwon validity index.

The rest of this paper is organized as follows: The clustering methods are reviewed in Section 2. Section 3 presents the historical review of Type-II Fuzzy Logic. Section 4 is dedicated to the proposed method and Section 5 presents the experimental results. Finally, conclusions are presented in Section 6.

2. Clustering Methods

The general philosophy of clustering is to divide the initial set into homogenous groups [6] and to reduce the data [1]. Clustering is useful in several exploratory decision-making, machine learning, data mining, image segmentation, and pattern classification [7]. In literature, most of the clustering methods can be classified into two types: Crisp clustering and Fuzzy clustering. Crisp clustering assigns each data to a single cluster and ignores the possibility that these data may also belongs to other clusters [8]. However, as the boundaries between clusters could not be defined precisely, some of the data could belong to more than one cluster with different positive degrees of memberships [6]. This clustering method considers each cluster as a fuzzy set and the membership function measures the degree of belonging of each feature in a cluster. So, each feature may be assigned to multiple clusters with some degree of belonging [8]. Two important applied models of fuzzy clustering, Fuzzy C-Means, and Possibilistic C-Means are described as follows:

Fuzzy C-Means (FCM): Fuzzy C-Means clustering model can be defined as follows [9]:

$$\min\left\{J(x,\mu,c)=\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}^{m}\,d_{ij}^{2}\right\} \quad (1)$$

$$\text{ST:}\begin{cases}0<\sum_{j=1}^{N}\mu_{ij}<N \quad \forall i\in(1,2,\dots,c) & (2)\\ \sum_{i=1}^{c}\mu_{ij}=1 \qquad \forall j\in(1,2,\dots,N) & (3)\end{cases}$$

where, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, d_{ij} is the distance between the j^{th} data and the i^{th} cluster center, m is the degree of fuzziness, c is the number of clusters, and N is the number of the data.

Although FCM is a very good clustering method, it has some disadvantages: the obtained solution may not be a desirable solution and the FCM performance might be inadequate, specially when the data set is contaminated by noise. In addition, the membership values show degrees of

ISBN: 978-989-95079-6-8

probabilities of sharing [10] and not only depend on the distance of that point to the cluster center, but also on the distance from the other cluster centers [11]. In addition, when the norm used in the FCM method is different to the Euclidean, introducing restrictions is necessary, e.g., Gustafson and Kessel [12] and Windham limit the volume of the groups using fuzzy covariance and scatter matrices, respectively [13].

Possibilistic C-Mean (PCM): To improve the FCM weaknesses, Krishnapuram and Keller created a possibilistic approach, which uses a possibilistic type of membership function to describe the degree of belonging. It is desirable that the memberships for representative feature points be as high as possible and unrepresentative points have low membership. The objective function, which satisfies the requirements, is formulated as follows [3]:

$$\min\left\{J_m(x,\mu,c)=\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}^{m}\,d_{ij}^{2}+\sum_{i=1}^{c}\eta_i\sum_{j=1}^{N}(1-\mu_{ij})^{m}\right\} \quad (4)$$

where, d_{ij} is the distance between the j^{th} data and the i^{th} cluster center, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, m is the degree of fuzziness, η_i is a suitable positive number, c is the number of the clusters, and N is number of the data. μ_{ij} can be obtained by using (5) [3]:

$$\mu_{ij}=\frac{1}{1+\left(\frac{d_{ij}^{2}}{\eta_i}\right)^{\frac{1}{m-1}}} \quad (5)$$

where, d_{ij} is the distance between the j^{th} data and the i^{th} cluster center, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, m is the degree of fuzziness, η_i is a suitable positive numbers. The value of η_i determines the distance at which the membership value of a point in a cluster becomes 0.5. In practice, (6) is used to obtained η_i values. The value of η_i can be fixed or changed in each iteration by changing the values of μ_{ij} and d_{ij}, but the care must be exercised, since it may lead to instabilities [3]:

$$\eta_i=\frac{\sum_{j=1}^{N}\mu_{ij}^{m}\,d_{ij}^{2}}{\sum_{j=1}^{N}\mu_{ij}^{m}} \quad (6)$$

The PCM is more robust in the presence of noise, in finding valid clusters, and in giving a robust estimate of the centers [14].

Updating the membership values depends on the distance measurements [11].The Euclidean and Mahalanobis distance are two common ones. The Euclidean distance works well when a data set is compact or isolated [7] and Mahalanobis distance takes into account the correlation in the data by using the inverse of the variance-covariance matrix of data set which could be defines as follows [15]:

$$\mathrm{D}=\sum_{\mathrm{i,j=1}}^{\mathrm{i,j=p}}A_{ij}\,(x_i-y_i)(x_i-y_j) \quad (7)$$

$$A_{ij}=\rho_{ij}\,\sigma_i\sigma_j \quad (8)$$

where, x_i and y_i are the mean values of two different sets of parameters, X and Y. σ_i^2 are the respective variances and ρ_{ij} is the coefficient of correlation between the i^{th} and j^{th} variates. Gustafson and Kessel proposed a new approach based on Mahalanobis distance, which enables the detection of ellipsoidal clusters. Their approach focused on the case where the matrix A is different for each cluster [12].

Satisfying the underlying assumptions, such as shape and number, is another important issue in clustering methods, which could be obtained by validation indices. Xie & Beni's (XB) and Kwon are two common validity indices [1]. Xie and Beni defined a cluster validity, (9), which aims to quantify the ratio of the total variation within clusters and the separation of the clusters [1]:

$$XB(c)=\frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^{m}\left\|x_j-v_i\right\|^{2}}{N\min\limits_{i\neq j}\left\|v_i-v_j\right\|^{2}} \quad (9)$$

where, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, v_j is the center of the j^{th} cluster, m is the degree of fuzziness, c is the number of clusters, and N is number of the data. The optimal number of clusters should minimize the value of the index [1]. However, in practice, when $c\rightarrow N\Rightarrow XB\rightarrow 0$ and it usually does not generate appropriate clusters. The $V_k(c)$, (10), was proposed by Kwon based on the improvement of the XB index [16];

$$V_k(c)=\frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^{m}\left\|x_j-v_i\right\|^{2}+\frac{1}{c}\sum_{i=1}^{c}\|v_i-\bar{v}\|^{2}}{\min\limits_{i\neq j}\left\|v_i-v_j\right\|^{2}} \quad (10)$$

where, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, v_j is center of the j^{th} cluster, $\bar{v}$ is the mean of the cluster centers, m is the degree of fuzziness, c is the number of clusters, and N is number of the data. To assess the effectiveness of clustering method, the smaller the V_k, the better the performance [16].

All of the clustering methods and validation indices, mentioned above, are based on Type-I fuzzy set. However, in real world, there exists many uncertainties, which Type-I fuzzy could not model them. Type-II fuzzy set, on the other hand, is able to successfully model these uncertainties [4].

3. Type-II Fuzzy Clustering

The concept of a Type-II fuzzy set, was introduced by Zadeh as an extension of Type-I fuzzy set [17]. A Type-II fuzzy set is characterized by fuzzy membership function, i.e., the membership grade for each element of this set is a fuzzy set in interval [0,1]. Such sets can be used in situations where there are uncertainties about the membership values [18]. Type-II fuzzy logic is applied in many clustering methods e.g., [19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. There are essentially two types of Type-II fuzziness: Interval-Valued Type-II and generalized Type-II. Interval-Valued Type-II fuzzy is a special Type-II fuzzy, where the upper and lower bounds of membership are crisp and the spread of membership distribution is ignored with the assumption that membership values between upper and lower values are

ISBN: 978-989-95079-6-8

uniformly distributed or scattered with a membership value of 1 on the $\mu(\mu(x))$ axis (Figure 1.a). Generalized Type-II fuzzy identifies upper and lower membership values as well as the spread of membership values between these bounds either probabilistically or fuzzily. That is there is a probabilistic or possibilistic distribution of membership values that are between upper and lower bound of membership values in the $\mu(\mu(x))$ axis (Figure 1.b) [29].

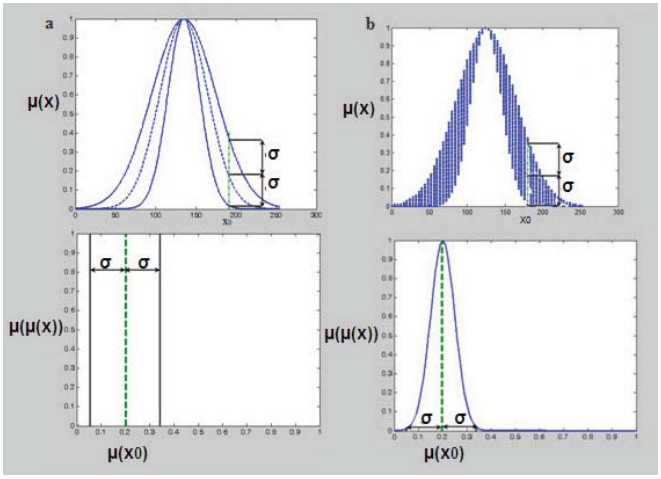


Figure 1:(a) Interval Valued Type-II (b) Generalized Type-II

4. Proposed Type-II PCM Method

Considering the growing application areas of Type-II fuzzy logic, designing a Type-II clustering method is essential. Several researchers designed a Type-II fuzzy clustering method based on FCM but FCM itself has some weaknesses, which make some of the developed methods ineffective in situations in which the data set is contaminated by noise, the norm used is different from the Euclidean, or the pixels on an input data are highly correlated. PCM could improves these weaknesses.

The proposed method is the extension of Krishnapuram and Keller Possibilistic C-Mean (PCM). Here, the membership functions are Type-II Fuzzy, the distance is assumed to be Euclidean and Mahalanobis and Type-II Kwon validity index is used to find the optimal degree of fuzziness (*m)* and number of clusters (*c)*. The proposed Type-II PCM model is as follows:

$$J_m(x,\tilde{\mu},c) = min\left[\sum_{i=1}^{c}\sum_{j=1}^{N}\tilde{\mu}_{ij}^{m} D_{ij} + \sum_{i=1}^{c}\eta_i\sum_{j=1}^{N}\left(1-\tilde{\mu}_{ij}\right)^m\right] \quad (11)$$

$$S.T:\begin{cases} 0 < \sum_{j=1}^{N}\tilde{\mu}_{ij} < N & (12) \\ \tilde{\mu}_{ij} \in [0,1] \quad \forall i,j & (13) \\ \max\tilde{\mu}_{ij} > 0 \quad \forall j & (14) \end{cases}$$

where, $\tilde{\mu}_{ij}$ is Type-II membership for the i^{th} data in the j^{th} cluster, D_{ij} is the Mahalanobis distance of the i^{th} data to the j^{th} cluster's center, η_i is positive numbers, c is the number of the clusters, and N is the number of input data. The first term make the distance to the cluster's center be as low as possible and the second term make the membership values in a cluster to be as large as possible. The membership values for data in each cluster must lie in the interval [0,1], and their sum are restricted to be smaller than the number of input data, as shown in (12), (13), and (14).

Minimizing $J_m(x,\tilde{\mu},c)$ with respect to $\tilde{\mu}_{ij}$ is equivalent to minimizing the individual objective function defined in (15) with respect to $\tilde{\mu}_{ij}$ (provided that the resulting solution lies in the interval [0,l]).

$$J_m^{ij}(x,\tilde{\mu},c) = \tilde{\mu}_{ij}^{m} D_{ij} + \eta_i\left(1-\tilde{\mu}_{ij}\right)^m \quad (15)$$

Differentiating (15) with respect to $\tilde{\mu}_{ij}$ and setting it to 0, leads to (16) which satisfies (12), (13), and (14).

$$\tilde{\mu}_{ij} = \frac{1}{1+\left(\frac{D_{ij}}{\eta_i}\right)^{\frac{1}{m-1}}} \quad \forall i = 1,\dots,c \quad (16)$$

$\tilde{\mu}_{ij}$ is updated in each iteration and depends on the D_{ij} and η_i. As mentioned in [3], the value of η_i determines the distance at which the membership value of a point in a cluster becomes 0.5. In general, it is desirable that η_i relate to i^{th} cluster and be of the order of D_{ij} [3].

$$\eta_i = \frac{\sum_{j=1}^{N}\tilde{\mu}_{ij}^{m} D_{ij}}{\sum_{j=1}^{N}\tilde{\mu}_{ij}^{m}} \quad \forall i = 1,\dots,c \quad (17)$$

where, D_{ij} is the distance measure and number of clusters (c) and degree of fuzziness (m) are unknown. Since the parameter $\eta_i$is independent of the relative location of the clusters, the membership value $\tilde{\mu}_{ij}$ depends only on the distance of a point to the cluster centre. Hence, the membership of a point in a cluster is determined solely by how far a point is from the centre and is not coupled with its location with respect to other clusters [11].

The clustering method needs a validation index to define the number of clusters (c) and degree of fuzziness (m), which are used in (15), (16), and (17). Therefore a Type-II Kwon Index based on Kwon index is designed, which is represented by (18):

$$\tilde{V}_k(c) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}\tilde{\mu}_{ij}^{m}\left\|x_j-\tilde{v}_i\right\|^2 + \frac{1}{c}\sum_{i=1}^{c}\|v_i-\bar{\tilde{v}}\|^2}{\min_{i\neq j}\left\|\tilde{v}_i-\tilde{v}_j\right\|^2} \quad (18)$$

where, $\tilde{\mu}_{ij}$ is Type-II possibilistic membership values for the i^{th} data in the j^{th} cluster, $\tilde{v}_i$ is the i^{th} center of cluster, $\bar{\tilde{v}}$ is the mean of centers, N is the number of input data, c is the number of the classes and m is the degree of fuzziness. The first term in the numerator denotes the compactness by the sum of square distances within clusters and the second term denotes the separation between clusters, while denominator denotes the minimum separation between clusters, so the smaller the $\tilde{V}_k(c)$, the better the performance.

In sum, the steps of the proposed clustering method are described bellow and are shown in Figure 2.

Step 1: Define the initial parameters including:

- Maximum iteration of the method (R)
- Number of the clusters (*c=2* is the initial value)

ISBN: 978-989-95079-6-8

- Degree of fuzziness (m=1.5 is the initial value)
- Primary membership functions ($\tilde{\mu}_{ij}{}^{0}$) (Note that these membership functions are Type-II)

Step 2: Estimate η_i by using (17).

Step 3: Calculate the membership functions for each data in each cluster ($\tilde{\mu}_{ij}{}^{r}$) by using (16).

Step 4: If the difference between two membership functions for each data is bigger than the threshold, defined by user, ($\left|\tilde{\mu}_{ij}{}^{r} - \tilde{\mu}_{ij}{}^{r-1}\right| > \varepsilon$) go to step 4.1 and 4.2. On the other hand, go to step 4.3.

Step 4.1: $r = r + 1$

Step 4.2: Recalculate $\tilde{\mu}_{ij}{}^{r}$

Step 4.3: Compute the Kwon index ($\tilde{V}_k$).

Step 5: If the difference between two Kwon indexes for each membership functions is bigger than the threshold, ($\left|\tilde{V}_k{}^{r} - Vkr{-}1{>}\varepsilon\right)$, go to step 5.1. On the other hand, go to step 5.2.

Step 5.1: Increase degree of fuzziness, m and run the method for another iteration.

Step 5.2: If the number of clusters are smaller than the number of data, ($c < N$), go to step 5.2.1. On the other hand, go to step 5.2.2.

Step 5.2.1: Run the method for another iteration

Step 5.2.2: Returns the value of *c, m*, and $\tilde{\mu}_{ij}{}^{r}$.

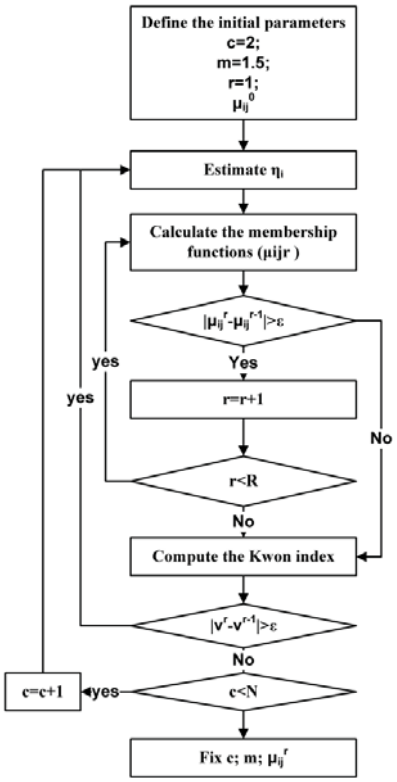


Figure 2: Type-II PCM Algorithm

5. Expand and Compare

In order to show the behavior of the proposed method, an image is used as input data. There may have been many uncertainties in images such as uncertainties caused by projecting a 3D object into a 2D image or digitizing analog pictures into digital images, and the uncertainty related to boundaries and non-homogeneous regions. Therefore, Type-II fuzzy logic can provide better performance than Type-I, this is shown by generating two models based on Type-I and Type-II Possibilistic C-Mean (PCM), each has two kind of distance measure, Euclidean and Mahalanobis. The Kwon Validity index is used to validate the results (m and c) as shown in Figure 2 and table 1, 2, 3 and 4. The results show that Type-II PCM using Mahalanobis distance can obtain better values for degree of fuzziness and number of clusters, which both are used for calculating the membership functions.

Table 1 shows the results of Kwon Validity index for Type-I PCM and Table 2 shows the results for Type-II PCM. In both of these tables Euclidean distance is used as the distance function. The elements of the tables are number of clusters (c) and degree of fuzziness (m) as input variables and the Kwon index values ($\tilde{V}_k$) as results, e.g., for m=2.7 and c=3, (2.7,3), the Kwon index for Type-I PCM is 4.5696 and is 750.34 for Type-II PCM. For m=3.3 and c=2, (3.3,2), the Kwon index could not be defined in both Type-I and Type-II PCM.

Table 3 shows the results of Kwon Validity index for Type-I PCM and Table 4 shows the results for Type-II PCM. In both of these tables Mahalanobis distance are used. The elements of the tables are number of clusters *(c)* and degree of fuzziness *(m)* as an input variables and the Kwon index values as results, i.e., for m=4.1 and c=3, (4.1,3), the Kwon index is 1.88 for Type-I PCM and is 33.686 for Type-II PCM.

ISBN: 978-989-95079-6-8

Table 1- Kwon Values for Type-I PCM with Euclidean Distance

c\m	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5
2	0.250	0.25	0.25	0.2500	0.2501	0.2509	0.2515	0.2662	0.2758	NaN	NaN
3	3.734	3.958	3.956	3.9877	4.0142	4.0762	4.5696	5.7768	6.889	NaN	NaN
4	5.95	5.944	6.05	6.3488	6.4236	2.464	987.99	7.0394	NaNi	NaN	NaN
5	5.351	5.367	5.474	5.7481	5.818	5.4528	991.14	11769	NaNi	NaN	NaN

Table 2- Kwon Values for Type-II PCM with Euclidean Distance

c\m	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5
2	57.82	38.36	14.53	14.436	30.184	6.2069	2.6636	1.5969	2.1884	NaN	NaN
3	514.7	113.3	92.05	37.547	484.97	201.14	750.34	392.91	10.381	NaN	NaN
4	9.17E+05	2.62E+05	8654.5	3.78E+05	1979.5	92.563	1854.1	331.16	NaN	NaN	NaN
5	1.01E+06	77060	6454	1534.8	386.57	130.98	9276.8	1240.2	NaN	NaN	NaN

Table 3- Kwon Values for Type-I PCM with Mahalanobis Distance

c\m	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1
2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.27	0.28	0.28	0.28	0.28	0.27	0.27	0.27	0.27	0.27	0.26	0.26
3	3.73	3.96	3.95	3.99	4.01	4.08	4.57	5.78	6.88	7.56	2.53	1.02	0.98	1.88	3.76	8.04	16.6	37.5	84.2
4	5.95	5.94	6.05	6.35	6.42	2.46	988	7.04	10.44	32.1	28.6	46.4	117	351	716	3080	5262	12599	28180
5	5.35	5.37	5.47	5.75	5.82	5.45	991	11769	587	282	56.5	49.3	121	356	716	3061	5235	12625	28521

Table 4- Kwon Values for Type-II PCM with Mahalanobis Distance

c\m	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1
2	57.8	38.3	14.5	14.4	30.1	5.75	2.69	1.60	2.74	4.42	6.2	9.09	9.26	9.54	9.71	9.23	11.2	8.58	8.96
3	515	113	92.1	37.5	485	345	600	481	1043	29.5	11.1	9.24	13.9	33.7	76.2	232	399	853	2318
4	9.17E+05	2.62E+05	8654.5	3.78E+05	1979.5	92.563	192.2	417.1	220.20	61.774	177.69	539.7	132.5	490.3	774.6	500.61	570.76	1.41E+05	4.21E+05
5	1.01E+06	77060	6454.5	1534.8	386.57	130.98	956.4	480.6	230.29	885.7	278.64	1.03E+05	6.50E+05	1.93E+06	1.26E+07	2.32E+07	4.90E+07	6.78E+07	1.11E+08

By comparing Type-I and Type-II PCM, the following conclusions can be obtained:

- In the same distance function cases, Kwon index is ascending for Type-I PCM, and its procedure is not clear for Type-II PCM case. Therefore, in Type-I PCM in all conditions the optimum *(m,c)* pairs are (1.5,2), (1.5,3), (1.5,4), (1.5,5), which may not be good results. However, in Type-II PCM the optimum *(m,c)* pairs are (2.9,2), (3.7,3), (3.3,4), (2.5,5), which seems to be good results. Figures 3 and 4 show the Kwon index results for *c=2* measured by Mahalanobis and Euclidean distance.

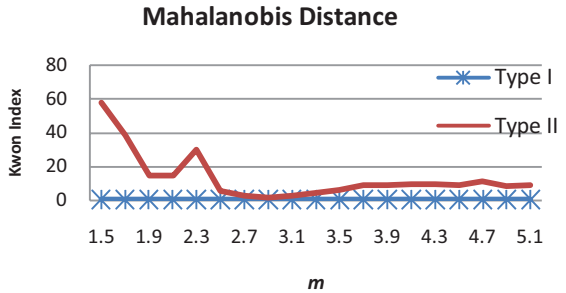


Figure 3: Kwon Index Result for c=2 and Mahalanobis Distance (based on Table 3 and 4)

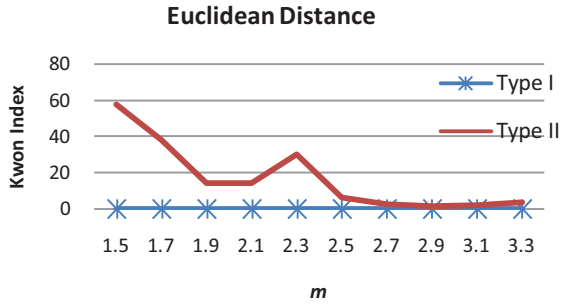


Figure 4- Kwon Index Result for c=2 and Euclidean Distance (based on Table 3 and 4)

- In the same type of fuzzy logic (Type-I or Type-II) case, for m<3.1 using two different distance functions did not show many differences for Kwon index results. However, for m>3.1, the Kwon index could be calculated by Mahalanobis but it is not defined for Euclidean distance, as shown in Figures 5.

ISBN: 978-989-95079-6-8

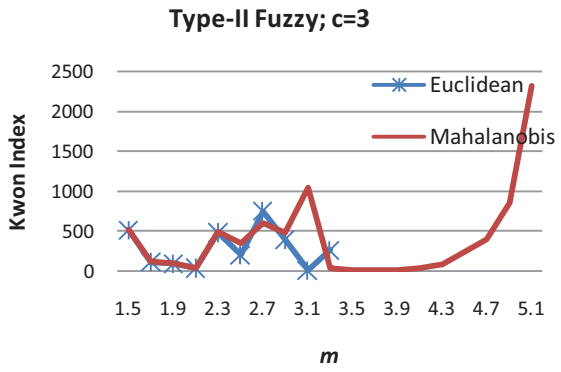


Figure 5: Kwon Index for c=3 and Different Distance Functions (based on Table 2 and 4)

6. Conclusions

This paper has presented a Type-II Possibilistic C-Mean (PCM) method for clustering purposes. The results of the proposed method are compared with Type-I PCM using an image as an input data and two kind of distance functions, Euclidean and Mahalanobis. The results show that Type-II PCM using Mahalanobis distance can provide better values for degree of fuzziness and number of clusters, which both are used in calculating the membership functions. Therefore the proposed clustering method is more accurate, can provide better performance and can handle uncertainties that exist in the data efficiently.

References

[1] J.V. Oliveira and W. Pedrycz, *Advances in Fuzzy Clustering and Its Applications*. John Wiley & Sons Ltd., 2007.

[2] J.M. Mendel and R. John, Type 2 Fuzzy Sets Made Simple. *IEEE Transactions On Fuzzy Systems*, 10(2):117-127, 2002.

[3] R. Krishnapuram and J.M. Keller, A Possibilistic Approach to Clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98-110, 1993.

[4] R. Seising (Ed.), *Views on Fuzzy Sets and Systems from Different Perspectives*. Springer-Verlag, 2009.

[5] J.M. Mendel et al., Interval Type 2 Fuzzy Logic Systems Made Simple. *IEEE Transactions on Fuzzy Systems,* 14(6):808-821, 2006.

[6] E. Nasibov and G. Ulutagay, A New Unsupervised Approach for Fuzzy Clustering. *Fuzzy Sets and Systems*, Article in Press.

[7] A.K Jain et al., Data clustering: A review. *ACM Computing Surveys*, 31(3):264-323, 1999.

[8] M. Menard and M. Eboueya Extreme Physical Information and Objective Function in Fuzzy Clustering. *Fuzzy Sets and Systems*, 128(3):285-303, 2002.

[9] I.B. Turksen, *An Ontological and Epistemological Perspective of Fuzzy Set Theory*. Elsevier Inc., 2006.

[10] H. Frigui, R. Krishnapuram, A Robust Competitive Clustering Algorithm with Applications in Computer Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):450-465 1999.

[11] K.P. Detroja et al., A Possibilistic Clustering Approach to Novel Fault Detection and Isolation. *Journal of Process Control,* 16(10):1055-1073, 2006.

[12] D.E. Gustafson and W.C. Kessel, Fuzzy Clustering with Fuzzy Covariance Matrix. *Proc. IEEE CDC, San Diego, CA,* 761-766, 1979.

[13] A. Flores-Sintas et al., A Local Geometrical Properties Application to Fuzzy Clustering. *Fuzzy Sets and Systems*, 100(3): 245-256, 1998.

[14] O. Nasraoui and R. Krishnapuram, Crisp Interpretations of Fuzzy and Possibilistic Clustering Algorithms. *In Proceedings of the 3rd European Congress on Intelligent Techniques and Soft Computing*, 1312-1318, 1995.

[15] P.C. Mahalanobis, On the generalized distance in statistics. *Proceedings National Institute of Science*, 2, 1936.

[16] C. Duo et al., An Adaptive Cluster Validity Index for the Fuzzy C-Means. *International Journal of Computer Science and Network Security*, 7(2)146-156, 2007.

[17] L.A. Zadeh, The Concept of a Linguistic Variable and its Applications to Approximate Reasoning-I. *Information Sciences*, 8:199-249, 1975.

[18] O. Castillo and P. Melin, *Type-2 Fuzzy Logic: Theory and Applications*. Springer-Verlag, 2008.

[19] A. Celikyilmaz and I.B. Turksen, Enhanced Type 2 Fuzzy System Models with Improved Fuzzy Functions. *Annual Conference of the North American Fuzzy Information Processing Society*, 140-145, 2007.

[20] B.I. Choi and F.C. Rhee, Interval Type-2 Fuzzy Membership Function Generation Methods for Pattern Recognition. *Information Sciences, Article in Press*, 2008.

[21] C. Hwang and F.C. Rhee, An interval type-2 Fuzzy C-Spherical Shells algorithm. *IEEE International Conference on Fuzzy Systems*, 2:1117-1122, 2004.

[22] C. Hwang, F.C. Rhee, Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach To C-Means. *IEEE Transactions on Fuzzy Systems,* 15(1):107-120, 2007.

[23] D.C. Lin and M.S. Yang, A Similarity Measure between Type 2 Fuzzy Sets with its Application to Clustering. *4th International Conference on Fuzzy Systems and Knowledge Discovery*, 1:726-731, 2007.

[24] F.C. Rhee, Uncertain Fuzzy Clustering: Insights and Recommendations. *IEEE Computational Intelligence Magazine*, 2(1):44-56, 2007.

[25] F.C. Rhee and C. Hwang, A Type-2 Fuzzy C-Means Clustering Algorithm. *Annual Conference of the North American Fuzzy Information Processing Society*, 4:1926-1929, 2001.

[26] O. Uncu and I.B. Turksen, Discrete Interval Type 2 Fuzzy System Models using Uncertainty in Learning Parameters. *IEEE Transactions on Fuzzy Systems*, 15(1):90-106, 2006.

[27] W.B. Zhang et al., Rules Extraction of Interval Type-2 Fuzzy Logic System based on Fuzzy C-Means Clustering. *4th International Conference on Fuzzy Systems and Knowledge Discovery*, 2:256-260, 2007.

[28] W.B. Zhang and W.J. Liu, IFCM: Fuzzy Clustering for Rule Extraction of Interval Type-2 Fuzzy Logic System. *Proceedings of 46th IEEE Conference on Decision and Control,* 5318-5322, 2007.

[29] I.B. Turksen, Type 2 Representation and Reasoning for CWW. *Fuzzy Sets and Systems*, 127(1):17-36, 2002.

ISBN: 978-989-95079-6-8

Opposite Fuzzy Sets with Applications in Image Processing

Hamid R. Tizhoosh
Pattern Analysis and Machine Intelligence Laboratory
Systems Design Engineering, University of Waterloo
200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada
Email: tizhoosh@uwaterloo.ca

Abstract— *Diverse forms of the concept of opposition are already existent in philosophy, linguistics, psychology and physics. The interplay between entities and opposite entities is apparently fundamental for balance maintenance in almost a universal manner. However, it seems that we have failed to incorporate oppositional thinking in engineering, mathematics and computer science. Especially, the set theory in general, and fuzzy set theory in particular, do not offer a formal framework to incorporate opposition in inference engines. Considering sets along with their opposites can establish a new computing scheme with a wide range of applications. In this work, preliminary definitions for opposite fuzzy sets will be established. The underlaying idea of opposition-based computing is the simultaneous consideration of guess and opposite guess, and estimate and opposite estimate, in order to accelerate learning, search and optimization. To demonstrate the applicability and usefulness of opposite fuzzy sets, a new image segmentation algorithm will be proposed as well.*

Keywords— Fuzzy sets, opposition, opposite fuzzy sets, antonym, antonymy, complement

1 Motivation

Wherever we observe the physical world and the human society, some notions of opposition exist. The intensity of *oppositeness* and the interplay between entities and opposite entities may vary greatly by closer observation. However, the antipodality, duality, complementarity, antithetics, polarity, in one word, opposition is, in different forms and at various intensity levels, present almost everywhere.

Strikingly, the human communication, expressed verbally and in written form, is without opposition unimaginable. How often we hear "on the one hand" followed by "and on the other hand" when humans weigh opposite alternatives, or how frequently we make use of antonyms: cold vs. hot, short vs. tall, true vs. false etc., or how impressively we make a point by using sayings such as "actions speak louder than words" (activity vs. passivity), "All for one and one for all" (union of opposites), "the pen is mightier than the sword" (spirit versus might). In spite of the virtual omnipresence of opposition, there is no formal framework in logic and set theory to integrate oppositional concepts. This preliminary work aims at providing elementary definitions and thoughts in this very direction.

2 What is Opposition?

Opposition is related to entities, objects or their abstractions of the same nature that are completely different in some manner. For instance, *cold* and *hot* both describe a certain temperature perception (are of the same kind), however, are completely different since they are located at opposite spots of temperature scale. Transition from one entity to its opposite is understandably accompanied with rapid and fundamental changes. Social revolutions, for instance, mainly aim at attaining opposite circumstances, dictatorship vs. democracy, by initiating sudden transitions. On the other hand, in many cases the balance due to interplay between entities and opposite entities is more dominant than rapid transition from one state to its opposite. These may be taken as initial thoughts for the purpose to improve algorithms – by smart switching between estimates and opposite estimates. However, in many other cases, this is the coexistence, the synergy of things and opposite things, and not their antagonism, that creates changes and enhances the system state.

In natural language, opposition can be detected at different levels [1]. For instance, directional opposition (north-south, up-down, left-right), adjectival opposition (ugly-handsome, long-short, high-low), and prefix opposition (thesis vs. antithesis, revolution vs. counter-revolution, direction vs. opposite direction) are all different levels of linguistic opposition. Further, one can distinguish complements (mutually exclusive properties: dead-alive, true-false), antonyms (two corresponding points or ranges of a scale: long-short, hot-cold), directional converses (two directions along an axis: east-west, up-down), and relational converses (the relative positions of two entities on opposite sides: above-below, teacher-pupil) [1]. The human communication without utilizing linguistic opposition is unimaginable.

Many examples also exist in the philosophy which deal with opposition. The concept of **yin** and **yang** originates in ancient Chinese philosophy and metaphysics, which describes two primal *opposing* but complementary forces found in all things in the universe. Also in the dualism delivered by Avesta, the book of Zoroastrians, the everlasting fight between *Ahura* (God=good) and *Ahriman* (evil) defines the worldlook. The **Table of Opposites** of Pythagoras, delivered by Aristotle, has his own list of opposites: finite-infinite, odd-even, one-many, right -left, rest-motion, straight-crooked, light-darkness, good-evil, and square-oblong. Many philosophers have focused on *one vs. many* as one of the most fundamental questions in philosophy.

In physics, *antiparticles* are subatomic particles having the same mass as one of the particles of ordinary matter but with opposite electric charge and magnetic moment. The positron (positive electron), hence, is the antiparticle of the negative electron (negatron). As another physical example, we define *electric polarization* as slight relative shift of positive and negative electric charge in opposite directions within an insulator, or dielectric, induced by an external electric field. Polarization occurs when an electric field distorts the negative cloud

ISBN: 978-989-95079-6-8

of electrons around positive atomic nuclei in a direction opposite the field[1]. And we could go back in time to the Newton's Mechanics with its principal of *action* and *reaction* which is a major example for discovering and exploiting oppositional behavior.

In mathematics we have several concepts employing some notions of opposition. For instance, the bisection method for solving equations makes use of *positive* vs. *negative* sign change in order to shrink the search interval. In probability theory, the probability of *contrary* situation is given by $1 - p$ if the initial event occurs with probability p. In Monte Carlo simulation, *antithetic* random numbers are used to reduce the variance, and in category theory there is the concept of *duality*.

These examples show that opposition plays a central rule in many fields ranging from linguistics to physics. In spite of all examples, however, it should be mentioned that understanding and defining opposition may not be trivial in some cases.

3 Opposition-Based Computing

The concept of opposition-based computing (OBC) [2, 4] is based on a quite simple idea: Whenever we are looking for a solution, we should consider the opposite solution as well. This may appear too simple, however, since opposition is almost universal, it can be understood and implemented in many different ways such that the application fields and potential impacts are immense. The idea of opposition-based computing is not entirely new in machine intelligence. Numerous works have treated and employed oppositional concepts in different ways and under different labels [4].

Past research on oppositional concepts has not been conducted under a unified framework. As such, a universal view of opposition-based learning should regard and treat entities and their opposites in a much more general framework, particularly when for application in the context of machine intelligence schemes. Among others, we should not forget that opposition considers both qualitative and abstract entities, and not only numbers.

In a preliminary paper [2], OBC-extensions of neural nets (with opposite weights), genetic algorithms (with anti-chromosomes) and reinforcement agents (with opposite actions) were provided, with experimental results indicating modest improvements in learning/searching speed. More detailed reports, on improvement of the learning behavior of evolutionary algorithms, reinforcement learning, ant colonies and neural nets have been reported as well [3, 5, 6, 7, 9, 10]. Among other benefits of incorporating of OBC into existing learn and search algorithms, total computational time has decreased by 1.6 times for differential evolution (DE) by incorporating anti-chromosomes. (This becomes even more significant when we recall that DE is generally considered a fast algorithm to begin with).

OBC may be regarded as a new philosophy for integration of a-priori knowledge (interplay between things and opposite things during search and learning procedures). However, this is a very special type of a-priori knowledge. Among other potential benefits, working with opposite entities is not accompanied with uncertainties, if the initial entities and their evaluation/fitness are known. In contrast, any work based on similarity and dissimilarity is always to a certain degree uncertain. The main question, however, is how we should find/define the opposites for every given problem. This work will attempt to find some answers within fuzzy set theory.

[1] Encyclopdia Britannica

4 Oppositional Ideas in Fuzzy Systems

Opposite fuzzy sets have been a part of fuzzy systems from the beginning. Fuzzy inferencing is without the concept of opposition virtually impossible. Linguistic terms are generally considered as pairs such as "short-tall", "cold-warm", "positive-negative" etc., something that is so natural and customary that we have overlooked its significance and failed to develop a more systematic paradigm to exploit its true potential.

Some explicit notions of employing opposition can be found in previous works published on fuzzy sets and their applications. One may immediately say that the concept of "negation" is the natural translation of opposition. A set A, describing a set of objects, is negated, $\neg A$, to quantify "not A". However, this is a general misconception. Negation cannot represent opposition, for it is too general to illustrate the oppositeness of attributes. For instance, "not very young" encompasses the entire spectrum of "young" to "very old", whereas the opposite of "very young" is clearly "very old". Most likely Roland Yager was the first one who used the term "opposition" in relation to fuzzy sets [11]. In literature on fuzzy systems, the opposition has been scarcely studied, and that only with respect to *antonyms* and their significance from a pure linguistic perspective [12, 13]. As we will see, antonyms are what we call in this paper "type-I opposite fuzzy sets".

5 Opposite Fuzzy Sets

In this section we attempt to establish a generic framework for opposite fuzzy sets. For all following definitions we consider a general class of membership functions (MFs) $\mu(x) = f(x; \mathbf{a}, \delta_1, \delta_2)$ where the vector $\mathbf{a}$ denotes the points with $\mu(a_i) = 1$, and δ_1 and δ_2 are the function widths to the left and right, respectively (Fig. 1).

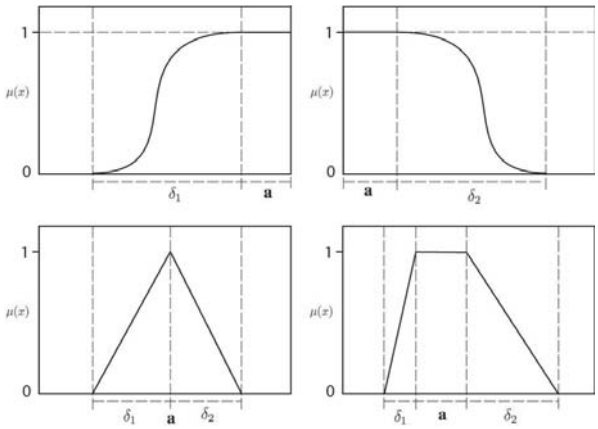


Figure 1: General MFs $\mu(x) = f(x; \mathbf{a}, \delta_1, \delta_2)$

Let X be the universe of discourse and $x \in X$ the elements, objects to be classified.

Definition (Fuzzy Set) - A fuzzy set $A \subset X$ with membership function $\mu_A(x)$ is defined as

$$A = \{(x, \mu_A(x)) | x \in X, \mu(x) \in [0, 1]\} \quad (1)$$

ISBN: 978-989-95079-6-8

The membership function is given as $\mu_A(x) = f(x; \mathbf{a}, \delta)$ where $\mu_A(\mathbf{a}) = 1 \quad \forall a_i \in \mathbf{a}$ and δ is the somatic parameter that changes the shape of the membership function.

Definition (Opposite Fuzzy Set) - Given a fuzzy set $A \subset X$, the opposite fuzzy set $\breve{A} \subset X$ with membership function $\mu_{\breve{A}}(x)$ is defined as

$$\breve{A} = \{(x, \mu_{\breve{A}}(x)) | x \in X, \mu_{\breve{A}}(x) \in [0,1]\} \tag{2}$$

where $\mu_{\breve{A}}(x) = f(x; \breve{\mathbf{a}}, \breve{\delta})$.

The vector $\mathbf{a} = [a_1, a_2, \dots]$ and its opposite vector $\breve{\mathbf{a}} = [\breve{a}_1, \breve{a}_2, \dots]$ represent the points on the universe of discourse with $\mu(a_i) = \mu(\breve{a}_i) = 1$; $\breve{\delta}$ will cause the opposite shape modification compared to original δ.

Example - If a Gaussian, bell curve, function is employed to represent $\mu_A(x) = f(x; \mathbf{a}, \delta)$, then $a = c$ and $\delta = \sigma$. The opposite membership function $\mu_{\breve{A}}(x) = f(x; \breve{\mathbf{a}}, \breve{\delta})$ will be defined with $\breve{a} = \breve{c}$ and $\breve{\delta} = \breve{\sigma}$

Since both elements of $\mathbf{a}$ and δ are real numbers, the definition of opposite fuzzy sets comes down to definition of the opposite of any arbitrary real number a.

Opposite numbers are generally defined for the set of irrational number $\mathbb{Z}$ on the interval $(-\infty, +\infty)$. For each $x \in \mathbb{Z}$ the opposite number $\breve{x}$ is defined by $\breve{x} = -x$. This definition, however, lacks generality. Following, we introduce more comprehensive definitions of opposite numbers [4].

Definition (Type-I Opposite Points[2]) - Let $P = (a_1, a_2, \cdots, a_n)$ be a point in an n-dimensional space, where $a_i \in [X^i_{min}, X^i_{max}] \in \mathbb{R}$. The type-I opposite point $\breve{P} = (\breve{a}_1, \breve{a}_2, \cdots, \breve{a}_n)$ is then completely defined where

$$\breve{a}_i = X^i_{max} + X^i_{min} - a_i. \tag{3}$$

For $n = 1$ following special cases exist:

$$\begin{array}{rcll} \breve{a} & = & -a & \text{for} \quad X_{max} = -X_{min}, \\ \breve{a} & = & 1 - a & \text{for} \quad X_{min} = 0 \quad \& \quad X_{max} = 1, \\ \breve{a} & = & a & \text{for} \quad a = \frac{X_{max} + X_{min}}{2}. \end{array} \tag{4}$$

Definition (Type I Opposite Fuzzy Sets) - The set $\breve{A}$ with membership function $\mu_{\breve{A}}(x) = f(x; \breve{a}, \breve{\delta})$ is type I opposite of the set A with membership function $\mu_A(x) = f(x; a, \delta)$ if $\breve{a}$ and $\breve{\delta}$ are type I opposites of a and δ, respectively.

Definition (Type-I Super-Opposite Points) - Let $P = (a_1, a_2, \cdots, a_n)$ be a point in n-dimensional space with $a_i \in [X^i_{min}, X^i_{max}]$ and its opposite point $\breve{P} = (\breve{a}_1, \breve{a}_2, \cdots, \breve{a}_n)$. Then all points $\breve{P}^s$ are type-I super-opposite of P when $d(\breve{P}^s, P) > d(\breve{P}, P)$ where $d(\cdot, \cdot)$ denotes a metric such as the Euclidian distance.

For $n = 1$ the type-I super-opposite of $a \in [X_{min}, X_{max}]$ can be given as

$$\breve{a}^s \in \begin{cases} [X_{min}, \breve{a}) & \text{for} \quad a > (X_{min} + X_{max})/2 \\ [X_{min}, X_{max}] - \{a\} & \text{for} \quad a = (X_{min} + X_{max})/2 \\ (\breve{a}, X_{max}] & \text{for} \quad a < (X_{min} + X_{max})/2 \end{cases}$$

[2]The terms "type I" and "type II" opposition have been taken from previously published works on OBC [4] and should not be confused with type-1 and type-2 fuzzy sets. Type-I opposites reflect linear opposition whereas type-II opposites constitute nonlinear opposition.

In other words, for $a = \frac{X_{min} + X_{max}}{2}$ the entire interval except a becomes the super-opposite of a. This means that for $a \to \frac{X_{min} + X_{max}}{2}$ the opposition generally loses its benefit against random guesses, since any random guess will be a super-opposite of the initial guess.

Definition (Type I Super-Opposite Fuzzy Sets) - The set $\breve{A}^S$ with membership function $\mu_{\breve{A}^S}(x) = f(x; \breve{a}^S, \breve{\delta}^S)$ is type I super-opposite of the set A with membership function $\mu_A(x) = f(x; a, \delta)$ if $\breve{a}^S$ and $\breve{\delta}^S$ are type I super-opposites of a and δ, respectively.

Definition (Type-I Quasi-Opposite Points) - Let $P = (a_1, a_2, \cdots, a_n)$ be a point in n dimensional space with $a_i \in [X^i_{min} + X^i_{max}]$ and its opposite point $\breve{P} = (\breve{a}_1, \breve{a}_2, \cdots, \breve{a}_n)$. Then all points $\breve{P}^q$ are type-I quasi-opposite of P when $d(\breve{P}^q, \breve{P}) = d(\breve{P}^s, \breve{P})$ and $d(\breve{P}^q, P) < d(\breve{P}^s, P)$ for any super-opposite $\breve{P}^s$.

For $n = 1$ the type-I quasi-opposite of $a \in [X_{min}, X_{max}]$ can be given as

$$\breve{a}^q \in \begin{cases} (\breve{a}, 2\breve{a}] & \text{for} \quad a > (X_{min} + X_{max})/2 \\ [X_{min}, X_{max}] - \{a\} & \text{for} \quad a = (X_{min} + X_{max})/2 \\ [2\breve{a} - X_{max}, \breve{a}) & \text{for} \quad a < (X_{min} + X_{max})/2 \end{cases}$$

Definition (Type I Quasi-Opposite Fuzzy Sets) - The set $\breve{A}^q$ with membership function $\mu_{\breve{A}^q}(x) = f(x; \breve{a}^q, \breve{\delta}^q)$ is type I quasi-opposite of the set A with membership function $\mu_A(x) = f(x; a, \delta)$ if $\breve{a}^q$ and $\breve{\delta}^q$ are type I quasi-opposites of a and δ, respectively.

Figure 2 illustrates type-I super- and quasi-opposition for one dimensional cases. Figure 3 shows a fuzzy subset A, its opposite $\breve{A}$, quasi- and super-opposite.

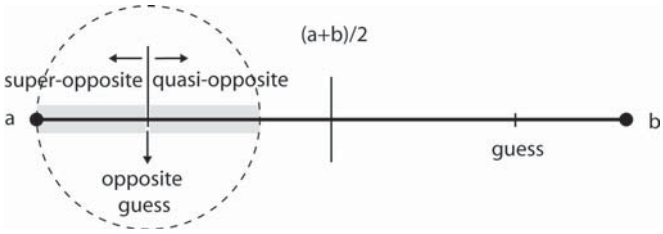


Figure 2: Definition of super- and quasi-opposition based on the distance of the opposite guess from the interval boundary.

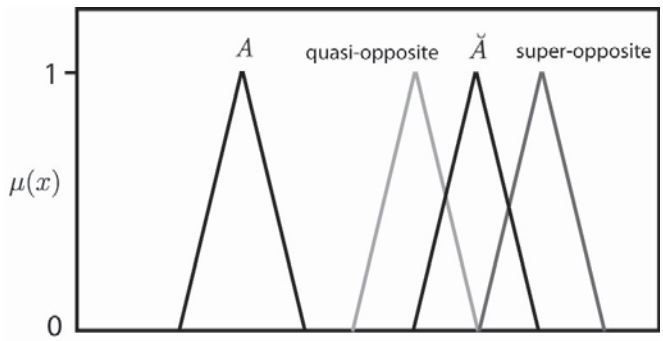


Figure 3: A fuzzy set and its opposites

Obviously type I opposition is a straightforward and a simple understanding of opposition. Generally, only linear,

ISBN: 978-989-95079-6-8

almost-linear and symmetric systems can benefit from type I opposition. In order to deal with complex and nonlinear $X-Y$-relationships we have to access higher levels of opposition.

Theorem - The entropy (fuzziness) of the fuzzy set A, denoted with $E(A)$ is equal to the entropy (fuzziness) of its type I opposite set $\breve{A}_I$: $E(A) = E(\breve{A}_I)$

Proof - since the membership function of $\breve{A}$ is just a reflection and translation of A, then $E(A) = E(\breve{A})$. □

Definition (Type-II Opposite Points) - Let $y = f(x_1, x_2, \cdots, x_n) \in \mathbb{R}$ be an arbitrary function with $y \in [y_{\min}, y_{\max}]$. For every point $P = (a_1, a_2, \cdots, a_n)$ the type-II opposite point $\breve{P} = (\breve{a}_1, \breve{a}_2, \cdots, \breve{a}_n)$ is defined by

$$\breve{a}_i = \{x| \quad \breve{y} = y_{\min} + y_{\max} - y\}. \tag{5}$$

This definition assumes that the function $f(x_1, x_2, \cdots, x_n)$ is not known but $y_{\min}$ and $y_{\max}$ are given or can be estimated. Alternatively, the temporal type-II opposite can be calculated according to

$$\breve{a}_i(t) = \left\{ \; x \; \middle| \quad \breve{y}(t) = \min_{j=1,\ldots,t} y_j + \max_{j=1,\ldots,t} y_j - y(t) \; \right\}. \tag{6}$$

If only the evaluation function $h(X)$ is available, then we may define a temporal *degree of opposition* $\breve{\Phi}$ for any two variables a_1 and a_2:

$$\breve{\Phi}(a_1, a_2, t) = \frac{|h(a_1) - h(a_2)|}{\max\limits_{j=1,\ldots,t} h(a_j) - \min\limits_{j=1,\ldots,t} h(a_j)} \in [0, 1]. \tag{7}$$

The type-II opposite $\breve{a}_i$ of any variable a_i can then be given as

$$\breve{a}_i = a_j|_{\breve{\Phi}(a_i,a_j,t)=\max\limits_k \breve{\Phi}(a_i,a_k,t)}. \tag{8}$$

Definition (Type II Opposite Fuzzy Sets) - The set $\breve{A}_{II}$ with membership function $\mu_{\breve{A}_{II}}(x) = f(x; \breve{a}, \breve{\delta})$ is type II opposite of the set A with membership function $\mu_A(x) = f(x; a, \delta)$ if $\breve{a}$ and $\breve{\delta}$ are type II opposites of a and δ, respectively.

Definition - Degree of Oppositeness for Type I Opposition For two real numbers $a, b \in X$ bounded in $[X_{min}, X_{max}]$, the degree of oppositeness can be given as

$$\breve{\varphi}_I(a, b) = \left(\frac{|a - b|}{X_{max} - X_{min}} \right)^{\beta} \tag{9}$$

where $\beta \in (0, 1]$ controls the opposition intensity.

Definition - Degree of Oppositeness for Type II Opposition Assuming the evaluation function $h(\cdot)$ is given, then for two real numbers $a, b \in X$, the degree of oppositeness can be given as

$$\breve{\varphi}_{II}(a, b) = \left(\frac{|h(a) - h(b)|}{h_{max} - h_{min}} \right)^{\beta} \tag{10}$$

where $\beta \in (0, 1]$ controls the opposition intensity. If the bounds of the evaluation function are not known, then a temporal type-II opposition may be calculated:

$$\breve{\varphi}^{(t)}_{II}(a, b) = \left(\frac{|h(a) - h(b)|}{h^{(t)}_{max} - h^{(t)}_{min}} \right)^{\beta} \tag{11}$$

Based on the introduced definitions, following observations can be made:

- Opposition, in general, can only be determined if the system parameters are bounded and *opposite* has a clear relationship with the finite interval, in which those parameters are defined.
- Type-I opposition delivers the real opposite (type-II) only for linear or quasi-linear functions. Type-I opposition can, however, be more easily calculated.
- Type-II opposition can, in absence of any information about the function $f(\cdot)$, only be calculated if an evaluation function $g(\cdot)$ (error, fitness, reward, cost etc.) is available.

Now that we have basic definitions available we can define the opposition-based computing.

> **Opposition-Based Computing (OBC) -** For any system behaving according a (unknown) function $y = f(x_1, x_2, \cdots, x_n)$ with an evaluation function $h(x_1, x_2, \cdots, x_n)$ with higher/lower values being desirable, for any estimate vector $X = (x_1, x_2, \cdots, x_n)$ to estimate the input-output relationship we calculate its opposite $\breve{X} = (\breve{x}_1, \breve{x}_2, \cdots, \breve{x}_n)$ and consider $\max(h(X), h(\breve{X}))$. The learning/search will continue with X if $h(X) \gtrless h(\breve{X})$ otherwise with $\breve{X}$.

The opposition can be calculated using the aforementioned definitions and based on the specific domain knowledge.

The definition of OBC is making two fundamental assumptions: 1) one of estimate and counter-estimate is always closer to the solution, and 2) considering the opposition is more beneficial than generating independent random solutions and taking the best among them. (this assumption has already been proved [8]).

The aforementioned definition of OBC is primarily meant to be used in the model-free context of machine intelligence, where an evaluation function such as error, cost, fitness, reward etc. is being used within multiple iterations/epochs/generations/episodes in order to find some optimal values. For pure fuzzy systems we may rely on the following definition.

> **Opposition-Based Fuzzy Inference Systems (OFIS) -** An inference mechanism which systematically uses fuzzy sets along with their opposites to construct membership functions and rules and/or carry out the inference is an opposition-based fuzzy inference system.

6 Applications

As mentioned before, we have virtually been using opposite fuzzy sets since Zadeh's first paper. However, due to lack of necessary formalism and due to investigative studies on the nature of opposition in its diverse forms, we have not exploited their potentials.

Most likely, one major application for opposite fuzzy sets is generation of membership functions (MFs) for an inference system. From the beginning days of fuzzy sets, determination of membership functions has been a challenge. Oppositional

ISBN: 978-989-95079-6-8

relationships can be considered to generate membership functions. Table 1 provides a simple algorithm for this purpose. The underlaying idea is as follows: Assuming we are looking for 3 membership functions, and assuming we can fix the position of one of them namely the set A. Then if we automatically determine the opposite fuzzy set $\breve{A}$, the third (middle) membership function, which quantifies "everything else" can be given by $\neg(A \cup \breve{A})$ (see Figure 4).

Table 1: Oppositional generation of membership functions

1.	Determine n to generate $2n+1$ membership functions
2.	For $i = 1 : n$
3.	$\mu_{A_i} = f(x; a_i, \delta_i)$
4.	end
5.	For $i = n+2 : 2n+1$
6.	Determine $\breve{a}_i$ subject to the constraint $\breve{a}_i < \breve{a}_{i+1}$
7.	Determine $\breve{\delta}_i$
8.	$\mu_{\breve{A}_i} = f(x; \breve{a}_i, \breve{\delta}_i)$
9.	end
10.	Determine the middle MF based on $\neg(A \cup \breve{A})$

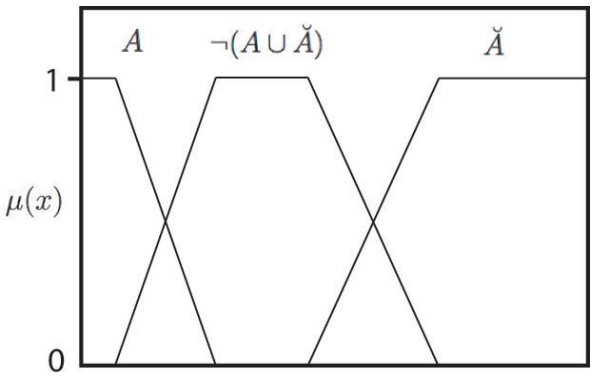


Figure 4: The set A, its opposite $\breve{A}$ and "everything else"

Processing of digital images is certainly an interesting application to verify the applicability of opposite fuzzy sets. The task of *image segmentation* is the most significant part in many vision-based application whereas segmentation is equivalent to classification of each pixel to different classes (segments). Here, based on proposed algorithm in Table 1 we can demonstrate how opposite fuzzy sets can help to segment a digital image.

Let us assume that we are interested in extracting only one object of interest in the image. The rest, regardless how many other objects, are irrelevant. Further assume we have the knowledge that the object of interest is *dark*. Then an oppositional fuzzy rule-based approach to object segmentation can be implemented as described in Table 2. As apparent from line 7 and 8 in this algorithm, we assume that the entropy of A and $\breve{A}$ are equal, which is, according to the theorem in section 5, only proved for type-I opposites. However, experimental results will show that this is a reasonable assumption, which still needs to be theoretically solidified.

Table 2: An OFIS for image segmentation

1. Define the set $A = \{$ *dark pixels* $\}$ with $\mu_A(x)$
2. Calculate the entropy of $A : E(A)$
3. Initialize the opposite membership $\breve{\mu}_A(x)$ at $\breve{a} = L-1$
4. For $i = L-1 : -1 : a+\delta$
5. Define opposite fuzzy set $\breve{A}_i$ at $\breve{a} = i$ (Fig. 5)
6. Calculate the entropy of $\breve{A}_i : E(\breve{A}_i)$
7. Calculate the entropy difference $d_i = \lvert E(\breve{A}_i) - E(A) \rvert$
8. Find the position j with $d_j = \min_i d_i$
9. Define the final opposite fuzzy set at $\breve{a} = j$
10. Define the middle membership function via $\neg(A \cup \breve{A})$
11. (OPTION 1) Create a fuzzy segmentation map via
$g' = \frac{\mu_A(g)+\mu_{\neg(A\cup\breve{A})}(g)\times(L-1)/2+\mu_{\breve{A}}(g)\times(L-1)}{\mu_A(g)+\mu_{\neg(A\cup\breve{A})}(g)+\mu_{\breve{A}}(g)}$
13. (OPTION 2) Threshold the image with $T = (a+\breve{a})/2$
14. $g' = 1$ if $g <= T$, and $g' = 0$ otherwise

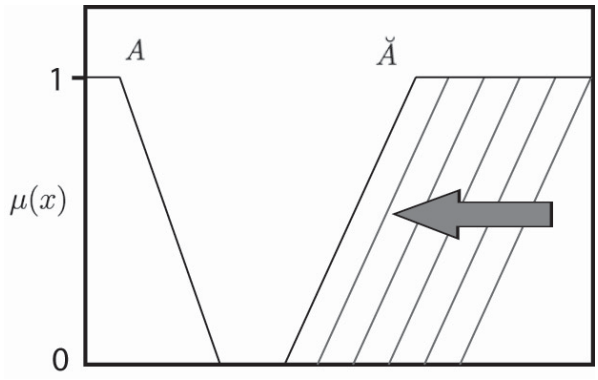


Figure 5: Moving an initial set toward A to find the opposite set $\breve{A}$

The results of this approach are presented in Figure 6. Needless to say that this simple algorithm should only demonstrate the usefulness of opposite fuzzy sets for a significant application. Most likely more sophisticated segmentation algorithms can be developed by incorporating oppositional thinking.

Breast ultrasound images have been used to verify the performance of this method (Fig. 6). The first two scan contain anechoic (dark) breast cysts. These are relatively easy to segment with almost any thresholding technique. The other two contain breast masses that are no cyst and should be examined on malignancy. These cases are challenging and cannot be easily segmented.

7 Conclusions

In this paper, we attempted to establish a formalism for opposite fuzzy sets and introduce some definitions. Preliminary results for a simple image segmentation method were provided as well. Opposition-based fuzzy inference systems (OFISs) exploit domain knowledge and seem to have real-world applications. Many questions, however, remain open with respect to the nature of opposition and how it can be embedded into existing fuzzy systems.

ISBN: 978-989-95079-6-8

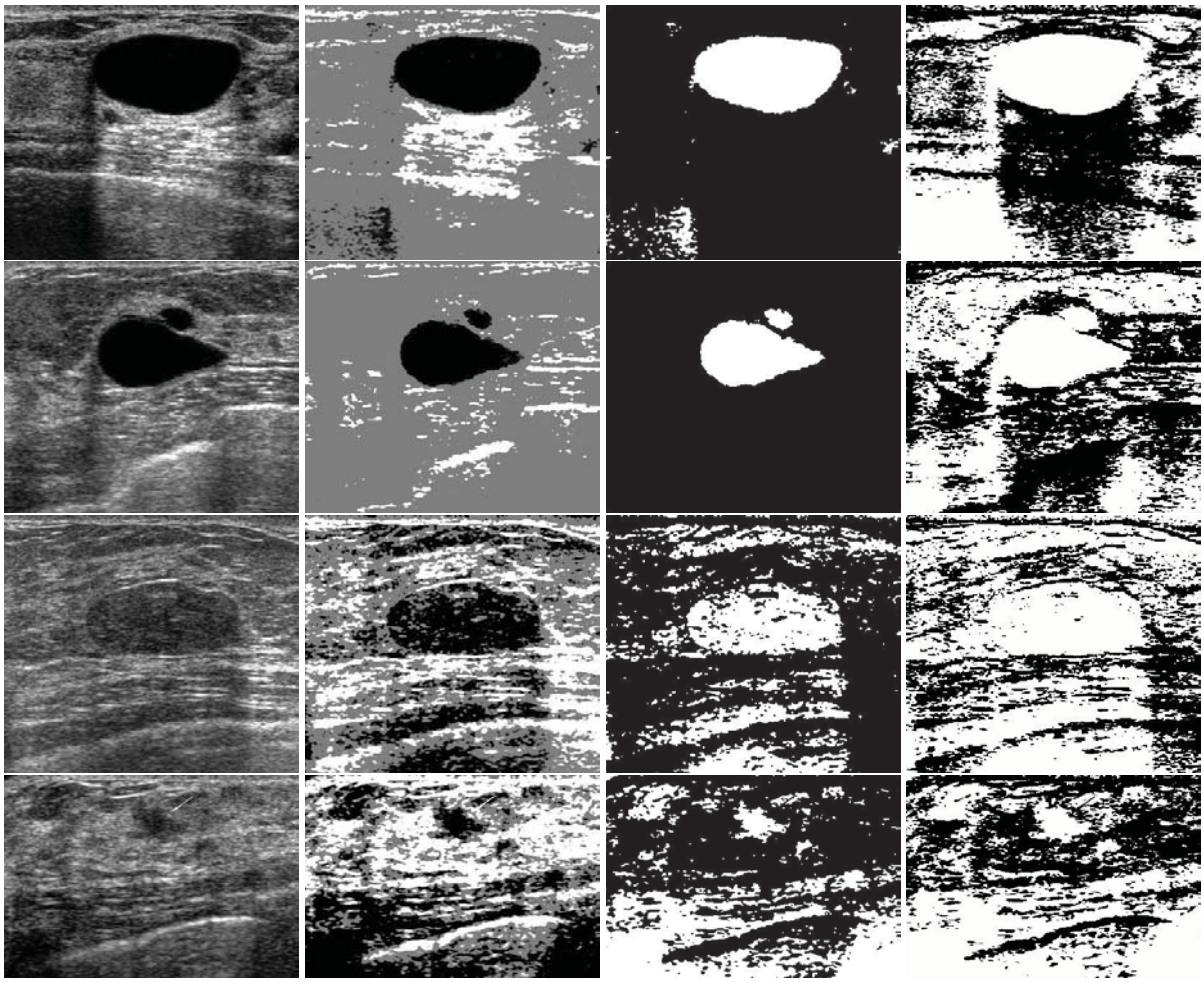


Figure 6: From left: first column: original breast ultrasound scans, second column: fuzzy segmentation with $A, \neg(A \cup \breve{A})$ and $\breve{A}$ (OPTION 1), third column: thresholding with $T = (a + \breve{a})/2$ (OPTION 2). The calculated values from top to bottom: $a = \{1, 1, 64, 62\}$ and $\breve{a} = \{143, 159, 127, 117\}$. The last column depicts the results by Otsu method for comparison [Source of images: Philips]

Acknowledgements - This work was supported by the *Natural Sciences and Engineering Research Council of Canada* (NSERC).

References

[1] Cruse, D.A., Lexical Semantics. Cambridge Univ. Press, 1986.

[2] H.R. Tizhoosh, *Opposition-based learning: a new scheme for machine intelligence*. International conference on computational intelligence for modeling, control and automation, Vienna, Austria, vol. I, pp. 695-701, 2005.

[3] H.R. Tizhoosh, *Opposition-Based Reinforcement Learning*, Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), vol. 10, no. 5, pp. 578-585, 2006.

[4] Tizhoosh, H.R. , Ventresca, M., *Oppositional Concepts in Computational Intelligence*. Springer, 2008.

[5] Ventresca, M., Tizhoosh, H.R., *A diversity maintaining population-based incremental learning algorithm*. Information Sciences 178(21), pp. 4038–405, 2008.

[6] Malisia, A., Tizhoosh, H.R., *Applying Opposition-Based Ideas to the Ant Colony system*. IEEE Symposium Series on Computational Intelligence (IEEE SSCI), pp. 182-189, 2007

[7] Montgomery, J., Randall, M., *Anti-pheromone as a tool for better exploration of search space*. In Dorigo et al. (Eds.) ANTS2002, LNCS 2463, Springer, pp. 100-110, 2002

[8] Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.A., *Opposition versus randomness in soft computing techniques*. Applied Soft Computing 8(2), pp. 906-918, 2008

[9] Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.A., *Opposition-Based Differential Evolution*. IEEE Transactions on Evolutionary Computation, vol. 12, issue 1, pp. 64–79, 2007

[10] Wang, R., Zhang, X., *Particle swarm optimization with opposite particles*., proceedings of 4th Mexican international conference on artificial intelligence, Monterrey, Mexico, November 14-18, 2005, pp. 633-640, 2005

[11] R.R. Yager, *Opposites and Measures of Extremism in Concepts and Constructs*. International Journal of Man-Machine Studies 18(4), pp. 307–322, 1983.

[12] A.R. de Soto, E. Trillas, *On Antonym and Negate in Fuzzy Logic*. International Journal of Intelligent Systems, vol. 14, 295-303, 1999.

[13] V. Novak, *Antonyms and linguistic quantifiers in fuzzy logic*. Fuzzy Sets and Systems 124, pp. 335–351, 2001.

ISBN: 978-989-95079-6-8

Genetic Cooperative-Competitive Fuzzy Rule Based Learning Method using Genetic Programming for Highly Imbalanced Data-Sets

Alberto Fernández[1] Francisco J. Berlanga[2] María J. del Jesus[3] Francisco Herrera[1]

1.Department of Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
2.Department of Computer Science and Systems Engineering, University of Zaragoza
Zaragoza, Spain
3.Department of Computer Science, University of Jaén
Jaén, Spain
Email: alberto@decsai.ugr.es, berlanga@unizar.es, mjjesus@ujaen.es, herrera@decsai.ugr.es

Abstract— *Classification in imbalanced domains is an important problem in Data Mining. We refer to imbalanced classification when data presents many examples from one class and few from the other class, and the less representative class is the one which has more interest from the point of view of the learning task. The aim of this work is to study the behaviour of the GP-COACH algorithm in the scenario of data-sets with high imbalance, analysing both the performance and the interpretability of the obtained fuzzy models. To develop the experimental study we will compare this approach with a well-known fuzzy rule learning algorithm, the Chi et al.'s method, and an algorithm of reference in the field of imbalanced data-sets, the C4.5 decision tree.*

Keywords— Fuzzy Rule-Based Classification Systems, Genetic Fuzzy Systems, Genetic Programming, Imbalanced Data-Sets, Interpretability

1 Introduction

In the area of Data Mining, real world classification problems present some features that can diminish the accuracy of Machine Learning algorithms, such as the presence of noise or missing values, or the imbalanced distribution of classes.

Specifically, the problem of imbalanced data-sets has been considered as one of the emergent challenges in Data Mining [1]. This situation occurs when one class is represented by a large number of examples (known as negative class), whereas the other is represented by only a few (positive class).

Our objective is to develop an empirical analysis in the context of imbalance classification for binary data-sets when the class imbalance ratio is high. In this study, we will make use of Fuzzy Rule Based Classification Systems (FRBCSs), a very useful tool in the ambit of Machine Learning, since they provide a very interpretable model for the end user [2].

We will employ a novel approach, GP-COACH (Genetic Programming-based evolutionary algorithm for the learning of COmpact and ACcurate FRBCS) [3], that learns disjunctive normal form (DNF) fuzzy rules (generated by means of a context-free grammar) and obtains very interpretable FRBCSs, with few rules and conditions per rule, with a high-generalization capability.

We want to analyse whether this model is accurate for data-sets with high imbalance in contrast with an FRBCS, the Chi et al.'s approach [4] and with C4.5 [5], a decision tree algorithm that has been used as a reference in the imbalanced data-sets field [6, 7]. We will also focus on the tradeoff between accuracy and interpretability [8] for the final obtained models. We will employ the Area Under the Curve (AUC) metric [9] to compute the classification performance, whereas we will measure the interpretability of the system by means of the number of rules in the system.

We have selected a large collection of data-sets with high imbalance from UCI repository [10] for developing our empirical analysis. In order to deal with the problem of imbalanced data-sets we will make use of a preprocessing technique, the "Synthetic Minority Over-sampling Technique" (SMOTE) [11], to balance the distribution of training examples in both classes. In this manner, we will analyse the positive synergy between the GP-COACH model and the SMOTE preprocessing technique for dealing with imbalanced data-sets. Furthermore, we will perform a statistical study using non-parametric tests [12, 13, 14] to find significant differences among the obtained results.

This contribution is organized as follows. First, Section 2 introduces the problem of imbalanced data-sets, describing its features, how to deal with this problem and the metric we have employed in this context. Next, in Section 3 we present the GP-COACH algorithm, explaining in detail the characteristics of this novel approach. Section 4 contains the experimental study for GP-COACH, Chi et al.'s and C4.5 algorithms regarding performance and interpretability. Finally, Section 5 summarizes and concludes the work.

2 Imbalanced Data-Sets in Classification

Learning from imbalanced data is an important topic that has recently appeared in the Machine Learning community [15, 16, 17]. The significance of this problem consists in its presence in most of the real domains of classification, such as fraud detection [18], risk management [19] and medical applications [20] among others.

This problem occurs when the number of instances of one class is much lower than the instances of the other classes. In this situation, the class of interest is often the one with the smaller number of examples, whereas the other class(es) represent(s) the counterpart of that concept and, in that manner, include(s) a high amount of data.

Standard classifier algorithms have a bias towards the majority class, since the rules that predicts the larger number of

ISBN: 978-989-95079-6-8

examples are positively weighted during the learning process in favour of the accuracy metric. Consequently, the instances that belongs to the minority class are misclassified more often than those belonging to the majority class [21]. Other important issue of this type of problem is the small disjuncts that can be found in the data-set [22] and the difficulty of most learning algorithms to detect those regions. Furthermore, the main handicap on imbalanced data-sets is the overlapping between the examples of the positive and the negative class [23]. These facts are depicted in Fig. 1.a and 1.b.

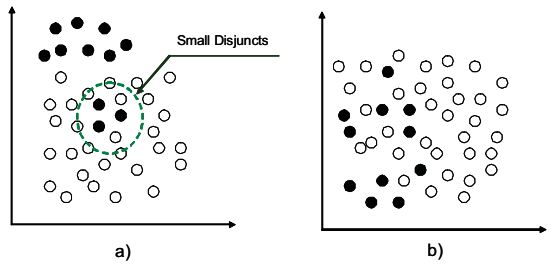


Figure 1: Example of the imbalance between classes: a) small disjuncts b) overlapping between classes

In this contribution we will focus on the data-sets with a higher degree of imbalance, using the imbalance ratio (IR) [24] as a threshold to categorize the different imbalanced scenarios. This measure is defined as the ratio of the number of instances of the majority class and the minority class.

In our previous work on this topic [25], we analysed the cooperation of some preprocessing methods with FRBCSs, showing a good behaviour for the oversampling methods, specially in the case of the SMOTE methodology [11]. According to this, we will employ in this contribution the SMOTE algorithm in order to deal with the problem of imbalanced data-sets.

In short, its main idea is to form new minority class examples by interpolating between several minority class examples that lie together. Thus, the overfitting problem is avoided and causes the decision boundaries for the minority class to spread further into the majority class space.

Regarding the empirical measure, instead of using accuracy, a more correct metric is considered. This is due to the fact that accuracy can lead to erroneous conclusions, since it does not take into account the proportion of examples for each class. Because of this, in this work we use the AUC metric [9], which can be defined as

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \tag{1}$$

where TP_{rate} is the percentage of positive cases correctly classified as belonging to the positive class and FP_{rate} is the percentage of negative cases misclassified as belonging to the positive class.

3 GP-COACH Algorithm

In this work we will make use of GP-COACH, a new FRBCS proposal [3]. The main features of this approach are listed below:

- It uses a context-free grammar that allows the learning of DNF fuzzy rules and the absence of some input features.
- It follows the *cooperative-competitive* approach, that is, it encodes a single rule per individual and the rule base (RB) is formed by the whole population. That makes necessary the use two different fitness functions in GP-COACH:
  - On the one hand, a local fitness function that evaluates the goodness of each one of the different rules in the population of individuals. From now on, we will refer it simply as *fitness function*.
  - On the other hand, a global fitness function that evaluates the goodness of a whole population of individuals (a rule set). From now on, we will refer it as *global fitness score*.

    This last fitness function has been introduced in GP-COACH in order to obtain the best rule set generated during the evolutionary process.
- It includes a mechanism to promote the diversity into the population, in order to avoid that all individuals converge to the same area of search space. Specifically, it uses the *Token Competition* diversity mechanism which makes rules compete among themselves during the evolutionary process, deleting irrelevant rules, and thus giving out a smaller number of rules that present a high-generalization capability.
- Finally, GP-COACH uses a two level hierarchical inference process because it learns rule sets containing two different types of rules: *primary rules*, which are strong and general rules generated by the genetic operators, and *secondary rules*, which are weaker and more specific rules, generated after the token competition procedure to increase the diversity in the population.

In the following subsections we will explain each one of these components and we will describe the way of working of this algorithm.

3.1 Context-free grammar for learning DNF fuzzy rules

GP-COACH learns DNF fuzzy rules:

$$R_k: \quad If\ X_1\ is\ \hat{A}_{k1}\ and\ \ldots\ and\ X_{n_v}\ is\ \hat{A}_{kn_v} \quad then\ Class\ is\ C_k\ with\ RW_k \tag{2}$$

where each input variable X_i takes as a value a set of linguistic terms or labels $\hat{A}_{ki} = \left\{ L_i^1\ or\ \ldots\ or\ L_i^{l_i} \right\}$ joined by a disjunctive operator, $(Class)$ is one of the class labels and RW_k is the rule weight [26]. We use triangular membership functions as antecedent fuzzy sets.

In GP-COACH, these DNF fuzzy rules are generated according to the production rules of a context-free grammar. In Table 1, an example of the grammar for a classification problem with two features (X_1, X_2), three linguistic labels per feature (*Low*, *Medium*, *High*) and three classes (C_1, C_2, C_3) is shown.

3.2 Evaluating an individual: Fitness function

Each one of the individuals in the population is evaluated according to a fitness function based on the estimation of:

ISBN: 978-989-95079-6-8

Table 1: Grammar example

$Start \longrightarrow [If],\ antec,\ [then],\ conseq,\ [.].$
$antec \longrightarrow descriptor1,\ [and],\ descriptor2.$
$descriptor1 \longrightarrow [any].$
$descriptor1 \longrightarrow [X_1\ is]\ label.$
$descriptor2 \longrightarrow [any].$
$descriptor2 \longrightarrow [X_2\ is]\ label.$
$label \longrightarrow \{member(?a,\ [L,\ M,\ H,\ L\ or\ M,\ L\ or\ H,$ $M\ or\ H,\ L\ or\ M\ or\ H])\},\ [?a].$
$conseq \longrightarrow [Class\ is]\ descriptorClass.$
$descriptorClass \longrightarrow \{member(?a,\ [C_1,\ C_2,\ C_3])\},\ [?a].$

- *Confidence*, which measures the accuracy of an individual, that is, the confidence of the consequent to be true if the antecedent is verified

$$Conf(R_k) = \frac{\mu_{tp}}{(\mu_{tp} + \mu_{fp})} \tag{3}$$

- *Support*, which measures the coverage of the knowledge represented in the individual

$$Supp(R_k) = \frac{\mu_{tp}}{N_{C_k}} \tag{4}$$

where μ_{tp} and μ_{fp} are the sums of the matching degrees for true and false positives, and N_{C_k} is the number of examples belonging to the class indicated in the consequent of the individual (R_k).

Both measures are combined to make up the fitness function in the following way

$$raw_fitness = (\alpha * Conf) + ((1 - \alpha) * Supp) \tag{5}$$

where α parameter allow us to give more importance to any of these both measures.

3.3 Evaluating a population: Global fitness score

Global fitness score measure is defined as follows:

$$Global_fitness = \begin{array}{l} (w_1 * AUC_{Tra}) + (w_2 * \overline{\#V}) + \\ (w_3 * \overline{\#C}) + (w_4 * \overline{\#R}) \end{array} \tag{6}$$

The global fitness score measure is formed by other four measures: a) AUC_{Tra}, the normalized value of the AUC metric for the training examples, b) *#V*, the normalized value of the number of variables per individual (rule) in the population, c) *#C*, the normalized value of the number of labels (or conditions) per individual, and d) *#R*, the normalized value of the number of rules in the population. We must remark that in the previous formula we employ the complement of these values.

Furthermore, we have included some weights (w_i) to give more importance to any of these four measures.

3.4 Token competition: Maintaining the diversity of the population

The idea of this mechanism is that each example in the training set can provide a resource called "token", for which all chromosomes in the population will compete to capture. If an individual (i.e. a rule) can match the example, it sets a flag to indicate that the token is seized preventing other weaker individuals to get this token.

The priority of receiving tokens is determined by the strength of the individuals. The individuals with higher fitness scores will exploit their niches by seizing as many tokens as they can. The other ones entering the same niches will have their strength decreased because they cannot compete with the stronger ones. This is done introducing a penalization in the fitness score of each individual. This penalization is based on the number of tokens that each individual has seized:

$$Penalized_fitness = raw_fitness * \frac{count}{ideal} \tag{7}$$

where *raw_fitness* is the fitness score obtained from the evaluation function, *count* is the number of tokens that the individual actually seized and *ideal* is the total number of tokens that it can take, which is equal to the number of examples that the individual matches.

As a result of the token competition, there exist individuals that cannot grab any token. These individuals are considered as irrelevant, and they can be eliminated from the population due to all of their examples are covered by other stronger individuals.

3.5 Secondary rules: Improving population diversity

Once the token competition mechanism has finished, it is possible that some training examples remain uncovered. The generation of new specific rules covering these examples improves the diversity in the population, and helps the evolutionary process to easily find stronger and more general rules covering these examples.

Therefore, GP-COACH learns rule sets having two different types of fuzzy rules: A core of strong and general rules (*primary rules*) that covers most of the examples, and a small set of weaker and more specific rules (*secondary rules*) that are only taken into account if there not exist any primary rule matching with some of the examples. This two level hierarchical inference process allows GP-COACH to obtain rule sets having a better interpretability-accuracy trade-off.

3.6 Genetic operators

GP-COACH makes use of four different genetic operators to generate new individuals during the evolutionary process:

1. *Crossover*: A part in the first parent is randomly selected and exchanged by another part, randomly selected, in the second one.
2. *Mutation*: It operates on label sets level. A variable in the rule is randomly chosen and then one of the next three different actions is carried out:
   (a) A new label is added to the label set.
   (b) A label is removed from the label set.
   (c) A label in the label set is exchanged by another one not included in it.
3. *Insertion*: It looks for all the variables in the rule and adds another different one that has not been included in it with a linguistic label set randomly chosen, although it must have at least one label and it must be different from the "any" set (see Table 1).
4. *Dropping Condition*: It randomly selects one variable in the rule and then turns it into "any". The label set associated with this variable is also removed.

We must remark that we generate only one child by using the genetic operators described above.

ISBN: 978-989-95079-6-8

3.7 Description of the Algorithm

GP-COACH algorithm begins creating a random initial population according to the rules in the context-free grammar. Each individual in this population is then evaluated. After that, the initial population is kept as the best evolved population and its global fitness score is calculated. Then, the initial population is copied to the current population and the evolutionary process begins:

1. An offspring population, with the same size than the current one, is created. Parents are selected by using the binary tournament selection mechanism and children are created by using one of the four genetic operators. The genetic operator selection is done in a probabilistic way. Specifically, the probabilities of the four genetic operator are added in a single measure $P = P_c + P_m + P_i + P_{dp}$ (where P_c, P_m, P_i and P_{dp} represent the crossover, mutation, insertion and dropping condition probabilities, respectively) and then a random value $u \in [0, P]$ is obtained to choose the genetic operator to be applied.

2. Once the offspring population is created, it is joined to the current population, creating a new population whose size is double the current population size. Individuals in this new population are sorted according to their fitness and Token Competition mechanism is applied. Secondaries rules are created if some examples remain uncovered.

3. Global fitness score measure is then calculated for this new population. We check whether this new fitness is better than the one stored for the best population, updating the best population and fitness if necessary. In any case, the new population is copied as the current population in order to be able to apply the evolutionary process again.

The evolutionary process ends when the stop condition is verified (i.e. number of evaluations). Then, the population kept as the best one is returned as a solution to the problem and GP-COACH finishes.

4 Experimental Study

In this study, our aim is to analyse the behaviour of the GP-COACH approach in the context of data-sets with high imbalance. We will compare the performance of this method against one state-of-the-art FRBCS algorithm, the Chi et al.'s approach, and the C4.5 decision tree, employing a large collection of imbalanced data-sets.

Specifically, we have considered twenty-two data-sets from UCI repository [10] with different IR, as shown in Table 2, where we denote the number of examples (#Ex.), number of attributes (#Atts.), class name of each class (minority and majority), class attribute distribution and IR. This table is in ascending order according to the IR. Data-sets with more than two classes have been modified by taking one against the others or by contrasting one class with another.

In order to reduce the effect of imbalance, we will employ the SMOTE preprocessing method [11] for all our experiments, considering only the 1-nearest neighbour to generate the synthetic samples, and balancing both classes to the 50% distribution.

In the remaining of this section, we will first present the experimental framework and all the parameters employed in this study and then we will show the results and all the statistical study for the GP-COACH approach.

4.1 Experimental Framework

To develop the different experiments we consider a *5-fold cross-validation model*, i.e., 5 random partitions of data with a 20%, and the combination of 4 of them (80%) as training and the remaining one as test. Since GP-COACH is a probabilistic method, we perform three executions per partition with different random seeds. For each data-set we consider the average results of the five partitions per three executions. Furthermore, Wilcoxon's Signed-Ranks Test [27] is used for statistical comparison of our empirical results. In all cases the level of confidence (α) will be set at 0.05.

The configuration for the FRBCSs approaches, GP-COACH and Chi et al.'s, is presented in Table 3. This parameter selection has been carried out according to the results achieved by GP-COACH and Chi et al.'s in our former studies in [3] and [25] respectively.

Table 3: Configuration for the FRBCS approaches

Parameter	GP-COACH	Chi et al.'s
Conjunction operator	Minimum T-norm	Product T-norm
Rule Weight	Certainty Factor	Penalized Certainty Factor
Fuzzy Reasoning Method	Additive Combination	Winning Rule
Number of Labels	5 Labels	5 Labels

The specific parameters setting for the GP-COACH algorithm, is listed below:

- Number of evaluations: 20000 evaluations.
- Initial Population Size: 200 individuals.
- α: 0.7.
- Crossover Probability P_c: 0.5.
- Mutation Probability P_m: 0.2.
- Dropping Probability P_{dp}: 0.15.
- Insertion Probability P_i: 0.15.
- Tournament Size: 2.
- w_1: 0.8, w_2 = w_3: 0.05, w_4: 0.1.

4.2 Analysis of the GP-COACH Behaviour in Data-sets with High Imbalance

The main objective in this study is to analyse the behaviour of the GP-COACH approach in the context of data-sets with high imbalance. According to this, Table 4 shows the results in performance (using the AUC metric) for GP-COACH and the algorithms employed for comparison, that is, Chi et al.'s learning method and C4.5.

We observe that the performance obtained by GP-COACH is higher than the one for Chi et al.'s and C4.5. This situation is represented statistically by means of a Wilcoxon test (Table 5) which shows a higher ranking in both cases for the GP-COACH algorithm.

Furthermore, when we compare the results for the fuzzy methods in each data-set, we observe that, as the IR increases, GP-COACH achieves a better performance than Chi et al.'s method. In this manner, we may state that GP-COACH is a very robust method according to the IR.

ISBN: 978-989-95079-6-8

Table 2: Summary Description for Imbalanced Data-Sets.

Data-set	#Ex.	#Atts.	Class (min.; maj.)	%Class(min., maj.)	IR
Data-sets with High Imbalance (IR higher than 9)					
Yeast2vs4	514	8	(cyt; me2)	(9.92, 90.08)	9.08
Yeast05679vs4	528	8	(me2; mit,me3,exc,vac,erl)	(9.66, 90.34)	9.35
Vowel0	988	13	(hid; remainder)	(9.01, 90.99)	10.10
Glass016vs2	192	9	(ve-win-float-proc; build-win-float-proc, build-win-non_float-proc,headlamps)	(8.89, 91.11)	10.29
Glass2	214	9	(Ve-win-float-proc; remainder)	(8.78, 91.22)	10.39
Ecoli4	336	7	(om; remainder)	(6.74, 93.26)	13.84
Yeast1vs7	459	8	(nuc; vac)	(6.72, 93.28)	13.87
Shuttle0vs4	1829	9	(Rad Flow; Bypass)	(6.72, 93.28)	13.87
Glass4	214	9	(containers; remainder)	(6.07, 93.93)	15.47
Page-blocks13vs2	472	10	(graphic; horiz.line,picture)	(5.93, 94.07)	15.85
Abalone9vs18	731	8	(18; 9)	(5.65, 94.25)	16.68
Glass016vs5	184	9	(tableware; build-win-float-proc, build-win-non_float-proc,headlamps)	(4.89, 95.11)	19.44
Shuttle2vs4	129	9	(Fpv Open; Bypass)	(4.65, 95.35)	20.5
Yeast1458vs7	693	8	(vac; nuc,me2,me3,pox)	(4.33, 95.67)	22.10
Glass5	214	9	(tableware; remainder)	(4.20, 95.80)	22.81
Yeast2vs8	482	8	(pox; cyt)	(4.15, 95.85)	23.10
Yeast4	1484	8	(me2; remainder)	(3.43, 96.57)	28.41
Yeast1289vs7	947	8	(vac; nuc,cyt,pox,erl)	(3.17, 96.83)	30.56
Yeast5	1484	8	(me1; remainder)	(2.96, 97.04)	32.78
Ecoli0137vs26	281	7	(pp,imL; cp,im,imU,imS)	(2.49, 97.51)	39.15
Yeast6	1484	8	(exc; remainder)	(2.49, 97.51)	39.15
Abalone19	4174	8	(19; remainder)	(0.77, 99.23)	128.87

Table 4: Detailed results in performance (AUC metric) for GP-COACH, Chi et al.'s learning method and C4.5

Dataset	GP-COACH		Chi et al.'s		C4.5	
	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}
Yeast2vs4	83.46 ± 2.01	78.26 ± 5.92	90.51 ± 1.43	86.85 ± 6.68	98.14 ± 0.88	85.88 ± 8.78
Yeast05679vs4	82.76 ± 1.25	79.92 ± 5.54	87.97 ± 0.65	76.42 ± 6.17	95.26 ± 0.94	76.02 ± 9.36
Vowel0	96.74 ± 0.76	92.20 ± 5.80	99.64 ± 0.19	97.89 ± 1.83	99.67 ± 0.48	94.94 ± 4.95
Glass016vs2	74.57 ± 2.98	59.35 ± 17.29	76.16 ± 2.11	60.02 ± 8.41	97.16 ± 1.86	60.62 ± 12.66
Glass2	78.45 ± 2.82	66.11 ± 13.76	75.50 ± 1.80	52.06 ± 11.20	95.71 ± 1.51	54.24 ± 14.01
Ecoli4	97.74 ± 0.66	91.91 ± 6.04	98.14 ± 0.65	92.30 ± 8.13	97.69 ± 1.96	83.10 ± 9.90
Yeast1vs7	76.44 ± 2.17	64.03 ± 8.50	84.08 ± 2.14	65.24 ± 10.47	93.51 ± 2.20	70.03 ± 1.46
Shuttle0vs4	99.88 ± 0.08	99.99 ± 0.04	100.0 ± 0.00	98.72 ± 1.17	99.99 ± 0.02	99.97 ± 0.07
Glass4	95.94 ± 1.87	86.94 ± 14.19	98.88 ± 0.56	82.85 ± 10.20	98.44 ± 2.29	85.08 ± 9.35
Page-Blocks13vs4	98.03 ± 0.82	96.39 ± 4.31	98.71 ± 0.23	93.41 ± 8.53	99.75 ± 0.21	99.55 ± 0.47
Abalone9vs18	77.91 ± 2.31	74.27 ± 7.15	71.22 ± 3.09	67.44 ± 9.88	95.31 ± 4.44	62.15 ± 4.96
Glass016vs5	95.83 ± 1.45	94.29 ± 8.21	98.43 ± 0.41	84.86 ± 21.91	99.21 ± 0.47	81.29 ± 24.44
Shuttle2vs4	97.36 ± 3.34	97.43 ± 3.78	100.0 ± 0.00	88.38 ± 21.60	99.90 ± 0.23	99.17 ± 1.86
Yeast1458vs7	66.36 ± 1.61	58.21 ± 8.47	81.83 ± 1.70	59.32 ± 7.68	91.58 ± 2.78	53.67 ± 2.09
Glass5	98.11 ± 1.01	78.05 ± 24.24	98.78 ± 0.48	74.63 ± 20.52	99.76 ± 0.40	88.29 ± 13.31
Yeast2vs8	80.97 ± 3.28	78.77 ± 9.37	83.46 ± 1.68	80.66 ± 6.94	91.25 ± 1.84	80.66 ± 11.22
Yeast4	84.88 ± 1.42	81.95 ± 4.08	87.96 ± 1.54	83.25 ± 2.39	91.01 ± 2.64	70.04 ± 5.65
Yeast1289vs7	72.17 ± 2.75	66.26 ± 10.42	80.03 ± 2.33	70.27 ± 3.75	94.65 ± 1.13	68.32 ± 6.16
Yeast5	95.78 ± 0.68	93.53 ± 2.74	95.43 ± 0.54	93.72 ± 2.72	97.77 ± 1.45	92.33 ± 4.72
Ecoli0137vs26	90.13 ± 2.42	81.00 ± 18.30	96.85 ± 1.59	68.80 ± 22.87	96.78 ± 3.28	81.36 ± 21.68
Yeast6	90.43 ± 1.68	86.67 ± 7.92	89.60 ± 2.00	88.20 ± 8.55	92.42 ± 3.54	82.80 ± 12.77
Abalone19	74.00 ± 3.70	68.45 ± 8.38	77.19 ± 2.49	67.48 ± 10.77	85.44 ± 2.49	52.02 ± 4.41
Mean	86.72 ± 1.87	**80.64 ± 8.84**	89.56 ± 1.25	78.76 ± 9.65	95.93 ± 1.68	78.25 ± 8.38

Table 5: Wilcoxon test to compare GP-COACH with Chi et al.'s approach and C4.5 according to their performance. R^+ corresponds to GP-COACH and R^- to Chi or C4.5

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	p-value
GP-COACH vs. Chi	161.0	92.0	Not Rejected	0.263
GP-COACH vs. C4.5	162.0	91.0	Not Rejected	0.249

Regarding interpretability of the obtained models, we must stress that GP-COACH is designed to obtain few DNF rules to describe the concept accurately. Table 6 shows the average number of rules for the three algorithms considered in this study, together with the associated standard deviation. The results from this table clearly shows that GP-COACH is the most interpretable model.

In Table 7 we show an example of an RB generated with the GP-COACH algorithm for the "shuttle2vs4" data-set. We can observe that the problem is described using few rules and only three of nine variables, enhancing the readability for the end-user.

5 Conclusions

In this contribution, we have studied the behaviour of GP-COACH in the context of data-sets with high imbalance.

Our results have shown the good performance achieved by this approach in contrast with the Chi et al.'s method, a well-known fuzzy rule learning algorithm, and C4.5, an algorithm of reference in the area of imbalanced data-sets.

Furthermore, we have compared the interpretability of the obtained models by means of the size of the rule set, concluding that GP-COACH employs a more compact RB in comparison with Chi et al.'s and C4.5 algorithms.

We must remark that GP-COACH is a good methodology for imbalanced data-sets, since it obtains very good classification results according to the AUC measure, by using a very interpretable model with few linguistic fuzzy rules.

Acknowledgment

This work had been supported by the Spanish Ministry of Science and Technology under Projects TIN2008-06681-C06-

01 and TIN2008-06681-C06-02.

Table 6: Detailed results table in interpretability (number of rules) for GP-COACH, Chi et al.'s learning method and C4.5

Data-set	GP-COACH	Chi et al.'s	C4.5
Yeast2vs4	6.60 ± 0.74	164.80 ± 3.83	20.40 ± 4.56
Yeast05679vs4	9.87 ± 2.17	191.20 ± 9.12	30.00 ± 6.63
Vowel0	5.60 ± 1.12	694.60 ± 8.32	11.80 ± 2.17
Glass016vs2	6.60 ± 2.23	65.20 ± 5.67	15.20 ± 1.48
Glass2	5.53 ± 1.96	73.20 ± 2.28	16.80 ± 2.49
Ecoli4	5.67 ± 0.90	116.80 ± 6.42	9.20 ± 2.17
Yeast1vs7	8.80 ± 2.01	156.80 ± 6.46	33.20 ± 6.10
Shuttle0vs4	2.60 ± 0.63	79.20 ± 7.56	2.80 ± 1.10
Glass4	6.33 ± 1.59	98.40 ± 10.06	7.40 ± 2.88
Page-Blocks13vs4	5.07 ± 0.96	164.00 ± 7.78	7.00 ± 1.22
Abalone9vs18	9.20 ± 2.24	93.60 ± 5.03	47.60 ± 1.14
Glass016vs5	5.13 ± 0.74	99.60 ± 8.62	10.00 ± 2.83
Shuttle2vs4	3.60 ± 0.99	33.00 ± 4.95	4.00 ± 0.00
Yeast1289vs7	8.93 ± 1.91	160.80 ± 2.68	58.40 ± 6.35
Glass5	4.93 ± 0.70	90.80 ± 6.87	7.00 ± 1.00
Yeast2vs8	5.73 ± 1.53	110.00 ± 2.65	14.60 ± 2.07
Yeast4	8.07 ± 1.22	197.20 ± 7.01	40.40 ± 3.78
Yeast1458vs7	7.27 ± 1.22	181.00 ± 8.69	47.80 ± 6.38
Yeast5	3.47 ± 0.52	206.60 ± 5.32	11.60 ± 2.07
Ecoli0137vs26	5.13 ± 0.74	168.60 ± 5.41	8.00 ± 1.41
Yeast6	5.73 ± 1.49	198.80 ± 6.14	21.60 ± 6.47
Abalone19	6.80 ± 1.15	180.20 ± 7.40	69.20 ± 3.70
Mean	**6.21 ± 1.31**	160.20 ± 6.29	22.45 ± 3.09

Table 7: Example of a DNF RB extracted using GP-COACH for the shuttle2vs4 data-set

Rule1: IF X_7 is $(L_4|L_5)$ THEN Class = negative with RW = 1
Rule2: IF X_1 is $(L_3|L_4|L_5)$ AND X_7 is $(L_1|L_2)$
THEN Class = positive with RW = 0.934089
Rule3: IF X_7 is $(L_2|L_4)$ AND X_8 is L_4
THEN Class = positive with RW = 0.495765
Rule4: IF X_7 is $(L_3|L_5)$ THEN Class = positive
with RW = 0.562457

References

[1] Q. Yang and X. Wu. 10 challenging problems in data mining research. *International Journal of Information Technology and Decision Making*, 5(4):597–604, 2006.

[2] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer-Verlag, 2004.

[3] F.J. Berlanga, M.J. del Jesus, and F. Herrera. A novel genetic cooperative-competitive fuzzy rule based learning method using genetic programming for high dimensional problems. In *3rd International Workshop on Genetic and Evolving Fuzzy Systems (GEFS08)*, pages 101–106, 2008.

[4] Z. Chi, H. Yan, and T. Pham. *Fuzzy algorithms with applications to image processing and pattern recognition*. World Scientific, 1996.

[5] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo-California, 1993.

[6] A. Estabrooks, T. Jo, and N. Japkowicz. A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence*, 20(1):18–36, 2004.

[7] C.T. Su and Y.H. Hsiao. An evaluation of the robustness of MTS for imbalanced data. *IEEE Transactions on Knowledge Data Engineering*, 19(10):1321–1332, 2007.

[8] J. Casillas, F. Herrera, R. Pérez, M.J. del Jesus, and Pedro Villar. Special issue on genetic fuzzy systems and the interpretability-accuracy trade-off - editorial. *International Journal of Approximate Reasoning*, 44(1):1–3, 2007.

[9] J. Huang and C. X. Ling. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3):299–310, 2005.

[10] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007. University of California, Irvine, School of Information and Computer Sciences. URL `http://www.ics.uci.edu/~mlearn/MLRepository.html`

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligent Research*, 16:321–357, 2002.

[12] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[13] S. García and F. Herrera An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. *Journal of Machine Learning Research*, 9:2607–2624, 2008.

[14] S. García, A. Fernández, J. Luengo and F. Herrera A Study of Statistical Techniques and Performance Measures for Genetics-Based Machine Learning: Accuracy and Interpretability. *Soft Computing*, 13:10 959–977, 2009.

[15] N. V. Chawla, N. Japkowicz, and A. Kolcz. Editorial: special issue on learning from imbalanced data sets. *SIGKDD Explorations*, 6(1):1–6, 2004.

[16] R. Barandela, J.S. Sánchez, V. García, and E. Rangel. Strategies for learning in class imbalance problems. *Pattern Recognition*, 36(3):849–851, 2003.

[17] M.C. Chen, L.S. Chen, C.C. Hsu, and W.R. Zeng. An information granulation based data mining approach for classifying imbalanced data. *Information Sciences*, 178(16):3214–3227, 2008.

[18] T. Fawcett and F. J. Provost. Adaptive fraud detection. *Data Mining and Knowledge Discovery*, 1(3):291–316, 1997.

[19] Y. M. Huang, C. M. Hung, and H. C. Jiau. Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem. *Nonlinear Analysis: Real World Applications*, 7(4):720–747, 2006.

[20] M.A. Mazurowski, P.A. Habas, J.M. Zurada, J.Y. Lo, J.A. Baker, and G.D. Tourassi. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*, 21(2-3):427–436, 2008.

[21] G. M. Weiss. Mining with rarity: a unifying framework. *SIGKDD Explorations*, 6(1):7–19, 2004.

[22] G.M. Weiss and F. Provost. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, 19:315–354, 2003.

[23] V. García, R.A. Mollineda, and J. S. Sánchez. On the k-NN performance in a challenging scenario of imbalance and overlapping. *Pattern Analysis Applications*, 11(3–4):269–280, 2008.

[24] A. Orriols-Puig and E. Bernadó-Mansilla. Evolutionary rule-based systems for imbalanced datasets. *Soft Computing*, 13(3):213–225, 2009.

[25] A. Fernández, S. García, M.J. del Jesus, and F. Herrera. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems*, 159(18):2378–2398, 2008.

[26] H. Ishibuchi and T. Yamamoto, Rule Weight Specification in Fuzzy Rule-Based Classification Systems. *IEEE Transactions on Fuzzy Systems*, 13(4):428-435, 2005.

[27] D. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC, second edition, 2006.

ISBN: 978-989-95079-6-8

New Fuzzy Color Clustering Algorithm Based on *hsl* Similarity

Vasile Pătraşcu

Department of Informatics Technology, Tarom Company
Bucharest, Romania
Email: patrascu.v@gmail.com

Abstract—*In this paper, one presents a fuzzy color clustering algorithm that is based on a new measure of similarity. This new measure of color similarity is defined on a perceptual color system called hsl.*

Keywords— Color similarity, fuzzy color clustering, *hsl* perceptual color system.

1 Introduction

The color clustering algorithms play an important role in image analysis. The obtained results are very dependent on the coordinate system and similarity/dissimilarity functions used for color separation [1], [4], [5] and [6]. In this paper, one proposes the use of a perceptual system hsl. In the framework of this system, a new measure is proposed for the color similarity. In the following sections, the paper is thus organized: section 2 presents the system hsl for color representation based on new measures of luminosity L and saturation S; section 3 presents distances and similarities for each component of the hsl system; section 4 presents the new color similarity; section 5 presents the color clustering algorithm based on the proposed similarity; section 6 presents the experimental results and conclusions are in section 7.

2 The perceptual color system *hsl*

The most part of the color images are represented by the RGB color system. Image analysis is done in the perceptual coordinate systems. One of them is the hSL system where h is the hue and an angular value; S is the saturation and lastly, L is the luminosity. In this paper, we will use the following formulae for the definition of the hSL components:

$$\begin{cases} L = \dfrac{M}{1+M-m} \\ S = \dfrac{2\cdot(M-m)}{1+|M-0.5|+|m-0.5|} \\ h = \text{atan2}\left(\dfrac{B-G}{\sqrt{2}}, \dfrac{2R-B-G}{\sqrt{6}}\right) \end{cases} \tag{1}$$

where:

$$M = \max(R,G,B)$$
$$m = \min(R,G,B)$$

We suppose that $R,G,B \in [0,1]$. According to the considered formulae for the definition of the components hSL, it results that $L,S \in [0,1]$ and $h \in (-\pi,\pi]$. In order to have a unitary consideration for the three color components, hSL, we will scale the luminosity L and the saturation S from $[0,1]$ to $\left[0,\dfrac{\pi}{2}\right]$ by formulae:

$$\begin{cases} l = \dfrac{\pi}{2}\cdot L \\ s = \dfrac{\pi}{2}\cdot S \end{cases} \tag{2}$$

Thus, any color q from RGB space will have the following representation in the hsl space:

$$\begin{cases} l = \dfrac{\pi}{2}\cdot\dfrac{M}{1+M-m} \\ s = \pi\cdot\dfrac{M-m}{1+|M-0.5|+|m-0.5|} \\ h = \text{atan2}\left(\dfrac{B-G}{\sqrt{2}}, \dfrac{2R-B-G}{\sqrt{6}}\right) \end{cases} \tag{3}$$

3 Distances and similarities for the values *hsl*

We know that for two angular values α,β a good measure for distance is based on the *sin* function, namely:

$$d(\alpha,\beta) = \left|\sin\left(\frac{\alpha-\beta}{2}\right)\right| \tag{4}$$

Thus, for the hue, the distance is given by the function:

$$d_h(h_1,h_2) = \left|\sin\left(\frac{h_1-h_2}{2}\right)\right| \tag{5}$$

Using the formula (4) for the saturation s and luminosity l we obtain their values in the interval $\left[0,\dfrac{1}{\sqrt{2}}\right]$, because $l,s \in \left[0,\dfrac{\pi}{2}\right]$. Consequently, we will multiply the formula (4) with the factor $\sqrt{2}$ and thus we will obtain the values in

ISBN: 978-989-95079-6-8

the whole interval $[0,1]$. In this way, the distances for luminosity and saturation will be defined by:

$$d_l(l_1,l_2)=\sqrt{2}\cdot\left|\sin\left(\frac{l_1-l_2}{2}\right)\right| \tag{6}$$

$$d_s(s_1,s_2)=\sqrt{2}\cdot\left|\sin\left(\frac{s_1-s_2}{2}\right)\right| \tag{7}$$

In this paper, we will take into account that the square of distance is a good measure of dissimilarity and the negation of dissimilarity is a good measure of similarity. In other words, if d is the distance and σ is the similarity, then the following relation exists between them:

$$\sigma=1-d^2 \tag{8}$$

Having the distance definitions, we will then define the similarity function for luminosity and saturation,

$$\begin{cases}\sigma_l(l_1,l_2)=1-d_l^2(l_1,l_2)=\cos(l_1-l_2)\\ \sigma_s(s_1,s_2)=1-d_s^2(s_1,s_2)=\cos(s_1-s_2)\end{cases} \tag{9}$$

and next for the hue:

$$\sigma_h(h_1,h_2)=1-d_h^2(h_1,h_2)=\cos^2\left(\frac{h_1-h_2}{2}\right) \tag{10}$$

4 The *hsl* color similarity

We will initially define the chromaticity c and the achromaticity a for a color having as saturation the value s:

$$\begin{cases}c=\sin(s)\\ a=\cos(s)\end{cases} \tag{11}$$

The following relation between these two defined parameters is obvious:

$$a^2+c^2=1 \tag{12}$$

We will consider two colors q_1,q_2 defined by parameters $(h_1,s_1,l_1),(h_2,s_2,l_2)$. We will add to these parameters the chromaticity coefficients c_1,c_2 and achromaticity coefficients a_1,a_2, computed by formulae (11). Now, we define the similarity between two colors by formula:

$$\sigma(q_1,q_2)=c_1c_2\cdot\sigma_h(h_1,h_2)+a_1a_2\cdot\sigma_l(l_1,l_2) \tag{13}$$

Seeing the formula (13), one remarks that this new color similarity measure is a linear and adaptive combination, between the hues similarity and luminosities similarity. Thus for two chromatic colors, the main component is given by the hues similarity and for two achromatic colors the main component is given by the luminosities similarity. We can see that, when a color is chromatic and the other is achromatic, the similarity has small values. This is nothing else than the two considered colors are quite non-similar. Also, we can see that in formula (13), the saturations similarity does not appear directly, but it exists somewhere in background. The formula (13) has the following equivalent form:

$$\begin{aligned}\sigma(q_1,q_2)=\sin(s_1)\cdot\sin(s_2)\cdot\cos^2\left(\frac{h_1-h_2}{2}\right)+\\ +\cos(s_1)\cdot\cos(s_2)\cdot\cos(l_1-l_2)\end{aligned} \tag{14}$$

Taking into account that:

$$\begin{cases}\cos^2\left(\frac{h_1-h_2}{2}\right)\le 1\\ \cos(l_1-l_2)\le 1\end{cases} \tag{15}$$

it results from (14) this inequality:

$$\sigma(q_1,q_2)\le\sin(s_1)\cdot\sin(s_2)+\cos(s_1)\cdot\cos(s_2)$$

namely

$$\sigma(q_1,q_2)\le\cos(s_1-s_2)$$

and finally we obtain that:

$$\sigma(q_1,q_2)\le\sigma_s(s_1,s_2) \tag{16}$$

The formula (16) shows that if the saturations of the two colors have a low similarity then the two colors have a low similarity too and if the two colors are strongly similar then their saturations are strongly similar as well. Also, from (15) and (16) it results the following implications:

$$\sigma(q_1,q_2)=1\ \Rightarrow\ \begin{cases}\sigma_h(h_1,h_2)=1\\ \sigma_l(l_1,l_2)=1\\ \sigma_s(s_1,s_2)=1\end{cases}\ \Rightarrow\ \begin{cases}h_1=h_2\\ l_1=l_2\\ s_1=s_2\end{cases}$$

Using the formula (8) we can define the distance between two colors in the *hsl* space as:

$$d(q_1,q_2)=\sqrt{1-\sigma(q_1,q_2)} \tag{17}$$

We can state that formula (17) defines a metric because it verifies the following three metric properties:

- $d(q_1,q_2)=0\ \Leftrightarrow\ q_1=q_2$
- $d(q_1,q_2)=d(q_2,q_1)$
- $d(q_1,q_2)+d(q_2,q_3)\ge d(q_1,q_3)$

A detailed proof of these three properties is not a subject of this paper and thus not considered here.

ISBN: 978-989-95079-6-8

5 The fuzzy color clustering algorithm

We consider n colors $q_1, q_2, ..., q_n$ that must be separated into k clusters. Each cluster i is characterized by the membership coefficients $w_{i1}, w_{i2}, ..., w_{in}$ for the considered n colors and the cluster center defined by the color $\bar{q}_i = (\bar{h}_i, \bar{s}_i, \bar{l}_i)$. For color clustering we will construct an algorithm that is similar to the fuzzy c-means algorithm [1]. We will consider the following objective function:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij}^{\alpha} \cdot \sigma(\bar{q}_i, q_j) \tag{18}$$

where α is a fuzzification-defuzzification parameter and also, $\alpha \in (0,1)$. If α approaches 1, then the fuzzy algorithm approaches a crisp one. The objective function J (18) must be maximized. In order to have fuzzy partitions, we must add the following conditions, for $j = 1,2,...,n$:

$$w_{1j} + w_{2j} + ... + w_{kj} = 1 \tag{19}$$

In this case, considering for the condition (19) the Lagrange multipliers $\lambda_1, \lambda_2, ..., \lambda_n$, the objective function (18) becomes:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij}^{\alpha} \cdot \sigma(\bar{q}_i, q_j) + \sum_{j=1}^{n} \lambda_j \cdot \left(\sum_{i=1}^{k} w_{ij} - 1 \right) \tag{20}$$

The maximum value of the objective function J (20) results from the following conditions:

- $\forall i \in [1,k], \forall j \in [1,n]$,

$$\frac{\partial J}{\partial w_{ij}} = 0 \tag{21}$$

It results:

$$w_{ij} = \frac{\left(\sigma(\bar{q}_i, q_j)\right)^{\frac{1}{1-\alpha}}}{\sum_{m=1}^{k} \left(\sigma(\bar{q}_m, q_j)\right)^{\frac{1}{1-\alpha}}} \tag{22}$$

- $\forall i \in [1,k]$,

$$\frac{\partial J}{\partial \bar{h}_i} = 0 \tag{23}$$

It results:

$$\bar{h}_i = \operatorname{atan2}\left(\frac{\sum_{j=1}^{n} u_{ij} \sin(h_j)}{\sum_{j=1}^{n} u_{ij}}, \frac{\sum_{j=1}^{n} u_{ij} \cos(h_j)}{\sum_{j=1}^{n} u_{ij}} \right) \tag{24}$$

where

$$u_{ij} = w_{ij}^{\alpha} \cdot c_j \tag{25}$$

- $\forall i \in [1,k]$,

$$\frac{\partial J}{\partial \bar{l}_i} = 0 \tag{26}$$

It results:

$$\bar{l}_i = \operatorname{atan2}\left(\frac{\sum_{j=1}^{n} v_{ij} \sin(l_j)}{\sum_{j=1}^{n} v_{ij}}, \frac{\sum_{j=1}^{n} v_{ij} \cos(l_j)}{\sum_{j=1}^{n} v_{ij}} \right) \tag{27}$$

where

$$v_{ij} = w_{ij}^{\alpha} \cdot a_j \tag{28}$$

- $\forall i \in [1,k]$,

$$\frac{\partial J}{\partial \bar{s}_i} = 0 \tag{29}$$

It results:

$$\bar{s}_i = \operatorname{atan2}\left(\frac{\sum_{j=1}^{n} x_{ij} \sin(s_j)}{\sum_{j=1}^{n} w_{ij}^{\alpha}}, \frac{\sum_{j=1}^{n} y_{ij} \cos(s_j)}{\sum_{j=1}^{n} w_{ij}^{\alpha}} \right) \tag{30}$$

where

$$\begin{cases} x_{ij} = w_{ij}^{\alpha} \cdot \sigma_h(\bar{h}_i, h_j) \\ y_{ij} = w_{ij}^{\alpha} \cdot \sigma_l(\bar{l}_i, l_j) \end{cases} \tag{31}$$

6 Experimental results

The proposed algorithm was applied to the image „flower” (Fig. 1) and the clustered image can be seen in Fig. 2. For comparison, the fuzzy c-means algorithm [1], [3], [8] was applied to image „flower”, in the coordinate systems RGB [3], [9], Lab [7], Luv [3], [9], hSL [9], I1I2I3 [2], [3] and the results can be seen in Figs. 3, 4, 5, 6 and 7.



Figure 1: The image „flower”.

ISBN: 978-989-95079-6-8

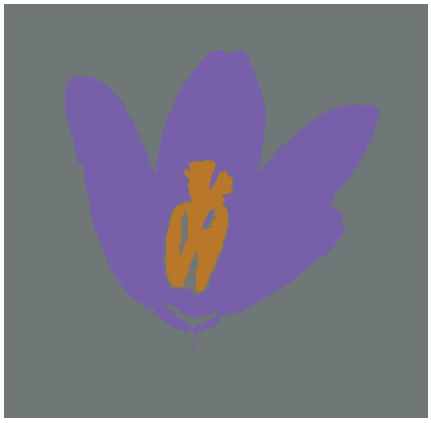


Figure 2: The image „flower” clustered with the proposed algorithm and *hsl*.

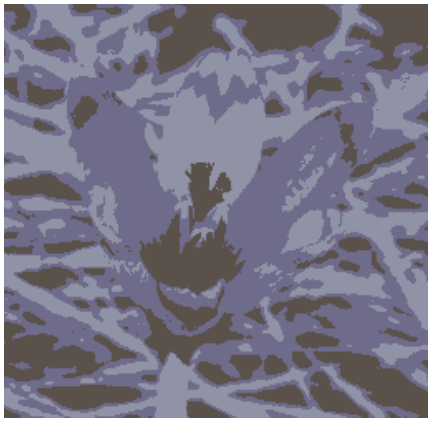


Figure 3: The image „flower” clustered with FCM algorithm and *RGB*.



Figure 4: The image „flower” clustered with FCM algorithm and *Lab*.



Figure 5: The image „flower” clustered with FCM algorithm and *Luv*.



Figure 6: The image „flower” clustered with FCM algorithm and *hSL*.



Figure 7: The image „flower” clustered with FCM algorithm and *I1I2I3*.

7 Conclusions

In this paper, a new measure for color similarity was defined. This new measure was defined in a new perceptual system *hsl* . Using this measure, an algorithm for color clustering was constructed.

In the section dedicated to the experimental results, one can see the advantage of using this new color similarity measure.

References

[1] J.C. Bezdek. Pattern Recognition with Fuzzy Objective Functions. New York: Plenum Press, 1981.

[2] Y.I. Ohta, T. Kanade and T. Sakai. Color information for region segmentation, *Computer and Image Processing*, vol. 13, pp.222-241, 1980.

[3] Y.W. Lim and S.U. Lee. On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques, *Pattern Recognition*, vol. 23, no.9 , pp.935-952, 1990.

ISBN: 978-989-95079-6-8

[4] V. Pătraşcu. A generalization of fuzzy c-means algorithm using a new dissimilarity function, The 11th International Fuzzy Systems Association World Congres, IFSA 2005, *Fuzzy Logic, Soft Computing and Computational Intelligence*, Vol II, pp. 591-596, Beijing China July, 28-31,2005.

[5] V. Pătraşcu. A Generalization of Gustafson-Kessel Algorithm Using a New Constraint Parameter*, In Proceedings of the 4th Conference of the European Society for Fuzzy Logic and Technology and 11th Rencontres Francophones sur la Logique Floue et ses Applications*, EUSFLAT-LFA 2005, pp.1250-1255, Barcelona, Spain 7-9 September, 2005.

[6] V. Pătraşcu. Fuzzy Image Segmentation Based on Triangular Function and Its n-dimensional Extension, in the volume *Soft Computing in Image Processing. Recent Advances, Series: Studies in Fuzziness and Soft Computing*, pp 187-208, Vol. 210 , (Eds.) M. Nachtegael, D. Van der Weken, E. E. Kerre, W. Philips, ISBN: 3-540-38232-1, Springer-Verlag, 2007.

[7] S. Tominga. A colour classification method for color images using a uniform color space, *Proc. 10th Int. Conf. On Pattern Recognition*, vol. I, Atlantic City, new Jersey, pp. 803-807, 16-21 June 1990.

[8] M. Trivedi and J.C. Bezdek. Low-level segmentation of aerial images with fuzzy clustering, *IEEE Trans. On Systems, Man, and Cybernetics*, vol. 16, no. 4, pp.589-598, 1986.

[9] S.E. Umbaugh, R.H. Moss, W.V. Stoecker and G.A. Hance. Automatic colour segmentation algorithms with application to skin tumor feature identification, IEEE Engineering in Medcine and Biology, vol. 12, no. 3, pp. 75-82, 1993.

ISBN: 978-989-95079-6-8

Multi-Dimensional Fuzzy Transforms for Attribute Dependencies

Ferdinando Di Martino[1,2] Vincenzo Loia[2] Salvatore Sessa[1]

[1]Dipartimento di Costruzioni e Metodi Matematici in Architettura, Università degli Studi di Napoli Federico II, 80134 Napoli, Italy
[2] Dipartimento di Matematica e Informatica, Università degli Studi di Salerno, 84084 Salerno, Italy
Email: fdimarti@unina.it, loia@unisa.it, sessa@unina.it

Abstract *— We explore attribute dependencies in the datasets by using direct and inverse fuzzy transforms. Our algorithm optimizes the fuzzy partitions of the universe of the attributes and moreover establishes if the set of the data points is sufficiently dense with respect to the chosen partitions: two specific regression indexes measure the reliability of our model. The known "El Nino" dataset is the basis of our experiments, whose results are consistent with the regression analysis made with the same data.*

Keywords — attribute dependence, fuzzy transform, index of determinacy, regression.

1 Introduction

Regression is a well known statistical supervised technique used in data mining (cfr. [5, 6, 7, 8, 9]. The dependency between numerical attributes is studied via an equation of regression: the dependent attribute is modeled as a function of the independent attributes or predictors. We present a fast regression algorithm in which we use the fuzzy transforms (for short, F-transforms) for exploring numerical attribute dependencies in datasets. A *direct fuzzy transform* [1, 11, 12, 13, 14] gives a correspondence between a set of continuous functions defined on the interval [a, b] and a set of n-dimensional vectors defined suitably. An *inverse fuzzy transform* gives the converse correspondence, that is an n-dimensional vector is transformed into a continuous function which approximates the original function up to a small quantity ε. The method based on the F-transforms has been used in other topics like image processing [2, 3, 4, 12], geology [11].
In [10] the attribute dependencies in data analysis are established via F-transforms: the dependency of an attribute X_z from the attributes $X_1,\ldots,X_k$ is made by setting $X_z = H(X_1,\ldots,X_k)$, where H: $[a_1,b_1] \times\ldots\times [a_k,b_k] \rightarrow [a_z,b_z]$ is a continuous function and $[a_i,b_i]$ is the domain of X_i, i= 1,…,k. The approximating function $H_{F,n}$ is obtained via the *discrete F-transforms* and the difference between H and $H_{F,n}$ at the their common points is evaluated by means of the index of determinacy. We devote particular attention to the solution of the following problems:

a) to find the best fuzzy partition of the attribute domains in the construction of the direct F-transform. We observe that the index of determinacy calculated with coarse grained partitions is less than that one calculated with finer partitions;

b) our method analyzes if the set of the assigned data points is sufficiently dense w. r. t. the chosen partitions.
Our algorithm makes a fuzzy partition of $[a_i,b_i]$, i=1,…,k, in n fuzzy sets and establishes a threshold value for the indexes of determinacy after some training tests. The attribute dependency is found if the index overcomes the threshold value, otherwise we pass to a partition of $[a_i,b_i]$ in n+1 fuzzy sets by checking additionally if the set of the assigned data points is sufficiently dense w.r.t. the chosen partition. Of course if this condition is violated, then the process is stopped.
In Section 2 we give the definitions of F-transforms in k (≥2) variables. In Section 3 we present our process, in Section 4 a sample simulation shows how to find the threshold value for the indexes of determinacy. The known dataset "El Nino" is the basis of our tests contained in Section 5. Section 6 concludes this paper.

2 Multi-dimensional F-transforms

In accordance to [12], let $n \geq 2$ and $x_1, x_2, \ldots, x_n$ be points of [a,b], called nodes, such that $x_1 = a < x_2 <\ldots< x_n = b$. The fuzzy sets $A_1,\ldots,A_n : [a,b] \rightarrow [0,1]$, called basic functions, form a *fuzzy partition* of [a,b] if the following hold:

(1) $A_i(x_i) =1$ for every i =1,2,…,n;

(2) $A_i(x) = 0$ if $x \notin (x_{i-1},x_{i+1})$ for every i = 2,…,n;

(3) $A_i(x)$ is a continuous function on [a,b] for every i = 1,2,…,n;

(4) $A_i(x)$ strictly increases on $[x_{i-1}, x_i]$ for every i = 2, …, n and strictly decreases on $[x_i,x_{i+1}]$ for every i = 1,…, n-1;

(5) $\sum_{i=1}^{n} A_i(x)=1$ for every $x \in [a,b]$.

The partition $\{A_1(x),\ldots,A_n(x)\}$ is said *uniform* if moreover the following hold:
(6) $n \geq 3$ and the nodes are equidistant, that is $x_i = a+ h\cdot(i-1)$ for every i = 1, 2, …, n where h = (b-a)/(n-1);

(7) $A_i(x_i – x)=A_i(x_i + x)$ for every $x \in [0,h]$ and i = 2,…, n-1;

ISBN: 978-989-95079-6-8

(8) $A_{i+1}(x) = A_i(x-h)$ for every $x \in [x_i, x_{i+1}]$ and i= 1,…, n-1.

By limiting ourselves to the discrete case (cfr. [12] in the case of a continuous function), let f be a real function defined on assigned points $p_1,...,p_m$ of [a,b]. Let P be the set of these points and suppose that it is *sufficiently dense* w. r. t. the fixed partition $\{A_1, A_2, \ldots, A_n\}$, that is for every i = 1,…,n there exists an index $j \in \{1,\ldots,m\}$ such that $A_i(p_j) > 0$. Thus we define the numerical vector $[F_1, F_2, ..., F_n]$ as the *direct F-transform* of f w.r.t. the basic functions $\{A_1, A_2, \ldots, A_n\}$, where each F_i is given by

$$F_i = \frac{\sum_{j=1}^{m} f(p_j) A_i(p_j)}{\sum_{j=1}^{m} A_i(p_j)} \quad (1)$$

for every i=1,…,n. Then we define the *inverse F-transform* of f w. r. t. $\{A_1, A_2, \ldots, A_n\}$ by setting for every j = 1,…,m:

$$f_{F,n}(p_j) = \sum_{i=1}^{n} F_i A_i(p_j) \quad (2)$$

Then the following theorem [12] holds:

Theorem 1. Let f(x) be a real function assigned on a set $P = \{p_1,..., p_m\} \subseteq [a,b]$. Then for every $\varepsilon > 0$, there exist an integer $n(\varepsilon)$ and a related fuzzy partition $\{A_1, A_2, \ldots, A_{n(\varepsilon)}\}$ of [a,b] such that P is sufficiently dense w. r. t. $\{A_1, A_2, \ldots, A_{n(\varepsilon)}\}$ and for every $p_j \in [a, b]$, j = 1,…,m, the following inequality holds:

$$|f(p_j) - f_{F,n(\varepsilon)}(p_j)| < \varepsilon \quad (3)$$

As in [15], we extend these definitions to the multi-dimensional case. Let k be given intervals $[a_i,b_i]$ (i=1,…,k) and $[a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_k,b_k]$ be our universe of discourse. Let $x_{11}, x_{12},...., x_{1n_1} \in [a_1,b_1], \ldots, x_{k1}, x_{k2},...., x_{kn_k} \in [a_k,b_k]$ be $n_1+\ldots+n_k$ assigned points, called nodes, such that $x_{i1} = a_i < x_{i2} < \ldots < x_{in_i} = b_i$ for every i = 1,…,k. Let $\{A_{i1}, A_{i2},...., A_{in_i}\}$ be a fuzzy partition of $[a_i,b_i]$ for every i = 1,…,k and $f(x_1,x_{2,\ldots,}x_k)$ be a real function by assuming determined values in m points $p_j = (p_{j1}, p_{j2},\ldots,p_{jk}) \in [a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_k,b_k]$ for every j=1,…,m. We say that the set $P=\{(p_{11}, p_{12}, \ldots, p_{1k}), (p_{21}, p_{22}, \ldots, p_{2k}),\ldots,(p_{m1}, p_{m2}, \ldots,p_{mk})\}$ is *sufficiently dense* w. r. t. the chosen partitions $\{A_{11}, A_{12},..., A_{1n_1}\},\ldots,\{A_{k1}, A_{k2},..., A_{kn_k}\}$ if for each k-tuple $\{h_1,\ldots,h_k\} \in \{1,\ldots,n_1\}\times\ldots\times\{1,\ldots,n_k\}$ there exists a point $p_j = (p_{j1},p_{j2},\ldots,p_{jk}) \in P$, $j \in \{1,\ldots,m\}$, such that $A_{1h_1}(p_{j1}) \cdot A_{2h_2}(p_{j2}) \cdot \ldots \cdot A_{kh_K}(p_{jk}) > 0$. Then we can define the $(h_1,h_2,\ldots,h_k)$-th component $F_{h_1h_2\ldots h_K}$ of the *direct F-transform* of f w.r.t. the basic functions $\{A_{11}, A_{12},..., A_{1n_1}\}, \ldots, \{A_{k1}, A_{k2},..., A_{kn_k}\}$ as

$$F_{h_1h_2\ldots h_K} = \frac{\sum_{j=1}^{m} f(p_{j1}, p_{j2},\ldots p_{jk}) \cdot A_{1h_1}(p_{j1}) \cdot A_{2h_2}(p_{j2}) \cdot \ldots \cdot A_{kh_K}(p_{jk})}{\sum_{j=1}^{m} A_{1h_1}(p_{j1}) \cdot A_{2h_2}(p_{j2}) \cdot \ldots \cdot A_{kh_K}(p_{jk})} \quad (4)$$

The *inverse F-transform* of f w.r.t. the basic functions $\{A_{11}, A_{12},..., A_{1n_1}\}, \ldots, \{A_{k1}, A_{k2},..., A_{kn_K}\}$ is the following function defined for each point $p_j=(p_{j1}, p_{j2},\ldots,p_{jk}) \in [a_1,b_1]\times\ldots\times[a_k,b_k]$ as

$$f^F_{n_1n_2\ldots n_K}(p_{j1}, p_{j2},..., p_{j,k}) = \sum_{h_1=1}^{n_1}\sum_{h_2=1}^{n_2}\ldots\sum_{h_K=1}^{n_k} F_{h_1h_2\ldots h_K} \cdot A_{1h_1}(p_{j1}) \cdot \ldots \cdot A_{kh_K}(p_{jk}) \quad (5)$$

for every j=1,…,m. The following extension of Theorem 1 holds:

Theorem 2. Let $f(x_1,x_{2,\ldots,}x_K)$ be a function assigned on the set of points $P = \{(p_{11},p_{12}, \ldots,p_{1k}) ,(p_{21}, p_{22}, \ldots, p_{2k}),\ldots,(p_{m1}, p_{m2}, \ldots,p_{mk})\} \subseteq [a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_k,b_k]$. Then for every $\varepsilon > 0$, there exist k integers $n_1(\varepsilon),\ldots, n_k(\varepsilon)$ and fuzzy partitions $\{A_{11}, A_{12},..., A_{1n_1(\varepsilon)}\},\ldots,\{A_{k1}, A_{k2},..., A_{kn_k(\varepsilon)}\}$ such that the set P is sufficiently dense w.r.t. the above fuzzy partitions and for every $p_j = (p_{j1}, p_{j2},\ldots,p_{jk}) \in P$, j=1,…,m, the following inequality holds:

$$\left| f(p_{j1}, p_{j2},..., p_{jk}) - f^F_{n_1(\varepsilon)n_2(\varepsilon)\cdots n_k(\varepsilon)}(p_{j1}, p_{j2},..., p_{jk}) \right| < \varepsilon \quad (6)$$

The proof follows the same lines of the similar Theorem 1.

3 Attribute dependencies

We modify slightly the algorithm of [10], where the data are represented in the following matrix:

	X_1	.	X_i	.	X_r
O_1	p_{11}	.	p_{1i}	.	p_{1r}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
O_j	p_{j1}	.	p_{ji}	.	p_{jr}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
O_m	p_{m1}	.	p_{mi}	.	p_{mr}

being $X_1,\ldots,X_i,\ldots,X_r$ the attributes, $O_1, \ldots,O_j, \ldots,O_m$ the objects and p_{ji} the value of the attribute X_i. By setting $a_i = \min\{p_{1i},\ldots,p_{mi}\}$ and $b_i = \max\{p_{1i},\ldots,p_{mi}\}$, we define the interval $[a_i,b_i]$ of the given values of the attribute X_i. We can consider the attribute X_z, $z \in \{1,\ldots, r\}$, dependent from k independent attributes $X_1, \ldots, X_k$, (with $k \le r < m$ and $z \notin \{1,\ldots,k\}$) via a function H defined as

ISBN: 978-989-95079-6-8

$$X_z = H(X_1,...,X_k) \tag{7}$$

We use the F-transforms method in the following process:

1) We assume $n_1 = n_2 = \ldots = n_k = n$ without loss of generality and we refer to the model (7). After normalization in [0,1] of the values of the independent and dependent attributes, we start with n = 3.

2) A uniform fuzzy partition $\{ A_{i1}, A_{i2}, ..., A_{in_i} \}$ is assigned in each interval $[a_i,b_i]$ by setting for every i = 1,…, k and j = 2,…, k-1:

$$A_{i1}(x) = \begin{cases} 0.5 \cdot (1 + \cos\frac{\pi}{h_i}(x - x_{i1})) & \text{if } x \in [x_{i1}, x_{i2}] \\ 0 & \text{otherwise} \end{cases}$$

$$A_{ij}(x) = \begin{cases} 0.5 \cdot (1 + \cos\frac{\pi}{h_i}(x - x_{ij})) & \text{if } x \in [x_{i(j-1)}, x_{i(j+1)}] \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$A_{in}(x) = \begin{cases} 0.5 \cdot (1 + \cos\frac{\pi}{h_i}(x - x_{in})) & \text{if } x \in [x_{i(n-1)}, x_{in}] \\ 0 & \text{otherwise} \end{cases}$$

where $h_i = (b_i - a_i)/(n - 1)$ and $x_{ij} = a_i + h_i \cdot (j-1)$.

3) For each k-tuple $\{h_1,\ldots,h_k\} \in \{1,\ldots,n\}^k$ we see that exists at least an object O_j with values $(p_{j1}, p_{j2}, \ldots p_{jk})$ such that $A_{1h_1}(p_{j1}) \cdot A_{2h_2}(p_{j2}) \cdot \ldots \cdot A_{kh_K}(p_{jk}) > 0$. If this condition is not verified, the set $P = \{(p_{j1},p_{j2},\ldots,p_{jk}), j=1,\ldots,m$ is not sufficiently dense w. r. t. the basic functions (8) and hence the process is stopped.

4) For simplicity, we put $H(p_{j1},p_{j2},\ldots,p_{jk}) = p_{jz}$ for every j=1,2,…,m. In accordance to (4), the $(h_1,h_2,\ldots,h_k)$-th component $F_{h_1h_2\ldots h_K}$ of the direct F-transform of H is defined as

$$F_{h_1h_2..h_k} = \frac{\sum_{j=1}^{m} p_{jz} \cdot A_{1h_1}(p_{j1}) \cdot \ldots \cdot A_{kh_K}(p_{jk})}{\sum_{j=1}^{m} A_{1h_1}(p_{j1}) \cdot \ldots \cdot A_{kh_K}(p_{jk})} \tag{9}$$

Then, in accordance to (5), the inverse F-transform $H^F_{n_1n_2\ldots n_k}$ of H is defined as

$$H^F_{n_1n_2\ldots n_k}(p_{j1}, p_{j2},\ldots p_{jk}) = \sum_{h_1=1}^{n_1}\sum_{h_2=1}^{n_2}\ldots\sum_{h_k=1}^{n_k} F_{h_1h_2\ldots h_K} \cdot A_{1h_1}(p_{j1}) \cdot \ldots \cdot A_{kh_K}(p_{jk}) \tag{10}$$

5) The difference between $H^F_{n_1n_2\ldots n_k}$ and H (or X_z) is estimated in the points $(p_{j1},p_{j2},\ldots,p_{jm})$, j=1,…,m, via the following statistical indexes of determinacy:

$$r_c^2 = \frac{\sigma^2_{H_F}}{\sigma^2_{X_Z}} = \frac{\sum_{j=1}^{m}\left(H^F_{n_1n_2\ldots n_k}(p_{j1}, p_{j2},\ldots p_{jk}) - \hat{p}_z\right)^2}{\sum_{j=1}^{m}\left(p_{jz} - \hat{p}_z\right)^2} \tag{11}$$

$$r'^2_c = 1 - \left[\left(1 - r_c^2\right) \cdot \frac{m-1}{m-k-1}\right] \tag{12}$$

where $\hat{p}_z$ is the mean of the values of X_z. We also look at a threshold value α determined experimentally: if r_c^2 and r'^2_c are less or equal to α, then we consider correct our model (7) and stop the process, otherwise we set n ← n+1 and go back to 2). We fix the threshold α at a confidence value.

Strictly speaking, Theorem 2 is unable to evaluate the correctness of our model (7) because it does not give a method of calculation of the k indexes $n_1(\varepsilon),\ldots, n_k(\varepsilon)$ and of the basic functions w. r. t. which the set $\{p_{j1}, p_{j2}, \ldots, p_{jk}$, j=1,…,m} is sufficiently dense in such a way

$$\left|H(p_{j1},p_{j2},\ldots p_{jk}) - H^F_{n_1(\varepsilon)\ldots n_k(\varepsilon)}(p_{j1},p_{j2},\ldots p_{jk})\right| < \varepsilon \tag{13}$$

In other words, the substitution of H obtained with $H^F_{n_1(\varepsilon)\ldots n_k(\varepsilon)}$ (up to an arbitrary quantity ε) is possible by Theorem 2, but we use the statistical index (12) from a practical point of view. Indeed it allows additionally to compare various regression models that intend to explain the same dependent variable from a different number k of explanatory variables.

4 A sample simulation

The calculation of the threshold value α for the indexes of determinacy (11) and (12) is made with training tests. As sample example, we present the case of a dataset with 1000 records and two fields, that is m=1000, r=2, k=1, X_1 = "Area of a circle", X_2 = "Radius of a circle", $a_1 = 0$, $b_1 = \pi$, $a_2 = 0$ and $b_2 = 1$. Hence the values of the attributes are $p_{j1} \in [0,\pi]$ and $p_{j2} \in [0,1]$ for every j = 1, …, 1000. If z = 2, we have $X_2 = H(X_1) = (X_1/\pi)^{1/2}$. In the interval [0, π] we consider a maximum number of 500 basic functions of the type (8) since we have seen that the set $P = \{p_{j1} : j = 1,\ldots,1000\}$ is not sufficiently dense w. r. t. the chosen partitions for $n \geq 500$. Table 1 contains the values of the indexes r_c^2 and r'^2_c for several values of n and it is plain that a good choice is α = 0.9 for $n \leq 7$.

ISBN: 978-989-95079-6-8

Table 1. Indexes of determinacy

n	r_c^2	r'^2_c
3	0.587000	0.588000
5	0.826898	0.827725
7	0.900207	0.901108
10	0.943000	0.944000
15	0.969749	0.970719
25	0.986085	0.987072
50	0.995084	0.996080
300	0.999000	1.000000
400	0.999000	1.000000
500	0.999000	1.000000

For sake of completeness, we show the plots in Figure 1 and Figure 2 of the values of X_2 and H_n^F for n = 3 and n = 300, respectively.

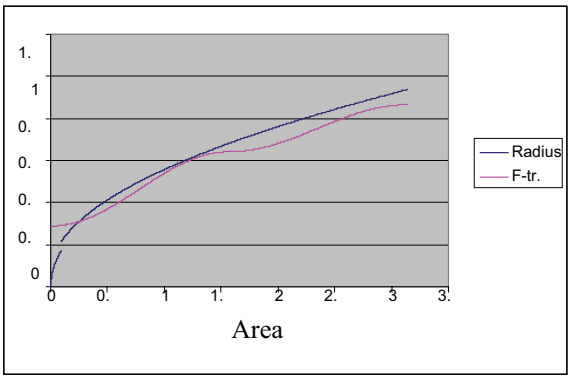


Figure 1. Graph of X_2 and H_3^F.

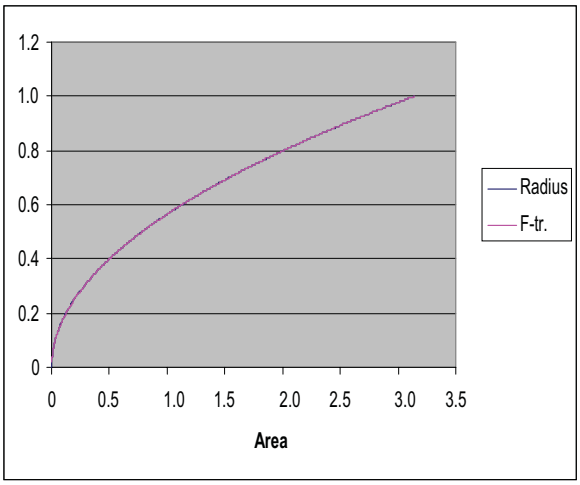


Figure 2. Graph of X_2 and H_{300}^F.

5 A complete experiment

We have downloaded from the known dataset "El Nino" (http://kdd.ics.uci.edu/databases/el_nino/el_nino.data.html) the oceanographic and surface meteorological data measured from a series of buoys positioned throughout the equatorial ocean Pacific. The data are formed from the following attributes:

- X_1 = date,
- X_2 = latitude,
- X_3 = longitude,
- X_4 = zonal winds (west < 0, east > 0),
- X_5 = meridional winds (south < 0, north > 0),
- X_6 = relative humidity,
- X_7 = air temperature in degrees Celsius,
- X_8 = sea surface temperature in degrees Celsius.

The data were measured from the buoys in various locations since 1980. The regression analysis establishes that significant relationships between the variables were not observed except for $X_8=H(X_7)$ and $X_8=H(X_1,X_7)$ which are of linear type like we will show in the sequel.
The training tests give $\alpha = 0.8$ as a threshold value. For single attribute dependencies, Table 2 shows the possible maximal number n_{max} of basic functions which can be used for deducing H_n^F. If n_{max} is overcome, the set of the attribute values is not sufficiently dense w. r. t. the chosen partitions. Moreover the absence of n_{max} in some entry of Table 2 means that the dependence model is not analyzable with the F-transforms method.

Table 2. n_{max} usable to explore single dependencies

	Date	Latit.	Lon.	Zon. wind	Mer. wind	Rel. hum.	Air temp	Sea sur. tem..
Date		13	3	14	48	5	14	3
Latitude	25		3	14	48	5	21	3
Longit.	25	13		14	48	5	21	3
Zonal winds	25	13	3		48	4	21	3
Merid. winds	25	13	3	14			21	3
Relative humid.	25	13		10	48		20	3
Air temper.	25	13	3	7	33			3
Sea sur. temper.	25	13	3	7	33	3	21	

The most meaningful model is $X_8=H(X_7)$ as Figure 3 shows: the plot is obtained for n = 10 and moreover we have $r_c^2 = r'^2_c = 0.83$. In the analysis of multiple attribute dependencies, Table 3 shows that only the model $X_8 = H(X_1,X_7)$ has the indexes of determinacy greater than $\alpha = 0.8$. Figure 4 shows the related 3D graph.

6 Conclusions

The F-transforms analyze all the possible dependencies between attributes in the datasets via a function H whose inverse F-transform is inside the value of the indexes of determinacy compared with a threshold parameter α. Our approach has the advantage to control always that the set of the attribute values is sufficiently dense w.r.t. the basic functions forming an uniform partition of the interval of context related to each attribute, otherwise the algorithm is stopped.

ISBN: 978-989-95079-6-8

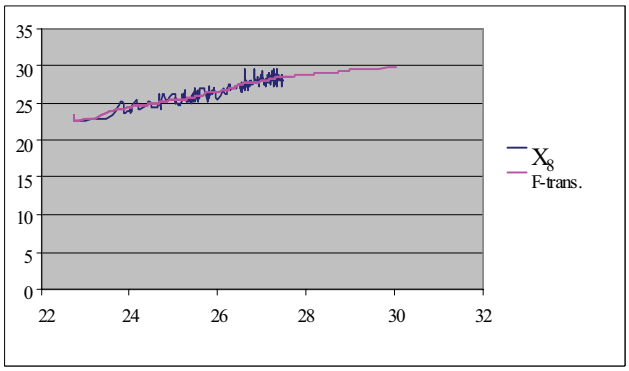


Figure 3. Graph of $X_8 = H(X_7)$ and H_{10}^F.

Table 3. Best values of indexes of determinacy for X_8

X_z	$X_1,\ldots, X_k$	r_c^2	r'^2_c
X_8	X_1, X_7	0.827	0.828
X_8	X_6, X_7	0.341	0.338
X_8	X_4, X_5	0.266	0.266
X_8	X_1, X_6, X_8	0.110	0.111

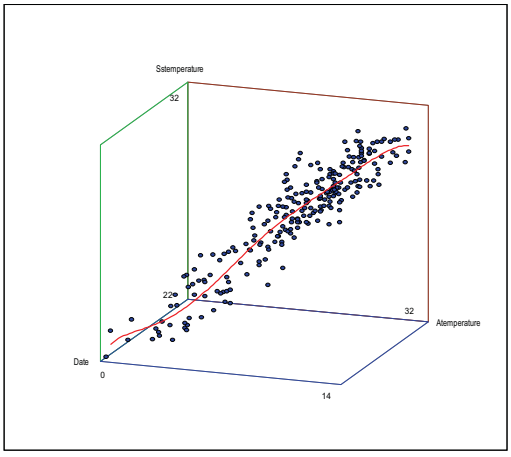


Figure 4. Plot 3D of $X_8 = H(X_1,X_7)$.

This method shall be integrated in future works with data mining classification and with results of regression data, such as decision tree learning and association mining algorithms.

References

[1] M. Daňková and M. Štěpnička, Fuzzy transform as an additive normal form, *Fuzzy Sets and Systems* 157 (8) (2006) 1024–1035.

[2] F. Di Martino, V. Loia, I. Perfilieva and S. Sessa, An image coding/ decoding method based on direct and inverse fuzzy transforms, *Internat. Journal of Approximate Reasoning* 48 (2008) 110–131.

[3] F.Di Martino, V. Loia and S. Sessa, Fuzzy transforms for compression and decompression of color videos, *Information Sciences,* submitted.

[4] F. Di Martino and S. Sessa, Compression and decompression of images with discrete fuzzy transforms, *Information Sciences* 177 (2007) 2349–2362.

[5] N.R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons Inc., New York, 1998.

[6] J. Han and M. Kamber, *Data Mining: Concepts and Techniques* (2nd Edition), Morgan-Kaufman, New York, 2006.

[7] R.A. Johnson and D.W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall International, London, 1992.

[8] Y. S. Lee and S. J. Yen, Classification based on attribute dependency, Proceedings of 6th Internat. Conference DaWaK' 04 (Zaragoza, 1-3/9/04) *Lecture Notes in Computer Sciences*, Vol. 5192, 2004, 259 – 268.

[9] S. Mitra, S.K. Pal and P. Mitra, Data mining in soft computing framework: a survey, *IEEE Transactions on Neural Networks* 13 (1) (2002) 3–14.

[10] I. Perfilieva, V. Novák and A. Dvořák, Fuzzy transforms in the analysis of data, *Internat. Journal of Approximate Reasoning* 48 (2008) 36–46.

[11] I. Perfilieva, Fuzzy transforms: application to reef growth problem, in: R. B. Demicco, G. J. Klir (eds.), *Fuzzy Logic in Geology*, Academic Press, Amsterdam, 2003, pp. 275–300.

[12] I. Perfilieva, Fuzzy transforms: theory and applications, *Fuzzy Sets and Systems* 157 (8) (2006) 993–1023.

[13] I. Perfilieva, Fuzzy transforms: a challenge to conventional transforms, In: P.W. Hawkes (Ed.), *Advances in Images and Electron Physics*, vol. 147, Elsevier Academic Press, San Diego, 2007, pp. 137–196.

[14] M. Štěpnička and R. Valášek, Fuzzy transforms for functions with two variables, in: J. Ramik, V. Novák (eds.), *Methods for Decision Support in Environment with Uncertainty-Applications in Economics*, *Business and Engineering*, IRAFM, Ostrava, Czech Republic, 2003, pp. 96–102.

[15] M. Štěpnička and R. Valášek, Numerical solution of partial differential equations with help of fuzzy transform, in: *Proc. Int. Conf. on FUZZ – IEEE*, Reno, 2005, pp. 1104–1109.

ISBN: 978-989-95079-6-8

Interval-valued Fuzzy propositions. An application to the L-Fuzzy contexts with absent values

C.Alcalde[1] A. Burusco, R. Fuentes-González[2]

1.Escuela Universitaria Politécnica. Departamento de Matemática Aplicada.
Universidad del País Vasco. Plaza de Europa 1, 20018 San Sebastián, Spain.
2.Departamento de Automática y Computación.
Universidad Pública de Navarra, Campus de Arrosadía, 31006 Pamplona, Spain.
Email: c.alcalde@ehu.es, {burusco,rfuentes}@unavarra.es

Abstract— *In order to obtain the information from an L-Fuzzy context, the complete relation between the objects and the attributes is needed. However, the contexts that model many situations have absent values.*
To solve this problem, at the beginning of the paper we remind the interval-valued linguistic variable definition and, later, we propose an extension of the fuzzy propositions to the interval-valued case.
In the second part, we apply these ideas to the problem of replacement of the absent values in an interval-valued L-Fuzzy Context. The use of interval-valued fuzzy propositions simplifies the process.

Keywords— Interval-valued fuzzy propositions, Interval-valued linguistic variables, Interval-valued L-Fuzzy Contexts.

1 Introduction

The first extension of the Formal Concept Analysis to the fuzzy case, the L-Fuzzy Concept Theory, is due to A. Burusco and R. Fuentes-González and was published in 1994 [1].

Later, to extract knowledge from a more general table, we defined the interval-valued L-Fuzzy context as an extension to the interval-valued case of the L-Fuzzy context defined in the L-Fuzzy Concept Theory [2]:

An *interval-valued L-Fuzzy context* is a tuple $(\mathcal{J}[L], X, Y, R)$, with X and Y two finite sets (of *objects* and *attributes*), R an interval-valued L-Fuzzy relation between X and Y and $\mathcal{J}[L]$ the set of the closed intervals of a lattice L with the usual order.

To replace the absent values, we have already developed a theory based on implications between attributes [3] that has given good results when has been applied to the Technique [4]. In these papers, in order to replace the absent values, we have used some labels that allow to obtain implications with high values of support and confidence.The good behavior of these labels, suggested us the application of the linguistic variables to our work [5].

In this work, we are going to define fuzzy propositions for the interval-valued case (they do not exist in the Literature of the subject) and later we will try to use them to replace the absent values in contexts instead of using implications between attributes. We will finally analyze the advantages of the method.

2 Linguistic variables defined in the set of closed intervals of $[0, 1]$

Let $(V, T(V), U, G, M)$ be a linguistic variable (Zadeh [6]) defined in the set $U = [0, 1]$, whose values or labels $T(V)$ are associated with the generalized trapezoidal fuzzy numbers defined by Yao and Lin [7]. We have extended the linguistic variable definition to the interval-valued case [5]:

Definition 1 *Taking as a departure point a linguistic variable V, we define an interval-valued linguistic variable associated with V as the linguistic variable $\mathbf{V}$ defined in the set $\mathcal{J}[U]$ of the closed intervals of U, characterized by the tuple $(\mathbf{V}, T(V), \mathcal{J}[U], G, M)$, where the compatibility function with each label $t \in T(V)$ is $\boldsymbol{c}_t : \mathcal{J}[U] \to \mathcal{J}[0, 1]$.*
Notation. *We denote the compatibility of the value $[\alpha, \beta] \in \mathcal{J}[U]$ with the label t by:*

$$\boldsymbol{c}_t([\alpha, \beta]) = [\alpha, \beta]_t = \\ = \begin{cases} [\min\{\alpha_t, \beta_t\}, \max\{\alpha_t, \beta_t\}] & \text{if } \alpha \geq b \text{ or } \beta \leq a \\ [\min\{\alpha_t, \beta_t\}, 1] & \text{in other case} \end{cases}$$

where a and b are the values that define the generalized trapezoidal fuzzy number associated with the label $t \in T(V)$.

With this definition we will try to represent the interval in which x_t have its values when $x \in [\alpha, \beta]$.

The defined interval-valued linguistic variable $\mathbf{V}$ is an extension of the linguistic variable V to the set of closed intervals of U.

Some properties of these interval-valued linguistic variables have been proven in [5].

3 Interval-valued fuzzy propositions

Before studying the interval-valued fuzzy propositions, we are going to remind the main aspects about fuzzy propositions.

3.1 Fuzzy propositions

As can be seen in [8], the main difference between classical and fuzzy propositions is that the truth value of the second ones belongs to the interval [0,1]. Moreover, four types of fuzzy propositions are defined by Klir [8]:

Unconditional and unqualified propositions: $p : \Phi$ *is* A, with Φ a variable and A a predicate attributed to the variable and represented by a fuzzy set. In other words, p : 'Φ *is* A' *is true*. The truth value of the proposition $p_\Phi : \Phi = \phi$ *is* A, is obtained by $T_r(p_\phi) = A(\phi), \forall \phi \in \Phi$.

Unconditional and qualified propositions: The difference with the previous ones is that the proposition has a qualified truth value that modifies the truth value. They are

ISBN: 978-989-95079-6-8

characterized by the form p : 'Φ *is* A*' is* S, with S qualified truth value.
Examples of truth qualifiers are expressions as *true*, *very true*, *fairly true*, *false*, *very false* or *fairly false* (See Fig. 1).

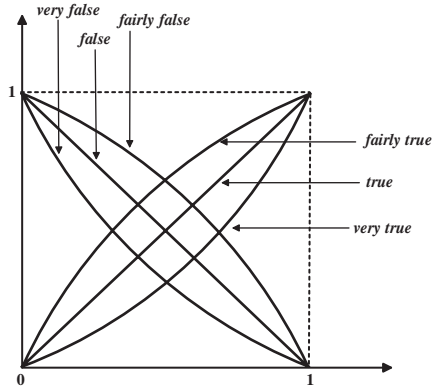


Figure 1: Some examples of truth qualifiers

Conditional and unqualified propositions: These propositions are expressed by p : *If* Φ *is* A*, then* Ψ *is* B, with Φ and Ψ variables and A and B predicates represented by fuzzy sets. That is, p : '*If* Φ *is* A*, then* Ψ *is* B*' is true*. The truth value of the proposition is calculated using an implication operator:

$$T_r(p_{\phi,\psi}) = I[A(\phi), B(\psi)]$$

Conditional and qualified propositions: p : '*If* Φ *is* A*, then* Ψ *is* B*' is* S, with S a qualified truth value. The truth value of the proposition is obtained by

$$T_{rS}(p_{\phi,\psi}) = S[T_r(p_{\phi,\psi})]$$

3.1.1 Fuzzy quantifiers

The two quantifiers of the predicate logic are *all* and *exists*. Fuzzy quantifiers are a tool for symbolizing quantified statements that minimizes the lost of information forced by the choice of quantifier.

The fuzzy propositions of any of the types introduced in the previous section may be quantified by a suitable fuzzy quantifier. In general, fuzzy quantifiers are fuzzy numbers which take part in the various propositional forms and affect the degrees of truth of specific fuzzy propositions. Each fuzzy quantifier expresses an approximate number of elements or an approximate proportion of elements in a given universal set that claims to satisfy a given property.

There are two types of fuzzy quantifiers: The absolute quantifiers expressed by fuzzy numbers defined on the set of real numbers or on the set of integers: *about a dozen*, *at most about 10*, *at least about 100...* And the relative quantifiers expressed by fuzzy numbers defined on [0,1]: *most*, *almost all*, *about half*, *about 20%*. These relative quantifiers are the most interesting for our study. So, we are going to see how the fuzzy propositions with these fuzzy quantifiers are.

Fuzzy propositions with quantifiers of the second kind have the general form ([8]):

p : *Among z's in* Z *such that* $\nu_1(z)$ *is* F_1 *there are* Q *z's in* Z *such that* $\nu_2(z)$ *is* F_2

Any propositions p can be expressed in a simplified form:

$$p' : QE_1's \text{ are } E_2's$$

where Q is the used quantifier, E_1 and E_2 are the fuzzy sets on Z defined by $E_1(z) = F_1(\nu_1(z))$, $E_2(z) = F_2(\nu_2(z)), \forall z \in Z$, with F_1 and F_2 fuzzy sets.

We may rewrite this proposition in the form:

$$p' : W \text{ is } Q$$

with W a variable that represents the degree of subsethood of E_2 in E_1, that is:

$$W = \frac{|E_1 \cap E_2|}{|E_1|}$$

where $|\quad|$ represents the cardinality of a set.

Using the standard fuzzy intersection, we obtain:

$$W = \frac{\sum_{z \in Z} \min\{F_1(\nu_1(z)), F_2(\nu_2(z))\}}{\sum_{z \in Z} F_1(\nu_1(z))}$$

for any E_1 and E_2.

Then, the truth value is $T_r(p) = T_r(p') = Q(W)$.

3.1.2 Linguistic hedges

Linguistic hedges are special linguistic terms by which other linguistic terms are modified: *Very*, *more or less*, *fairly* or *extremely*. They can be used for modifying fuzzy predicates, fuzzy truth values and fuzzy quantifiers.

Any linguistic hedge may be interpreted as an unary function h in [0,1]. For example, *Very* ($h(a) = a^2$), *fairly* ($h(a) = \sqrt{a}$). These functions h are also said to be modifiers or qualifiers.

3.2 Interval-valued fuzzy propositions

There are some situations where we work with interval-valued fuzzy variables and fuzzy propositions and we think that the truth value of an interval-valued fuzzy proposition must be also an interval in order to model in a more suitable way the represented data. An example of this situation can be seen in next section.

With this aim, we are going to study the interval-valued fuzzy propositions.
Then, we define an interval-valued fuzzy proposition with fuzzy quantifiers of the second kind as:

$$\mathbf{p}' : QE_1{}'s \text{ are } E_2{}'s$$

where Q is the quantifier, and E_1,$E_2 \in \mathcal{J}[L]^X$ are interval-valued fuzzy sets.

As in the previous section, we can rewrite this proposition as:

$$\mathbf{p}' : W \text{ is } Q$$

ISBN: 978-989-95079-6-8

with W a variable that represents the idea of degree of subsethood of E_2 in E_1.

Using the standard fuzzy intersection, we can express this degree as:

$$W = \left[\frac{\sum_{x\in X}\min\{l_{E_1}(x), l_{E_2}(x)\}}{\sum_{x\in X} u_{E_1}(x)}, \frac{\sum_{x\in X}\min\{u_{E_1}(x), u_{E_2}(x)\}}{\sum_{x\in X} l_{E_1}(x)} \wedge 1\right]$$

where u_{E_1}, u_{E_2}, l_{E_1} y l_{E_2} are the upper and lower bounds of the interval-valued observations of E_1 and E_2.

Then, we define the truth value of the proposition $\mathbf{p}'$ as

$$\mathbf{T_r}(\mathbf{p}') = Q(W) = \\ = [\min\{Q(l_W), Q(u_W)\}, \max\{Q(l_W), Q(u_W)\}]$$

where l_W and u_W are the lower and upper bounds of W respectively.

4 Use of the linguistic variables and fuzzy propositions to replace the absent values in Contexts

4.1 Interval-valued L-Fuzzy Contexts with absent values

To extract knowledge from a table with incomplete and ambiguous information, we use the interval-valued L-Fuzzy context as an extension of the L-Fuzzy context given by an implication operator [1, 2] to the interval-valued case.

Definition 2 *Let $\mathcal{J}[L]$ be the set of the closed intervals of a lattice L with the usual order. An* interval-valued L-Fuzzy context *is a tuple $(\mathcal{J}[L], X, Y, R)$, with X and Y two finite sets (of* objects *and* attributes*) and $R \in \mathcal{J}[L]^{X\times Y}$ an interval-valued L-Fuzzy relation between X and Y.*

We are going to see an example of an interval-valued L-Fuzzy context to understand the problem.

Example 1 *In Table 1 we have collected the amounts of the different ingredients (I_i) of a certain food item produced under several trademarks (M_j). As can be seen, there are two unknown values. The table values have been normalized and belong to $\mathcal{J}[L]$, the set of intervals of the lattice $L = \{0, 0.1, 0.2, \ldots 0.9, 1\}$.*
This is an example of an interval-valued L-Fuzzy context $(\mathcal{J}[L], X, Y, R)$, where the set of objects X is the set of trademarks or brands, the set of attributes Y is the set of ingredients and the relation R between both is represented in the table.

In the papers developed to study contexts with absent values [3, 4] we have analyzed the implications between attributes associated with labels to obtain good values to replace the absent ones.

In those papers, we represent these absent values by the interval [0,1] and analyze the implications between attributes with the intention of obtaining good values to reduce, as much as possible, the width of that interval.

The support and confidence definitions of an interval-valued implication are based on the idea of association rules [9].

Definition 3 *Let $(\mathcal{J}[L], X, Y, R)$ be an interval-valued L-Fuzzy context. Given the attribute sets $B, C \in \mathcal{J}[L]^Y$, we define the* support *of the implication $B \Rightarrow C$, $\mathrm{supp}(B \Rightarrow C)$, as the interval:*

$$\left[\frac{\sum_{x\in X} l_{(B\cup C)_2}(x)}{|X|}, \frac{\sum_{x\in X} u_{(B\cup C)_2}(x)}{|X|}\right]$$

where the membership function of the derived set B_2 is $B_2(x) = [l_{B_2}(x), u_{B_2}(x)]$ and represents the percentage of objects that share the attributes of B and C.

The confidence *of the implication, $\mathrm{conf}(B \Rightarrow C)$, is given by the interval:*

$$\left[\frac{\sum_{x\in X} l_{(B\cup C)_2}(x)}{\sum_{x\in X} u_{B_2}(x)}, \frac{\sum_{x\in X} u_{(B\cup C)_2}(x)}{\sum_{x\in X} l_{B_2}(x)} \wedge 1\right]$$

and represents the percentage of objects that verify the implication, that is, the percentage of objects that having the attributes of B to a certain degree also have those of C to the same degree.

One kind of these implications that we have used are the implications associated with labels [3]. They are denoted by $y_{i_r} \Longrightarrow y_{j_s}$ where y_i and y_j are attributes of the context and r, s labels. In the calculation process, a function G_s (based on the distance between intervals) to measure the degree of similarity between the different labels and the attributes is needed. In addition, once the implications with highest support and confidence have been chosen, the absent values are replaced by intervals, in general, with less amplitude than [0,1].

In this work, we are going to approach all this process, in an easy way, using interval-valued linguistic variables and fuzzy propositions.

The implication between attributes $y_{i_r} \Longrightarrow y_{j_s}$ can be considered as a quantified fuzzy proposition since the expression:

$\mathbf{p}$: *Most of those that have y_i(high, medium, low...) also have y_j (high, medium, low...)*

can be interpreted in the sense of section 3.2:

The quantifier Q is *most* and E_1,E_2 are interval-valued linguistic variables resulting of applying the linguistic variable (with values *high, medium, low...*) to the columns of the context. The quantifier Q will be applied to both of the bounds of the intervals. In this way, we can calculate the truth value of the proposition:

$$\mathbf{T_r}(\mathbf{p}) = Q(W) = \\ = [\min\{Q(l_W), Q(u_W)\}, \max\{Q(l_W), Q(u_W)\}]$$

4.2 Substitution of the absent values using the compatibility function of an interval-valued linguistic variable and the interval-valued fuzzy propositions.

In the study of contexts with absent values, we begin replacing these absent values by [0,1]. Next, we are going to define an interval-valued linguistic variable $(\mathbf{V}, T(V), \mathcal{J}[U], G, M)$, and use the interval-valued fuzzy propositions to estimate the

ISBN: 978-989-95079-6-8

Table 1: Amounts of the different ingredients under several trademarks.

R	$I1$	$I2$	$I3$	$I4$	$I5$	$I6$
$M1$	$[0.8, 1]$	$[0.1, 0.1]$	$[0.6, 0.6]$	$[0.2, 0.2]$	$[0.4, 0.4]$	$[0.4, 0.6]$
$M2$	$[0.8, 0.8]$		$[0.4, 0.6]$	$[0.4, 0.4]$	$[0.2, 0.2]$	$[0.6, 0.6]$
$M3$	[1,1]	[0,0]	$[0.2, 0.2]$	$[0.5, 0.5]$	$[0.4, 0.4]$	$[0.4, 0.4]$
$M4$	$[0.9, 0.9]$	$[0.2, 0.2]$	$[0, 0]$	$[0.4, 0.4]$	$[0.1, 0.2]$	
$M5$	$[0.8, 0.8]$	$[0, 0.2]$	$[0.4, 0.4]$	$[0.2, 0.2]$	$[0.2, 0.4]$	$[0.2, 0.2]$

absent values $R(x_i, y_j)$ of the interval-valued L-Fuzzy context by intervals with less amplitude than that of [0,1].

The compatibility function of the interval-valued linguistic variable will be used to establish the new columns in the table that measure the similarity between some attributes and the values or labels $t \in T(V)$ of the linguistic variable.

Example 2 *In the previous context, the two absent values ($R(M2, I2)$ and $R(M4, I6)$) have been replaced by the interval [0,1].*

We are going to define a linguistic variable ***V****, with labels* $T(V)$ =*{high, medium, low}.*

Now, we take the label t =medium $\in T(V)$ *represented by Fig. 2,*

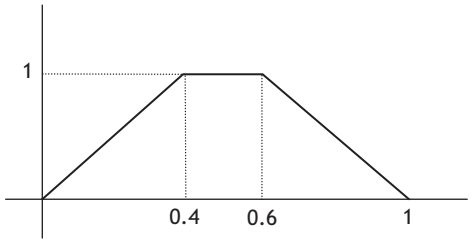


Figure 2: L-Fuzzy set associated with the label t

where the compatibility of the value $x \in X = [0, 1]$ *with the label* t =medium *has the following expression:*

$$x_t = \begin{cases} \dfrac{x}{0.4} & \text{if } x \leq 0.4 \\ 1 & \text{if } 0.4 \leq x \leq 0.6 \\ \dfrac{1-x}{1-0.6} & \text{if } 0.6 \leq x \end{cases}$$

and the interval-valued compatibility is defined, $\forall [\alpha, \beta] \in \mathcal{J}[X]$, *by:*

$$[\alpha, \beta]_t = \begin{cases} [\min\{\alpha_t, \beta_t\}, \max\{\alpha_t, \beta_t\}] & \text{if } \alpha \geq 0.6 \text{ or } \beta \leq 0.4 \\ [\min\{\alpha_t, \beta_t\}, 1] & \text{in other case} \end{cases}$$

From this point, we can extend the table with a new column $I6_{medium}$ *where the compatibility of the attribute* $I6$ *with the label* $t = medium$ *is valued, obtaining Table 2.*

Remark. Sometimes, the obtained values do not belong to the lattice $L = \{0, 0.1, 0.2, \ldots 0.9, 1\}$. Then, we will round the fractions to the nearest element of L.

One of the advantages of the use of linguistic variables to establish the compatibility between the values of the relation R and the labels is the maintenance of the unknowledge. It is possible to observe this property in the example where $R(M4, I6) = [0, 1] = \widehat{R}(M4, I6_{medium})$ and it is certain for any label $t \in T(V)$, as it was proven in [5].

On the other hand, the values here obtained are not very different from which turned out applying the techniques developed in [3] which can be seen in the Table 3. Nevertheless, we consider that the use of linguistic variables solves the problem in a simpler and more natural way.

Table 3: Values obtained from a previous method.

	$I6_{medium}$
$M1$	$[0.8, 1]$
$M2$	$[0.8, 1]$
$M3$	$[0.8, 1]$
$M4$	$[0.4, 1]$
$M5$	$[0.6, 0.8]$

Once extended the table, the interval-valued fuzzy propositions related to the new attribute $I6_{medium}$ can be calculated to find those that have highest values. In this case, this is the proposition $I1 \Rightarrow I6_{medium}$:

$\mathbf{p}$: *Most of those that have* $I1$ *also have* $I6_{medium}$

Then, Q =*most*, $E_1 = I1$, $E_2 = I6_{\text{medium}}$, where

$$W = \left[\frac{3.1}{4.5}, \frac{4.2}{4.3} \wedge 1\right] = [0.68, 0.97]$$

Now, if we take the quantifier Q defined by ([10]):

$$Q(x) = \begin{cases} 1.25x & \text{if } x \leq 0.8 \\ 1 & \text{in other case} \end{cases}$$

and we apply to both bounds of the interval, then we have:

$$Q([0.68, 0.97]) = [Q(0.68), Q(0.97)] = [0.85, 1]$$

that is the truth value of the proposition $\mathbf{p}$.

The obtained results are similar to those with implications between attributes but, in this case, the process is easier.

We can conclude that, in many cases (at least 85%), the ingredient $I1$ is associated with a medium amount of ingredient $I6$.

Given the high percentage of cases in which the implication is valid, we suppose that it also holds for brand $M4$. Therefore we expect that:

$$\widehat{R}(M4, I6_{medium}) \geq \widehat{R}(M4, I1) = [0.9, 0.9]$$

ISBN: 978-989-95079-6-8

Table 2: New relation $\widehat{R}$.

$\widehat{R}$	$I1$	$I2$	$I3$	$I4$	$I5$	$I6$	$I6_{medium}$
$M1$	$[0.8, 1]$	$[0.1, 0.1]$	$[0.6, 0.6]$	$[0.2, 0.2]$	$[0.4, 0.4]$	$[0.4, 0.6]$	$[1, 1]$
$M2$	$[0.8, 0.8]$	$[0, 1]$	$[0.4, 0.6]$	$[0.4, 0.4]$	$[0.2, 0.2]$	$[0.6, 0.6]$	$[1, 1]$
$M3$	[1,1]	[0,0]	$[0.2, 0.2]$	$[0.5, 0.5]$	$[0.4, 0.4]$	$[0.4, 0.4]$	$[1, 1]$
$M4$	$[0.9, 0.9]$	$[0.2, 0.2]$	$[0, 0]$	$[0.4, 0.4]$	$[0.1, 0.2]$	$[0, 1]$	$[0, 1]$
$M5$	$[0.8, 0.8]$	$[0, 0.2]$	$[0.4, 0.4]$	$[0.2, 0.2]$	$[0.2, 0.4]$	$[0.2, 0.2]$	$[0.5, 0.5]$

The election of the attributes and labels to obtain interval-valued fuzzy propositions with high truth values is very important because we will have then good values of substitution for the absent ones.

At this point, we have obtained the value $\widehat{R}(M4, I6_{medium}) = [0.9, 0.9]$. Now, using this value we want to get the best value to replace the initial absent value $R(M4, I6)$. To do this, denoting by $R(x_i, y_j) = [\alpha, \beta]$ the unknown value, we will use the following proposition associated with compatibility functions to obtain this value:

Proposition 1 *$\forall x \in [0, 1]$ and $[\alpha, \beta] \in \mathcal{J}[0, 1]$ it is verified that*

$$[\alpha, \beta]_t \subseteq [x, 1] \Leftrightarrow [\alpha, \beta] \subseteq [ax, 1 - x(1 - b)]$$

where $a, b \in [0, 1]$ are the values that appear in the fuzzy set associated with the label $t \in T(V)$.

Proof

$\Leftarrow$) Let us suppose that $[\alpha, \beta] \subseteq [ax, 1 - x(1 - b)]$.
In [5], we have proven that:

If $[\alpha, \beta] \subseteq [ax, 1 - x(1 - b)]$ holds,

then

$$[\alpha, \beta]_t \subseteq [ax, 1 - x(1 - b)]_t$$

Moreover, for all $x \in [0, 1]$, it is verified

$$\left.\begin{array}{l} ax \leq a \\ 1 - x(1 - b) \geq b \end{array}\right\} \Rightarrow$$

Therefore,

$$\begin{aligned} [\alpha, \beta]_t \subseteq [ax, 1 - x(1 - b)]_t = \\ = [\min\{(ax)_t, (1 - x(1 - b))_t\}, 1] = \\ = \left[\min\left\{\frac{ax}{a}, \frac{1 - (1 - x(1 - b))}{1 - b}\right\}, 1\right] = [x, 1] \end{aligned}$$

$\Rightarrow$) By reduction to the absurd, suppose that

$$[\alpha, \beta] \not\subseteq [ax, 1 - x(1 - b)]$$

then we have $\alpha < ax$ or $\beta > 1 - x(1 - b)$.

- If $\alpha < ax < a$, then:

$$\alpha_t = \frac{\alpha}{a} < \frac{ax}{a} = x \Rightarrow \min\{\alpha_t, \beta_t\} < x$$

- If $\beta > 1 - x(1 - b) > b$, then:

$$\beta_t = \frac{1 - \beta}{1 - b} < \frac{1 - (1 - x(1 - b))}{1 - b} = x \Rightarrow$$
$$\Rightarrow \min\{\alpha_t, \beta_t\} < x$$

Therefore $[\alpha, \beta]_t \not\subseteq [x, 1]$.

Remark. Really, we are only interested in the implication

$$[\alpha, \beta]_t \subseteq [x, 1] \Rightarrow [\alpha, \beta] \subseteq [ax, 1 - x(1 - b)]$$

To replace the absent value, we will take the widest interval verifying the condition, that is, $[ax, 1 - x(1 - b)]$.

As $[a, b] \subseteq [ax, 1 - x(1 - b)]$, we have to avoid to take labels with a wide amplitude of the interval $[a, b]$ because, in other case, the interval $[ax, 1 - x(1 - b)]$ will be also wide.

Example 3 *Returning to the example, if we apply the previous proposition associated with the label t=medium (with the values $a = 0.4$ and $b = 0.6$), then we obtain*

$$R(M_4, I_6) = [0.36, 0.64]$$

that will be the interval with which we will replace the absent value (also it will be necessary here to round the results to the nearest element of L). The final result given in [3] for this same absent value $R(M_4, I_6)$ was $[0.5, 0.5]$.

In both cases we have a fuzzy value distant from 0 and 1.

Finally, the other unknown value, $R(M2, I2)$ can be replaced using interval-valued linguistic variables and fuzzy propositions by the interval [0,0.2]. In this case, the proposition with the highest value is

p : *Most of those that have $I1_{high}$ also have $I2_{low}$*

5 Conclusions and future work

The use of linguistic variables and fuzzy propositions in the field of contexts with absent values supposes a simplification of the process of replacement of these values. The old method, based on implication between attributes, needs the calculus of functions to measure the similarity between attributes and labels that makes difficult the process.

In future works we will analyze its reliability.

Acknowledgment

This work has been partially supported by the Research Group "Intelligent Systems and Energy" of the University of the Basque Country, under grant GIU 07/45 and by the Research Project of the Government of Navarra (Resolution 2031 of 2008).

ISBN: 978-989-95079-6-8

References

[1] A. Burusco and R. Fuentes-González. The study of the L-fuzzy concept lattice. *Mathware and Soft Computing*, 1(3):209–218, 1994.

[2] A. Burusco and R. Fuentes-González. The study of the interval-valued contexts. *Fuzzy Sets and Systems*, 121(3):69–82, 2001.

[3] C. Alcalde, A. Burusco, R. Fuentes-González, and I. Zubia. Treatment of L-fuzzy contexts with absent values. *Information Sciences*, 179:1–15, 2009.

[4] C. Alcalde, A. Burusco, R. Fuentes-González, and I. Zubia. An application of the L-fuzzy concept theory to the short-circuit detection. *IEEE International Conference on Fuzzy Systems*, pages 547–555, 2007.

[5] C. Alcalde, A. Burusco, and R. Fuentes-González. Interval-valued linguistic variables. *The 8th International FLINS Conference on Computational Intelligence in Decision and Control*, pages 217–222, 2008.

[6] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, 8:199–249, 1975.

[7] J.S. Yao and Lin F.T. Constructing a fuzzy flow-shop sequencing model based on statistical data. *International Journal of Approximate Reasoning*, 29(3):215–234, 2002.

[8] G.J. Klir, U.H. St.Clair, and B. Yuan. *Fuzzy Set Theory. Foundations and Applications*. Prentice Hall, 1997.

[9] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data.*, pages 207–216, 1993.

[10] J. Kacprzyk, M. Fedrizzi, and H. Nurmi. *Fuzzy Logic with Linguistic quantifiers in group decision making. In: An introduction to Fuzzy Logic applications in intelligent systems*. Kluwer Academic Publishers, 1992.

ISBN: 978-989-95079-6-8

On the structure of the k-aditive fuzzy measures

Elías F. Combarro[1] Pedro Miranda[2]

1.Artificial Intelligence Center, University of Oviedo, Spain
2.Department of Statistics and O. R. , Complutense University of Madrid, Spain
Email: elias@aic.uniovi.es, pmiranda@mat.ucm.es

***Abstract*—** The family of k-additive measures has been introduced as a midterm between probabilities and general fuzzy measures and finds a wide number of applications in practice. However, its structure is different from other families of fuzzy measures and is certainly more complex (for instance, its vertices are not always $\{0,1\}$-valued), so it has not been yet fully studied.

In this paper we present some results concerning the extreme points of the k-additive fuzzy measures. We give a characterization of these vertices as well as an algorithm to compute them. We show some examples of the results of this algorithm and provide lower bounds on the number of vertices of the $n-1$-additive measures, proving that it grows much faster than the number of vertices of the general fuzzy measures. This suggests that k-additive measures might not be a good choice in modeling certain decision problems when the value of k is high but not equal to n.

***Keywords*—** Fuzzy measures, k-additive measures, polytope vertices.

1 Introduction

Fuzzy measures (also known as capacities or non-additive measures) are a generalization of probability distributions. More concretely, they are measures in which the additivity axiom has been relaxed to a monotonicity condition. This extension is needed in many practical situations, in which additivity is too restrictive. Fuzzy measures have proved themselves to be a powerful tool in many different fields, as Decision Theory ([1], [2]), Game Theory ([3], [4]), and many others (see, for example, [5],[6]).

However, despite all these advantages, the practical application of fuzzy measures is limited by the increased complexity of the measure. If we have a finite space of cardinality n, only $n-1$ values are needed in order to completely determine a probability, while 2^n-2 coefficients are needed to define a fuzzy measure on the same referential. This exponential growth is the actual *Achilles' heel* of fuzzy measures and translates in big complexity when identifying the fuzzy measure modelling a situation.

With the aim of reducing this complexity several of subfamilies have been defined. In these families some extra restrictions are added in order to decrease the number of coefficients but trying to keep the modelling capabilities of the measures. Examples of subfamilies include the λ-measures [7], the k-intolerant measures [8, 9], the p-symmetric measures [10], the decomposable measures [11], etc. In this paper we will focus on the complexity of k-additive measures (see Definition 9 below) introduced by Grabisch in [4].

Consider a situation that can be modeled by a k-additive measure (possibly through Choquet integral [12]); to this extent, an axiomatic characterization of such a model can been found in [13]. Next step is the identification of the measure. Suppose that sample information is available. Using Decision Theory terminology, this information consists in a collection of objects whose overall score and valuation on each criterium are known. We assume that this information is numerical; otherwise, we should previously apply a tool to transform ordinal into numerical data, such as MACBETH [14] or TOMASO [15]. The desired measure (that might not be unique) is the one that best fits the data.

If the considered proximity criterion is the squared error, different techniques exist to solve the problem [16, 17], leading to a problem whose complexity is very high. In a previous work [18], we have proposed a learning method based on genetic algorithms [19]. In this method, the crossover operator is the convex combination (as it will become clearer below, the set of fuzzy measures being at most k-additive forms a convex polytope). This operator is simple and natural and allows to decrease the computational cost. However, it has the drawback that the search regions are embedded one in other. This can be overcome with a suitable mutation operator, but even in that case, the search region is determined by the initial population. Thus, this initial population must be wide enough to include the solution and, obviously, the best option is to select the set of vertices of the polytope.

In this paper we study this set of vertices for the family of k-additive fuzzy measures. The results in the paper seem to mean that the number of k-additive vertices is even greater than the number of vertices for general fuzzy measures. The rest of the paper is organized as follows. In Section 2 we briefly introduce definitions and results that will be used throughout the paper. In Section 3 we provide an algorithm to determine the vertices of the polytope of k-additive (at most) measures from those of the general fuzzy measures and show some examples of the results of the computation of the algorithm. Section 4 is devoted to the study of lower bounds on the number of vertices when $k=n-1$. Finally, in Section 5 some conclusions are drawn and ideas for future work are presented.

2 Basic concepts and previous results

Consider a finite referential set $X=\{1,...,n\}$ of n elements or *criteria*. Let us denote by $\mathcal{P}(X)$ the set of subsets of X. Subsets of X are denoted $A,B,...$ and also by $A_1,A_2,....$

Definition 1 *[12, 20, 21] A* **fuzzy measure**, **non-additive measure** *or* **capacity** *over X is a function $\mu:\mathcal{P}(X)\to[0,1]$ satisfying*

- $\mu(\emptyset)=0$, $\mu(X)=1$ *(boundary conditions).*

ISBN: 978-989-95079-6-8

- $\forall A, B \in \mathcal{P}(X)$, *if* $A \subseteq B$, *then* $\mu(A) \leq \mu(B)$ *(monotonicity).*

We will denote the set of all fuzzy measures over X by $\mathcal{FM}(X)$. Remark that $\mathcal{FM}(X)$ is a bounded convex polyhedron in $\mathbb{R}^{2^n-2}$ (or $\mathbb{R}^{2^n}$ if we include coordinates for X and $\emptyset$).

We can define a partial order on $\mathcal{FM}(X)$ in the following way:

Definition 2 *Let μ_1 and μ_2 be two fuzzy measures. Then we say that $\mu_1 \leq \mu_2$ if for every $A \subseteq X$, it holds $\mu_1(A) \leq \mu_2(A)$.*

A special class of fuzzy measures is the set of $\{0,1\}$-valued measures.

Definition 3 *A fuzzy measure is* **$\{0,1\}$-valued** *if it only takes values 0 and 1.*

Notice that for a $\{0,1\}$-valued measure μ, there are some subsets A satisfying the following conditions:

$$\begin{array}{ll} \mu(A) = 1, & \\ \mu(B) = 1, & \forall B \supseteq A, \\ \mu(C) = 0, & \forall C \subset A. \end{array}$$

We will call any subset satisfying these conditions a **minimal subset** for μ. Note that a $\{0,1\}$-valued measure is completely defined by its minimal subsets. Let us denote by $\mu_{\mathbf{C}}$ the fuzzy measure whose collection of minimal sets is $\mathbf{C}$.

An important instance of $\{0,1\}$-valued measures are the so-called unanimity games:

Definition 4 *A* **unanimity game** *over $A \subseteq X$, $A \neq \emptyset$ is a fuzzy measure defined by*

$$u_A(B) := \begin{cases} 1 & \text{if A} \subseteq \text{B} \\ 0 & \text{otherwise} \end{cases}$$

For $\emptyset$, we define the unanimity game by

$$u_\emptyset(B) := \begin{cases} 1 & \text{if B} \neq \emptyset \\ 0 & \text{if B} = \emptyset \end{cases}$$

Definition 5 *Let μ be a fuzzy measure over X; we define the* **dual measure of** *μ as the fuzzy measure $\overline{\mu}$ given by $\overline{\mu}(A) = 1 - \mu(A^c)$.*

In order to define the *extreme points* or *vertices* of a family of fuzzy measures we need to define the convex combination of measures.

Definition 6 *Given $\lambda \in [0,1]$ and two fuzzy measures μ_1 and μ_2 the fuzzy measure*

$$\mu = \lambda\mu_1 + (1-\lambda)\mu_2$$

is said to be a **convex combination** *of μ_1 and μ_2. A subset of fuzzy measures $\mathcal{F} \subseteq \mathcal{FM}(X)$ is* **convex** *if it contains every convex combination of its members.*

Definition 7 *Given a convex subset of fuzzy measures, $\mathcal{F} \subseteq \mathcal{FM}(X)$, we say that $\mu \in \mathcal{F}$ is a* **vertex** *or* **extreme point** *of $\mathcal{F}$ if μ can not be written as a convex combination of two measures $\mu_1, \mu_2 \in \mathcal{F}$ both different from μ.*

Obviously, the set of vertices completely determines the convex polytope $\mathcal{FM}(X)$, as any point in it can be written as a convex combination of the vertices. The vertices of $\mathcal{FM}(X)$ are given in the following result.

Proposition 1 *[22] The set of $\{0,1\}$-valued measures constitutes the set of vertices of $\mathcal{FM}(X)$.*

A similar result (see [23]) can be proved for the family of p-symmetric measures [10]. However, it does not hold [23] for the family of measures that we study in this paper: the k-additive measures, which are defined through the concept of Möbius transform.

Definition 8 *[6, 24] Let μ be a set function (not necessarily a fuzzy measure) on X. The* **Möbius transform (or inverse)** *of μ is another set function on X defined by*

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B), \, \forall A \subseteq X. \tag{1}$$

The Möbius transform given, the original set function can be recovered through the *Zeta transform* [25]:

$$\mu(A) = \sum_{B \subseteq A} m(B). \tag{2}$$

The value $m(A)$ represents the strength of the subset A in any superset which it appears. Remark that the Möbius transform can attain negative values; when it is a non-negative function, it corresponds to the *basic probability mass assignment* in Dempster-Shafer theory of evidence [26].

In order to determine a fuzzy measure, $2^n - 2$ values are necessary. The number of coefficients grows exponentially with n and so does the complexity of the problem of identifying the fuzzy measure. This drawback reduces considerably the practical use of fuzzy measures. Thus, some subfamilies of fuzzy measures have been defined in an attempt to reduce complexity, prominently k-additive measures [4].

Definition 9 *[4] A fuzzy measure μ is said to be* **k-order additive** *or* **k-additive** *if its Möbius transform vanishes for any $A \subseteq X$ such that $|A| > k$ and there exists at least one subset A with exactly k elements such that $m(A) \neq 0$.*

In this sense, a probability measure is just a 1-additive measure. Thus, k-additive measures generalize probability measures, that are very restrictive in many situations. They fill the gap between probability measures and general fuzzy measures. For a k-additive measure, the number of coefficients is reduced to

$$\sum_{i=1}^{k} \binom{n}{i}.$$

More about k-additive measures can be found e.g. in [2]. We will denote the set of all k'-additive measures on X with $k' \leq k$ by $\mathcal{FM}^k(X)$; we will use the fact that $\mathcal{FM}^k(X)$ is a convex polyhedron (the proof is straightforward considering the Möbius transform). Specially appealing is the 2-additive case, that provides a generalization of probability allowing interactions while keeping a reduced complexity.

In this paper we will be studying some properties of the extreme points of the k-additive measures. The following results are known:

ISBN: 978-989-95079-6-8

Theorem 1 *[23] There are vertices of the set $\mathcal{FM}^k(X), k > 2$, that are not $\{0,1\}$-valued measures.*

Proposition 2 *[23] The set of extreme points of $\mathcal{FM}^1(X)$ (resp. $\mathcal{FM}^2(X)$) are the $\{0,1\}$-valued measures that are in $\mathcal{FM}^1(X)$ (resp. $\mathcal{FM}^2(X)$).*

We will also need some results relating the group of isometries of $\mathcal{FM}(X)$ in next section.

Definition 10 *Let $\mathcal{F}$ be a family of fuzzy measures. A surjective function $f : \mathcal{F} \to \mathcal{F}$ is an* **isometry** *if*

$$d(\mu_1, \mu_2) = d(f(\mu_1), f(\mu_2)), \forall \mu_1, \mu_2 \in \mathcal{F}.$$

Remark that an isometry is a bijective mapping on $\mathcal{F}$ and that these isometries form a group under composition of functions. It can be seen [27] that isometries also map vertices into vertices. Let us denote by $\mathcal{G}(\mathcal{F})$ the group of isometries of $\mathcal{F}$.

In [27] we have determined the group of isometries of general fuzzy measures (a result which was later generalized on [28]). We need some previous definitions.

Definition 11 *Consider $\sigma : X \to X$ a permutation on X. We define the* **symmetry induced by** *σ, denoted S_σ, the transformation on $\mathcal{FM}(X)$ such that for any $\mu \in \mathcal{FM}(X)$, the fuzzy measure $S_\sigma(\mu)$ is defined by*

$$S_\sigma(\mu)(A) = \mu(\sigma(A)), \forall A \subseteq X.$$

Definition 12 *We define the* **dual transformation***, denoted D, the transformation on $\mathcal{FM}(X)$ given by*

$$D : \begin{array}{ccc} \mathcal{FM} & \to & \mathcal{FM} \\ \mu & \mapsto & \overline{\mu} \end{array}.$$

In most cases, the group of isometries of the general fuzzy measures is generated by symmetries and the dual transformation.

Theorem 2 *If $|X| > 2$, the group $\mathcal{G}(\mathcal{FM}(X))$ is given by symmetries and compositions of symmetries with the dual application.*

In fact, it can be seen that $\mathcal{G}(\mathcal{FM}(X))$ is the semi-direct product of the group of symmetries with the cyclic group or order 2 generated by the dual transformation, i.e. the group of isometries is given by a composition of symmetries composed with either the dual application or the identity map.

Theorem 3 *If $n = 2$, the group $\mathcal{G}(\mathcal{FM}(X))$ is isomorphic to the dihedral group D_4 (the group of isometries of the square).*

Finally, we define the concept of adjacency of extreme points of $\mathcal{FM}(X)$, which will be central in our study of the vertices of $\mathcal{FM}^k(X)$.

Definition 13 *Two vertices μ_1 and μ_2 of $\mathcal{FM}(X)$ are adjacent if the convex combination (their midpoint)*

$$\mu = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$$

can not be written as a convex combination of extreme points in any other way.

The following results have been proved in [29] and later generalized in [30].

Lemma 1 *If μ_1 and μ_2 are adjacent vertices then $\mu_1 \leq \mu_2$ or $\mu_2 \leq \mu_1$.*

However, this condition is not sufficient. The characterization of adjacency in $\mathcal{FM}(X)$ is based on the concept of $\mathcal{C}-$decomposability, that is given in next definition.

Definition 14 *Let $\mathbf{C}$ be a collection of subsets of X, μ an extreme point and let $A_1, \ldots, A_m$ be all its minimal sets. We say that μ is* **C-decomposable** *if there exists a partition of $\{A_1, \ldots, A_m\}$ in two non-empty subsets $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{A} \not\subseteq \mathbf{C}$ and $\mathbf{B} \not\subseteq \mathbf{C}$ and if $A \in \mathbf{A}$ and $B \in \mathbf{B}$ then there exists $C \in \mathbf{C}$ such that $C \subseteq A \cup B$.*

The following theorem characterizes adjacency on $\mathcal{FM}(X)$.

Theorem 4 *Let $\mu, \mu_\mathcal{C}$ be two vertices of $\mathcal{FM}(X)$ such that $\mu \geq \mu_\mathcal{C}$. Then μ and $\mu_\mathbf{C}$ are adjacent if and only if μ is not $\mathbf{C}$-decomposable.*

3 Generating the vertices of the k-additive measures

Consider the set $\mathcal{FM}(X)$. From a geometrical point of view, $\mathcal{FM}^k(X)$ is a subpolytope of $\mathcal{FM}(X)$ with additional restrictions; namely, these restrictions are $m(A) = 0, \forall |A| > k$.

Let $|A| > k$ and consider the restriction $m(A) = 0$. Then, it can be proved that the set of vertices of the polytope of fuzzy measures with this new condition consists in (see [31]):

1. Those $\{0,1\}$-valued measures for which it is $m(A) = 0$ and
2. the cut points of the hyper-plane $m(A) = 0$ with the edges of the polytope $\mathcal{FM}(X)$.

If we repeat the procedure imposing restrictions for every set A with size bigger than k we eventually obtain all the vertices of the polytope $\mathcal{FM}^k(X)$. Thus, we obtain a characterization of the extreme points of the k-additive measures (and of all intermediate polytopes). If $\mathbf{A}$ is a collection of subsets of X we denote by $\mathcal{FM}_\mathbf{A}(X)$ the set of fuzzy measures whose Möbius transform is zero on every set belonging to $\mathbf{A}$. The following theorem, well-known in the theory of polytopes, provides the new vertices when adding a new constraint.

Theorem 5 *Let $\mathbf{A}$ be a collection of subsets of X (eventually empty), B a subset of X not in $\mathbf{A}$ and μ a fuzzy measure of $\mathcal{FM}_{\mathbf{A}\cup\{B\}}(X)$ which is not a vertex of $\mathcal{FM}_\mathbf{A}(X)$. Then μ is a vertex of $\mathcal{FM}_{\mathbf{A}\cup\{B\}}(X)$ if and only if there exist μ_1 and μ_2 two adjacent vertices of $\mathcal{FM}_\mathbf{A}(X)$ not in $\mathcal{FM}_{\mathbf{A}\cup\{B\}}(X)$ and $\lambda \in (0,1)$ such that $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$.*

As a consequence of this result, for the case of $(n-1)$-additive measures we have the following:

Theorem 6 *Let μ be a $(n-1)$-additive measure which is not $\{0,1\}$-valued. Then μ is a vertex of $\mathcal{FM}^{n-1}(X)$ if and only*

$$\mu = \lambda\mu_1 + (1-\lambda)\mu_2,$$

where $\lambda \in (0,1)$ and μ_1, μ_2 are two adjacent vertices of $\mathcal{FM}(X)$ not in $\mathcal{FM}^{n-1}(X)$.

ISBN: 978-989-95079-6-8

In this particular case we can obtain all vertices of $\mathcal{FM}^{n-1}(X)$ just from adjacency relationships in $\mathcal{FM}(X)$. However, this is not possible in general; for other values of k there exist vertices not $\{0,1\}$-valued which cannot be expressed as a convex combination of exactly two $\{0,1\}$-valued vertices, as the following example shows.

Example 1 *Consider $|X| = 5$. Let μ_1 be the $\{0,1\}$-valued measure whose minimal sets are the singletons $\{1\},\ldots,\{5\}$ and μ_2 the fuzzy measure whose minimal sets are the sets of size 4. Also, let μ_3 be the fuzzy measure with minimal sets $\{(1,2),(1,3),(2,4),(3,5),(4,5)\}$.*

From Theorem 4 it follows that μ_1 and μ_2 are adjacent vertices in $\mathcal{FM}(X)$. Analogously, μ_2 and μ_3 are also adjacent vertices.

It is easy to check that

$$\frac{4}{5}\mu_1 + \frac{1}{5}\mu_2, \frac{4}{5}\mu_3 + \frac{1}{5}\mu_2 \in \mathcal{FM}^4(X). \tag{3}$$

Thus, these measures are extreme points of $\mathcal{FM}^4(X)$ and are not $\{0,1\}$-valued. Let us denote them by μ and μ' respectively. It can be checked that these measures are adjacent in $\mathcal{FM}^4(X)$. In fact, it easy to verify that their midpoint satisfies with equality exactly $2^n - 3$ linearly independent restrictions of those defining the polytope. Moreover,

$$\frac{1}{4}\mu + \frac{3}{4}\mu' \in \mathcal{FM}^3(X). \tag{4}$$

Thus, we have obtained a vertex of the 3-additive measures, which requires three vertices of $\mathcal{FM}(X)$ in order to obtain a convex combination which generates it.

As an application of Theorem 5 we have computed the vertices of the polytopes $\mathcal{FM}^k(X)$ for some values of k and n. The following tables show the number of vertices and the number of vertices being $\{0,1\}$-valued.

Table 1: Number of vertices of the $\mathcal{FM}^k(X)$ polytopes.

$n\backslash k$	1	2	3	4	5	6
1	1					
2	2	4				
3	3	9	18			
4	4	16	303	166		
5	5	25	584740	407201	7579	
6	6	36	?	?	232871070690	7828352

Table 2: Number of $\{0,1\}$-valued vertices of the $\mathcal{FM}^k(X)$ polytopes.

$n\backslash k$	1	2	3	4	5	6
1	1					
2	2	4				
3	3	9	18			
4	4	16	68	166		
5	5	25	195	1855	7579	
6	6	36	456	10986	1322954	7828352

To compute these numbers for the cases with $n \leq 5$ we repeatedly use Theorem 5 to calculate the vertices of the intermediate polytopes. This process is very time-consuming and, in fact, took several weeks in a 2.1GHz PC to complete the calculations in the case $n = 5$. One of the main difficulties is that we lack criteria to quickly test adjacency (except for general fuzzy measures, where we can use Theorem 4). The other main problem is that the number of extreme points grows very fast, making the number of comparisons needed grow even faster. All this makes unfeasible a direct approach to calculate the number of vertices of the k-additive measures when $n = 6$.

However, it is possible to use the results on the isometries of the polytope of fuzzy measures (see Section 2) to at least obtain the number of vertices of the 5-additive measures when $n = 6$. Just observe that if σ is a permutation of X then S_σ, the symmetry induced by σ (Definition 11), is an isometry that fixes u_X and if μ is a fuzzy measure then $m_\mu(X)$ and $m_{S_\sigma(\mu)}(X)$ are both either positive, negative or zero, by Eq. (1). Also, as S_σ is an isometry, if μ_1 and μ_2 are adjacent vertices, then so are $S_\sigma(\mu_1)$ and $S_\sigma(\mu_2)$. Consequently, if $\lambda\mu_1 + (1-\lambda)\mu_2$ is a vertex of the $n-1$-additive measures then so is $\lambda S_\sigma(\mu_1) + (1-\lambda)S_\sigma(\mu_2)$.

Take μ, μ' two $\{0,1\}$-valued measures such that there exists a symmetry S_σ for which $S_\sigma(\mu) = \mu'$. If μ_1 is a $\{0,1\}$-valued measure adjacent to μ and $\alpha \in (0,1)$ such that

$$\alpha\mu + (1-\alpha)\mu_1 \in \mathcal{FM}^{n-1}(X),$$

then

$$S_\sigma(\alpha\mu + (1-\alpha)\mu_1) = \alpha S_\sigma(\mu) + (1-\alpha)S_\sigma(\mu_1)$$

also belongs to $\mathcal{FM}^{n-1}(X)$. Consequently, every vertex of $\mathcal{FM}^{n-1}(X)$ obtained from μ' is the image under S_σ of another vertex obtained from μ.

Thus, if μ and μ' are taken one into the other by some symmetry S_σ (i.e., they are in the same orbit under the action of the group of symmetries), then they will generate exactly the same number of vertices of $\mathcal{FM}^{n-1}(X)$. Therefore, in order to count the number of vertices of the $n-1$-additive measures, we can proceed as follows:

1. List all the vertices of the general fuzzy measures.
2. Classify these vertices according to the sign of their Möbius transform on X.
3. Calculate the orbits of the vertices with positive Möbius on X.
4. Pick a representative of each orbit and test its adjacency to the vertices with negative Möbius on X.
5. Multiply the number of adjacent vertices by the size of the orbit, and sum all the values obtained.
6. Add to this value the number of vertices whose Möbius transform on X is 0.

After applying these methods in the case where $n = 6$ we obtained 1322954 vertices with zero Möbius transform on X,

ISBN: 978-989-95079-6-8

3252699 with negative Möbius and also 3252699 with positive Möbius. These last vertices are organized into 6806 different orbits. After 20 hours of computation on a 2.1GHz PC, the number of non $\{0,1\}$-valued vertices of $\mathcal{FM}^5(X)$ with $|X| = 6$ was found to be 232869747736, for a total of 232871070690 vertices.

The figures in Tables 1 and 2 seem to suggest that the number of non $\{0,1\}$-valued vertices of $\mathcal{FM}^k(X)$ for $k = 3, ..., n-1$ grows much more quickly than the number of extreme points of $\mathcal{FM}(X)$. In the following Section we show that, in fact, this is the case when $k = n-1$.

4 On the number of vertices of the $n-1$-additive measures

In this section we study the asymptotic behavior of the ratio of the number of vertices of $\mathcal{FM}^{n-1}(X)$ to the number of vertices of $\mathcal{FM}(X)$. We will denote by D_n the number of vertices of $\mathcal{FM}(X)$ (with $n = |X|$), by A_n the number of vertices of $\mathcal{FM}^{n-1}(X)$, and by B_n the number of vertices of $\mathcal{FM}(X)$ which are not in $\mathcal{FM}^{n-1}(X)$ and whose minimal sets all have size at least $\lceil \frac{n-3}{2} \rceil$.

Lemma 2 *It holds that*

$$\lim_{n\to\infty} \frac{B_n}{D_n} > 0$$

Sketch of Proof:

The proof of the Theorem rests on considering, for even n, the set M_0^n of vertices μ of $\mathcal{FM}(X)$ such that:

1. every minimal set of μ has size $\frac{n}{2}-1$, $\frac{n}{2}$ or $\frac{n}{2}+1$,
2. $\mu(A) = 1$ for every A of size greater than $\frac{n}{2}+1$,
3. $\mu(A) = 0$ for every A of size less than $\frac{n}{2}-1$,
4. the number of minimal sets of μ of size $\frac{n}{2}-1$ is at most $2^{\frac{n}{2}}$.

In [32] it is proven that the ratio of the number of measures in M_0^n to the total number of measures does not tend to zero when n tends to infinity. It can also be seen that the ratio of non $(n-1)$-additive measures in M_0^n does not tend to 0 when n tends to infinity and the Lemma follows (for even n).

For odd n, the argument is similar, but considering the set M_1^n of vertices μ of $\mathcal{FM}(X)$ such that:

1. every minimal set of μ has size $\frac{n-3}{2}$, $\frac{n-1}{2}$ or $\frac{n+1}{2}$,
2. $\mu(A) = 1$ for every A of size greater than $\frac{n+1}{2}$,
3. $\mu(A) = 0$ for every A of size less than $\frac{n-3}{2}$,
4. the number of minimal sets of μ of size $\frac{n-3}{2}$ is at most $2^{\frac{n}{2}}$.

■

We can now state the main Theorem of the section.

Theorem 7 *There exist $k > 0$ and n_0 such that*

$$A_n > k\frac{2^n}{\sqrt{n}} D_n$$

for all $n \geq n_0$.

Sketch of Proof:

Denote $q = \lceil \frac{n-3}{2} \rceil$. Consider n such that $q > 4$. Consider $\mathcal{P}_1$ the collection of all partitions P of X such that

1. exactly one set in P has size $q-2$,
2. the rest of the sets in P are singletons,

and $\mathcal{P}_2$ the collection of all partitions P of X such that

1. exactly one set in P has size $q-3$,
2. the rest of the sets in P are singletons.

Consider μ a vertex of $\mathcal{FM}(X)$ not in $\mathcal{FM}^{n-1}(X)$ whose minimal sets all have size at least q. Consider P in $\mathcal{P}_1 \cup \mathcal{P}_2$ and μ' the measure whose collection of minimal sets is P. It is possible to show, with the help of Theorem 4, that μ and μ' are adjacent.

The size of $\mathcal{P}_1$ is $\binom{n}{q-2}$ and the size of $\mathcal{P}_2$ is $\binom{n}{q-3}$. Also, the sizes of the partitions in $\mathcal{P}_1$ and of the partitions in $\mathcal{P}_2$ have different parity, so their Möbius transform on X will be 1 in one case and -1 in the other (see Proposition 4 in [23]). Since μ is not $(n-1)$-additive and it is adjacent to every μ' whose minimal sets are in $\mathcal{P}_1 \cup \mathcal{P}_2$, then the convex combinations of μ with all such μ' generate at least $\binom{n}{q-3}$ vertices of $\mathcal{FM}^{n-1}(X)$. It follows that

$$A_n \geq \binom{n}{q-3} B_n$$

for all n such that $q > 4$.

From Lemma 2 we know that there exist $k_1 > 0$ and n_1 such that

$$\frac{B_n}{D_n} > k_1$$

for all $n \geq n_1$. Then, it is easy to prove that there exists $k > 0$ such that for n big enough

$$A_n \geq k\frac{2^n}{\sqrt{n}} D_n$$

■

This Theorem shows that the growth of the number of vertices of $\mathcal{FM}^{n-1}(X)$ is much faster than that of $\mathcal{FM}(X)$. Theorem 5 together with the values on Table 1 hint that this may be the case also for other polytopes $\mathcal{FM}^k(X)$ with k big enough. This suggests that, despite the simple interpretation, k-additive measures have a complex structure and the choice of the value of k should be made with care.

5 Conclusions and open problems

In this paper we have studied the vertices of the polytope of k-additive measures. We have provided a description of these vertices and a algorithm to compute them. We have also shown some examples of the results of the computation of this algorithm. Finally, we have given a lower bound on the number of vertices of the polytope of fuzzy measures which are at most $(n-1)$-additive, showing that this number grows much faster than the number of vertices of the general fuzzy measures. Similar results for other values of k seem plausible and deeper investigation of the asymptotical behavior on

ISBN: 978-989-95079-6-8

those cases might provide further insights on the polytopes $\mathcal{FM}^k(X)$.

The results presented here show that, in general, the structure of $\mathcal{FM}^k(X)$ is more complex than that of the general fuzzy measures and of the p-symmetric measures (cf.[29, 30]). For instance, knowledge of a wide number of intermediate polytopes was needed in order to compute the extreme points of the k-additive measures, making this computation very time consuming. A study of conditions for the adjacency in the polytope of k-additive measures (and in the intermediate ones) could help to decrease this computing time. Knowing the group of isometries of the polytopes may also be useful.

We also intend to investigate subfamilies of the k-additive measures which retain the modeling power of these measures. For instance, it is interesting to restrict the vertices to those which are $\{0, 1\}$-valued and study the resulting polytope.

Finally, in order to obtain a suitable initial population (for measure identification with a genetic algorithm) it would be useful to study methods for the random generation of k-additive measures.

Acknowledgements

This research has been supported in part by grant numbers MTM2007-61193 and CAM- UCM910707, and by MEC and FEDER grant TIN2007-61273.

References

[1] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5:587–607, 1997.

[2] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, (92):167–189, 1997.

[3] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the theory of Games*, volume II of *Annals of Mathematics Studies*, pages 307–317. Princeton University Press, 1953.

[4] M. Grabisch. k-order additive discrete fuzzy measures. In *Proceedings of 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1345–1350, Granada (Spain), 1996.

[5] P. L. Hammer and R. Holzman. On approximations of pseudo-boolean functions. *Zeitschrift für Operations Research. Mathematical Methods of Operations Research*, (36):3–21, 1992.

[6] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, (2):340–368, 1964.

[7] M. Sugeno and T. Terano. A model of learning based on fuzzy information. *Kybernetes*, (6):157–166, 1977.

[8] J.-L. Marichal. k-intolerant capacities and Choquet integrals. In *Proceedings of Tenth International Conference of Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, pages 601–608, Perugia (Italy), July 2004.

[9] J.-L. Marichal. Tolerant or intolerant character of interacting criteria in aggregation by the Choquet integral. *European Journal of Operational Research*, 155(3):771–791, 2004.

[10] P. Miranda, M. Grabisch, and P. Gil. p-symmetric fuzzy measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 (Suppl.):105–123, 2002.

[11] D. Dubois and H. Prade. A class of fuzzy measures based on triangular norms. *Int. J. General Systems*, 8:43–61, 1982.

[12] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, (5):131–295, 1953.

[13] P. Miranda, M. Grabisch, and P. Gil. Axiomatic structure of k-additive capacities. *Mathematical Social Sciences*, 49:153–178, 2005.

[14] C. A. Bana e Costa and J.-C. Vansnick. MACBETH- An interactive path towards the construction of cardinal value function. *International Transactions in Operational Research*, 1(4):489–500, 1994.

[15] J.-L. Marichal, P. Meyer, and M. Roubens. Sorting multiattribute alternatives : the TOMASO method. *Computer and Operations Resarch*, 32(4):861–877, 2005.

[16] M. Grabisch and J.-M. Nicolas. Classification by fuzzy integral-performance and tests. *Fuzzy Sets and Systems, Special Issue on Pattern Recognition*, (65):255–271, 1994.

[17] J. Wang Z. Wang, K. Xu and G. Klir. Using genetic algorithms to determine nonnegative monotone set functions for information fusion in environments with random perturbation. *International Journal of Intelligent Systems*, 14:949–962, 1999.

[18] E. F. Combarro and P. Miranda. Identification of fuzzy measures from sample data with genetic algorithms. *Computers and Operations Research*, 33(10):3046–3066, 2006.

[19] J. H. Holland. *Adaptation in natural and artificial systems*. Ann Arbor: The University of Michigan Press, 1975.

[20] D. Denneberg. *Non-additive measures and integral*. Kluwer Academic, 1994.

[21] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[22] D. Radojevic. The logical representation of the discrete Choquet integral. *Belgian Journal of Operations Research, Statistics and Computer Science*, 38(2–3):67–89, 1998.

[23] P. Miranda, E.F. Combarro, and P. Gil. Extreme points of some families of non-additive measures. *European Journal of Operational Research*, 33(10):3046–3066, 2006.

[24] J. H. van Lint and R. M. Wilson. *A course in Combinatorics*. Cambridge University Press, 1998.

[25] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, (17):263–283, 1989.

[26] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, (USA), 1976.

[27] P. Miranda and E.F. Combarro. On the structure of some families of fuzzy measures. *IEEE Transactions on Fuzzy Systems*, 15(6):1068–1081, 2007.

[28] E. F. Combarro and P. Miranda. Adjacency on the order polytope with applications to the theory of fuzzy measures. *Submitted*.

[29] E. F. Combarro and P. Miranda. On the polytope of non-additve measures. *Fuzzy Sets and Systems*, 159(16):2145–2162, 2008.

[30] E. F. Combarro and P. Miranda. Characterizing the isometries on the order polytope with an application to the theory of fuzzy measures. *Submitted*.

[31] K. G. Murty. Adjacency on convex polyhedra. *SIAM Review*, 13(3):377–386, 1971.

[32] A. D. Korshunov. Monotone boolean functions. *Russian Mathematical Survey*, 58(5):929–1001, 2003.

Type-2 Fuzzy Trust Aggregation

John T. Rickard[1] Sam Hamilton[2] Wendy Hamilton[3]

1 Distributed Infinity, Inc., 4637 Shoshone Drive,
Larkspur, CO 80118, USA
2, 3 Distributed Infinity, Inc., 4033 Nabal Drive
La Mesa, CA 91941, USA
Email: {trickard, shamilton, whamilton@distributedinfinity.com}

Abstract— *Trust is a fundamental concept that is critical in human decision processes in almost all domains, but of particular relevance in the domain of computer security. In many computer systems deployed today, trust is only modeled indirectly through a serious of formal rules and regulations describing when privileges are to be granted or revoked. Recently, research has been conducted into the area of fuzzy trust modeling in order to allow a more intelligent tradeoff analysis by the computer security software. In this paper, we extend this work to support Type-2 fuzzy reasoning. The Type-2 input membership functions can be derived from multiple human experts' judgments. This allows us to incorporate the inherent imprecision in trust judgments in a mathematically principled manner.*

Keywords— Computer security, fuzzy systems, trust modeling, Type-2 fuzzy logic.

1 Introduction

In computer security, trust is described as the belief in the competence of an entity to act dependably, reliably and securely within a specific context, and distrust is described as the belief in the incompetence of an entity to act dependably, securely and reliably within a specific context [1]. In many computer systems deployed today, trust is only modeled indirectly through a serious of formal rules and regulations describing when privileges are to be granted or revoked. Manual human judgment is invoked whenever the system falls into an ambiguous state. For example, if a user tries to log on with an incorrect password too many times, many systems will revoke privileges for that user account and rely on the system administrator for the judgment of when it is appropriate to reissue those privileges.

This approach has significant weaknesses in that its strong reliance on a human in the decision process requires the Boolean security rule set to make inappropriately crisp statements, which either overly restrict users (reducing productivity) or are not restrictive enough (reducing security). Recently, research has been conducted into augmenting policy rules with Type-1 fuzzy trust modeling to help address this issue by softening the hard and fast set of computer security policy rules to allow a more intelligent tradeoff analysis [2-7].

The notion of trust relationships is traditionally modeled in a fuzzy sense, i.e., "entity A trusts entity B to a degree $x \in [0,1]$," where x is usually determined by the historical results of entity $A's$ interactions with entity B. Trust chains may also be involved in this determination, e.g., entity A can consult entity C as to its trust in entity B and factor that information into its decision. A discrete set K of contexts may also enter into this description, resulting in an ordered pair (x, y) describing trust, indicating that entity A trusts entity B to a degree $x \in [0,1]$ in a particular context $y \in K$.

The contribution of each context factor may be measured independently using two numbers, a *confidence* $c \in [-1,1]$ and a *plausibility* $p \in [0,1]$. The confidence c describes entity $A's$ belief in a particular outcome either *occurring* or *not occurring* (we use negative values to indicate the latter) in an interaction with entity B in a given context, and the plausibility p describes the degree to which entity A considers her confidence assessment to be reliable. For example, entity A might believe with high confidence (e.g., 0.8) in a particular outcome's occurrence, but if this belief is based upon little or no historical evidence, then her plausibility might be low (say, 0.1). Upon the accumulation of further direct evidence as a result of her interactions with entity B and/or consultation with other entities, she may simultaneously refine her confidence estimate and increase her plausibility in this refined confidence. Thus we treat these as independent variables.

Hamilton and Hamilton [8] developed an algebra for combining such assessments, based upon *singleton values* of input confidences and plausibilities, and producing a corresponding singleton aggregate confidence and plausibility. However, it is well known that human assessments of situations are pervasively imprecise, and thus humans are more reliable at estimating *intervals* of confidence or plausibility rather than point values. Likewise, the results of computer simulations are better described by an *interval* of outcomes than by a single number. Furthermore, it is often desirable to solicit such intervals from multiple experts, or to perform multiple simulations. In this case, the problem remains as to how to incorporate this multiplicity of interval inputs into the trust system. Multiple interval inputs can be combined into interval Type-2 input membership functions using the techniques described in [17].

In this paper, we extend the results of [8] initially to individual interval inputs of confidence and plausibility, to compute corresponding interval output membership functions for their aggregation. Then, employing α-cuts, we further extend these results to accommodate Type-1 and

ISBN: 978-989-95079-6-8

interval Type-2 fuzzy inputs. This allows us to propagate imprecision in our knowledge of the inputs to the corresponding imprecision in our knowledge of the output in a mathematically principled fashion.

2 Trust aggregation

Suppose that we have two assessments of confidence and plausibility relating to our trust in a particular context factor, which we denote by A_{c_1,p_1} and A_{c_2,p_2}, respectively, where c_1 and c_2 denote the confidences, and p_1 and p_2 the plausibilities, of the corresponding factor. Suppose further that we wish to perform an aggregation of these values, denoted by the operator "$\wedge$", to yield a sensible aggregate confidence and plausibility based upon these two inputs. Thus we wish to describe $A_{c_3,p_3} = A_{c_1,p_1} \wedge A_{c_2,p_2}$ in terms of its aggregate confidence c_3 and plausibility p_3.

Hamilton and Hamilton [8] propose the following algebraic computations for c_3 and p_3:

$$c_3 = \frac{c_1 p_1^{\alpha} + c_2 p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}} \tag{1}$$

$$p_3 = \frac{(p_1 + p_2 - p_1 p_2)\left(2 - |c_1 - c_2|\right)^{\beta} + (p_1 + p_2 - 2p_1 p_2)|c_1 - c_2|^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}} \tag{2}$$

The parameters $\alpha > 0$ and $\beta > 0$ govern the degree of influence of the input plausibilities p_1 and p_2 upon the aggregate confidence c_3, and the degree of influence of the *differences* in the two input confidences c_1 and c_2 upon the aggregate plausibility p_3, respectively. Typical values for these parameters might be $\alpha = 2$ and $\beta = 1/4$. It is straightforward to show that when the domains of the inputs are $c_1, c_2 \in [-1,1]$ and $p_1, p_2 \in [0,1]$, the resulting ranges for the aggregate confidence and plausibility are likewise $c_3 \in [-1,1]$ and $p_3 \in [0,1]$.

From (1), we see that the aggregate confidence c_3 is a simple convex combination of the input confidences, with weights proportional to the α-power of the plausibilities. The aggregate plausibility p_3 is a more complicated convex combination of two functions of the input plausibilities. At one extreme, when the input confidences are in perfect agreement, the aggregate plausibility is given by $p_3 = p_1 + p_2 - p_1 p_2$ $(c_1 = c_2)$. At the other extreme where the confidences are diametrically opposite and of unity magnitude, the aggregate plausibility is given by $p_3 = p_1 + p_2 - 2p_1 p_2$, $(c_1 = \pm 1,\ c_2 = -c_1)$.

To illustrate further the behaviors of (1) and (2), consider the following examples, where $\alpha = 2$ and $\beta = 1/4$:

$$A_{0.8,0} \wedge A_{0.3,0.5} = A_{0.3,0.5} \tag{3}$$

$$A_{-0.8,0.9} \wedge A_{0.8,0.9} = A_{0,0.5} \tag{4}$$

$$A_{0.8,0.3} \wedge A_{0.7,0.4} = A_{0.74,0.54} \tag{5}$$

In the first case, an assessment with zero plausibility adds no new information. In the second case, strongly opposing confidences reduce aggregate plausibility. In the final case, similar confidences reinforce aggregate plausibility.

A single fuzzy scalar value of our degree of trust $t\left(A_{c,p}\right)$ can be computed as the following convex combination involving confidence and plausibility:

$$t\left(A_{c,p}\right) = \frac{0.5(1-p)^{\gamma} + \frac{(1+c)}{2} p^{\gamma}}{(1-p)^{\gamma} + p^{\gamma}} \tag{6}$$

where $\gamma > 0$ controls the degree of influence of the plausibility upon the trust. A typical value might be $\gamma = 2$. Note that if $p = 0$, then $t\left(A_{c,p}\right) = 0.5$, i.e., we have a maximally ambiguous degree of trust (i.e., the "coin toss" situation), whereas if $p = 1$, then $t\left(A_{c,p}\right) = \frac{1+c}{2}$, (i.e., trust is equivalent to a rescaling of the confidence c to lie in the interval [0,1].) In the extremes of this latter case, if $c = -1$, we have zero trust in the given outcome, while if $c = 1$, we have complete trust in the given outcome.

3 Interval inputs

Suppose now that, rather than the point values of the inputs described above, we instead have interval values. Interval inputs might naturally arise from soliciting them from a human expert, or as the results of a computer simulation that includes the uncertainty of the simulations results. Specifically, let $c_1 \in [c_{1\ell}, c_{1r}]$, $c_2 \in [c_{2\ell}, c_{2r}]$, $p_1 \in [p_{1\ell}, p_{1r}]$ and $p_2 \in [p_{2\ell}, p_{2r}]$, where the subscripts ℓ and r refer to the left- and right-hand endpoints of the intervals, respectively. Bear in mind that the c intervals are subsets of $[-1,1]$, and the p intervals are subsets of $[0,1]$.

3.1 Confidence aggregation for interval inputs

From (1), c_3 is an interval weighted average [9], where the input variables take values continuously over interval ranges. Thus c_3 takes values continuously over an interval range, which we denote by $c_3 \in [c_{3\ell}, c_{3r}]$, where

$$c_{3\ell} = \min_{c_1,c_2,p_1,p_2} \frac{c_1 p_1^{\alpha} + c_2 p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}} = \min_{p_1,p_2} \frac{c_{1\ell} p_1^{\alpha} + c_{2\ell} p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}}, \tag{7}$$

$$c_{3r} = \max_{c_1,c_2,p_1,p_2} \frac{c_1 p_1^{\alpha} + c_2 p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}} = \max_{p_1,p_2} \frac{c_{1r} p_1^{\alpha} + c_{2r} p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}}. \tag{8}$$

It is well-known [10] how to solve this simple case of a two-element interval weighted average. We first re-index the triplets $(c_{i\ell}, p_{i\ell}, p_{ir})$ in order of increasing $c_{i\ell}$, i.e., to obtain $c_{1\ell} \le c_{2\ell}$. Then using these re-indexed triplets, $c_{3\ell}$ is given by

$$c_{3\ell} = \frac{c_{1\ell} p_{1r}^{\alpha} + c_{2\ell} p_{2\ell}^{\alpha}}{p_{1r}^{\alpha} + p_{2\ell}^{\alpha}}, \tag{9}$$

which is recognized as a closed-form two-element solution of the Karnik-Mendel (KM) algorithm [10], which is used more generally to find the endpoints of the output interval corresponding to an $n-$element interval weighted average. In similar fashion, after re-indexing the triplets $(c_{ir}, p_{i\ell}, p_{ir})$ in order of increasing c_{ir}, we have

ISBN: 978-989-95079-6-8

$$c_{3r} = \frac{c_{1r} p_{1\ell}^{\alpha} + c_{2r} p_{2r}^{\alpha}}{p_{1\ell}^{\alpha} + p_{2r}^{\alpha}}. \tag{10}$$

Thus we have a closed-form analytical solution for the output interval of $c_3 \in [c_{3\ell}, c_{3r}]$ in (1) when the inputs have interval values.

3.2 Plausibility aggregation for interval inputs

From (2), p_3 is a continuous function of its input values, so it too will take values over an interval $p_3 \in [p_{3\ell}, p_{3r}]$. By inspection of (2), we may re-write this equation as

$$p_3 = \lambda(p_1 + p_2 - p_1 p_2) + (1-\lambda)(p_1 + p_2 - 2p_1 p_2) \tag{11}$$

where $\lambda \in [0,1]$ is a function of c_1 and c_2 given by

$$\lambda(c_1, c_2) = \frac{\left(2 - |c_1 - c_2|\right)^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}}. \tag{12}$$

Taking partial derivatives in (11) with respect to p_1 and p_2, we obtain

$$\frac{\partial p_3}{\partial p_1} = 1 - (2-\lambda) p_2 \quad \begin{cases} > 0 \ \textit{for} \ p_2 < \dfrac{1}{2-\lambda} \\ < 0 \ \textit{for} \ p_2 > \dfrac{1}{2-\lambda} \end{cases}, \tag{13}$$

and

$$\frac{\partial p_3}{\partial p_2} = 1 - (2-\lambda) p_1 \quad \begin{cases} > 0 \ \textit{for} \ p_1 < \dfrac{1}{2-\lambda} \\ < 0 \ \textit{for} \ p_1 > \dfrac{1}{2-\lambda} \end{cases}. \tag{14}$$

Given values of c_1 and c_2, and hence a given value of λ, we see that p_3 is monotonically increasing or decreasing as a function of p_1 and p_2 away from the saddle point $\left(\frac{1}{2-\lambda}, \frac{1}{2-\lambda}\right)$, which lies in the upper right-hand quadrant of the unit square since $\lambda \in [0,1]$. Thus the minimum (maximum) of (11) as a function of p_1 and p_2 always occurs at one of the four input interval extremal points $(p_{1\ell}, p_{2\ell})$, $(p_{1\ell}, p_{2r})$, $(p_{1r}, p_{2\ell})$ or (p_{1r}, p_{2r}).

To continue this analysis, let us define four functions $f_{ij}(c_1, c_2)$, $i \in \{\ell, r\}$, $j \in \{\ell, r\}$ such that

$$f_{ij}(c_1, c_2) = \frac{(p_{1i} + p_{2j} - p_{1i} p_{2j})\left(2 - |c_1 - c_2|\right)^{\beta} + (p_{1i} + p_{2j} - 2p_{1i} p_{2j})|c_1 - c_2|^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}} \tag{15}$$

It then remains to calculate

$$p_{3\ell} = \min_{i,j \in \{\ell, r\}} \left\{ \min_{c_1, c_2 \in [-1,1]} \frac{\pi_{ij}^{(1)}\left(2 - |c_1 - c_2|\right)^{\beta} + \pi_{ij}^{(2)} |c_1 - c_2|^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}} \right\} \tag{16}$$

$$p_{3r} = \max_{i,j \in \{\ell, r\}} \left\{ \max_{c_1, c_2 \in [-1,1]} \frac{\pi_{ij}^{(1)}\left(2 - |c_1 - c_2|\right)^{\beta} + \pi_{ij}^{(2)} |c_1 - c_2|^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}} \right\} \tag{17}$$

where

$$\pi_{ij}^{(1)} = p_{1i} + p_{2j} - p_{1i} p_{2j} \tag{18}$$

$$\pi_{ij}^{(2)} = p_{1i} + p_{2j} - 2p_{1i} p_{2j} \tag{19}$$

To this end, we employ the continuous change of variables $u = |c_1 - c_2|$, which reduces these equations to:

$$p_{3\ell} = \min_{i,j \in \{\ell, r\}} \left\{ \min_{u \in [0,2]} \frac{\pi_{ij}^{(1)}\left(2-u\right)^{\beta} + \pi_{ij}^{(2)} u^{\beta}}{\left(2-u\right)^{\beta} + u^{\beta}} \right\} \triangleq \min_{i,j \in \{\ell, r\}} \left\{ \min_{u \in [0,2]} g_{ij}^{(\min)}(u) \right\} \tag{20}$$

$$p_{3r} = \max_{i,j \in \{\ell, r\}} \left\{ \max_{u \in [0,2]} \frac{\pi_{ij}^{(1)}\left(2-u\right)^{\beta} + \pi_{ij}^{(2)} u^{\beta}}{\left(2-u\right)^{\beta} + u^{\beta}} \right\} \triangleq \max_{i,j \in \{\ell, r\}} \left\{ \max_{u \in [0,2]} g_{ij}^{(\max)}(u) \right\} \tag{21}$$

It is obvious by inspection of (18) and (19) that for any of the four combinations of $(i,j): i,j = 1,2$, $\pi_{ij}^{(1)} \geq \pi_{ij}^{(2)}$, so from (20) and (21) we observe that each of the functions $g_{ij}^{(\min)}(u)$ and $g_{ij}^{(\max)}(u)$ is a convex combination (whose coefficients are nonlinear functions of u over the interval $[0,2]$) of the values $\pi_{ij}^{(1)}$ and $\pi_{ij}^{(2)}$, having its maximum at $u = 0$ and its minimum at $u = 2$. We can show that the derivatives of these functions are uniformly non-positive over the interval $[0,2]$, and thus the functions $g_{ij}^{(\min)}(u)$ and $g_{ij}^{(\max)}(u)$ are uniformly non-increasing over any sub-interval of $[0,2]$. Therefore their minimum with respect to u in (20) over an interval $[u_\ell, u_r] \subseteq [0,2]$ will correspond to u_r, the right endpoint point of this interval, and their maximum with respect to u in (21) will correspond to u_ℓ, the left endpoint of this interval.

To determine the associated interval for u from the intervals $c_1 \in [c_{1\ell}, c_{1r}]$ and $c_2 \in [c_{2\ell}, c_{2r}]$, we must consider six cases of different interval overlap scenarios, depicted in Table 1.

Table 1: Interval overlap cases.

Case	Interval Positions	Interval for $u = \lvert c_1 - c_2 \rvert$
1	$c_{1r} < c_{2\ell}$	$u \in [c_{2\ell} - c_{1r}, c_{2r} - c_{1\ell}]$
2	$c_{1\ell} \leq c_{2\ell} \leq c_{1r} \leq c_{2r}$	$u \in [0, c_{2r} - c_{1\ell}]$
3	$c_{1\ell} \leq c_{2\ell} \leq c_{2r} \leq c_{1r}$	$u \in [0, \max(c_{1r} - c_{2\ell}, c_{2r} - c_{1\ell})]$
4	$c_{2\ell} \leq c_{1\ell} \leq c_{2r} \leq c_{1r}$	$u \in [0, c_{1r} - c_{2\ell}]$
5	$c_{2\ell} \leq c_{2r} \leq c_{1\ell} \leq c_{1r}$	$u \in [c_{1\ell} - c_{2r}, c_{1r} - c_{2\ell}]$
6	$c_{2\ell} \leq c_{1\ell} \leq c_{1r} \leq c_{2r}$	$u \in [0, \max(c_{2r} - c_{1\ell}, c_{1r} - c_{2\ell})]$

As noted above, in all cases, $[u_\ell, u_r] \subseteq [0,2]$, and due to the monotone non-increasing nature of the functions $g_{\min}(u)$ and $g_{\max}(u)$, we finally have

ISBN: 978-989-95079-6-8

$$p_{3\ell} = \min_{i,j\in\{\ell,r\}} \left\{ \min_{u\in[u_\ell,u_r]} g_{ij}^{(\min)}(u) \right\} = \min_{i,j\in\{\ell,r\}} \left\{ \frac{\pi_{ij}^{(1)}\left(2-u_r\right)^\beta + \pi_{ij}^{(2)}u_r^\beta}{\left(2-u_r\right)^\beta + u_r^\beta} \right\}, \quad (22)$$

$$p_{3r} = \max_{i,j\in\{\ell,r\}} \left\{ \max_{u\in[u_\ell,u_r]} g_{ij}^{(\max)}(u) \right\} = \max_{i,j\in\{\ell,r\}} \left\{ \frac{\pi_{ij}^{(1)}\left(2-u_\ell\right)^\beta + \pi_{ij}^{(2)}u_\ell^\beta}{\left(2-u_\ell\right)^\beta + u_\ell^\beta} \right\}. \quad (23)$$

Thus to find the interval boundaries for p_3, we simply must find the minimum (maximum) of the four function evaluations involved in (22) and (23).

3.3 Interval calculations for trust

As a final step in extending our results to interval values, we consider the trust function of equation (6) in the event that $c\in[c_\ell,c_r]$ and $p\in[p_\ell,p_r]$ are described by interval values. Since c appears only in the numerator, it is obvious by inspection that

$$\min t\left(A_{c,p}\right) = \min_{p\in[p_\ell,p_r]} \frac{0.5(1-p)^\gamma + \frac{(1+c_\ell)}{2}p^\gamma}{(1-p)^\gamma + p^\gamma} \quad (24)$$

$$\max t\left(A_{c,p}\right) = \max_{p\in[p_\ell,p_r]} \frac{0.5(1-p)^\gamma + \frac{(1+c_r)}{2}p^\gamma}{(1-p)^\gamma + p^\gamma} \quad (25)$$

Analogous to the functions $g_{ij}^{(\min)}(u)$ and $g_{ij}^{(\max)}(u)$ of the previous sections, we are here dealing with the minimum or maximum of a nonlinear convex combination of the values 0.5 and $(1+c)/2$ over the interval $p\in[p_\ell,p_r]$. We can show that the derivative of (6) with respect to p is monotonic, so that the minimum (maximum) of the above expressions occurs either at $p=p_\ell$ or $p=p_r$ depending upon the value of c (for $c<0$ the derivative is negative and for $c>0$ it is positive), with the result that $t\in[t_\ell,t_r]$, where

$$t_\ell = \min_{p\in[p_\ell,p_r]} t\left(A_{c,p}\right) = \begin{cases} \dfrac{0.5(1-p_r)^\gamma + \frac{(1+c_\ell)}{2}p_r^\gamma}{(1-p_r)^\gamma + p_r^\gamma}, & c_\ell < 0 \\ \dfrac{0.5(1-p_\ell)^\gamma + \frac{(1+c_\ell)}{2}p_\ell^\gamma}{(1-p_\ell)^\gamma + p_\ell^\gamma}, & c_\ell \geq 0 \end{cases} \quad (26)$$

$$t_r = \max_{p\in[p_\ell,p_r]} t\left(A_{c,p}\right) = \begin{cases} \dfrac{0.5(1-p_\ell)^\gamma + \frac{(1+c_r)}{2}p_\ell^\gamma}{(1-p_\ell)^\gamma + p_\ell^\gamma}, & c_r < 0 \\ \dfrac{0.5(1-p_r)^\gamma + \frac{(1+c_r)}{2}p_r^\gamma}{(1-p_r)^\gamma + p_r^\gamma}, & c_r \geq 0 \end{cases} \quad (27)$$

4 Type-1 fuzzy inputs

Suppose that instead of interval membership functions, the fuzzy membership functions for c_1, c_2, p_1 and p_2 in equations (1) and (2) are general convex Type-1 membership functions $\mu_{c_1}(x)$, $\mu_{c_2}(x)$, $\mu_{p_1}(x)$ and $\mu_{p_2}(x)$, respectively, which may not necessarily be normal (i.e., the maximum degree, or height, of the membership function is not necessarily unity). The approach we take in this case is similar to that used to analyze fuzzy weighted averages [9] via α-cuts of the input membership functions, as corrected in [10]. (Note that the term "α-cuts", as commonly used in the fuzzy systems literature, bears no relationship to the α exponent used in equation (1).)

We recall that that an α-cut of a convex fuzzy membership function $\mu(x)$ is an interval on the independent variable axis defined by

$$\left[x(\alpha)_\ell, x(\alpha)_r\right] = \left\{x : \mu(x) \geq \alpha\right\}. \quad (28)$$

We may then represent $\mu(x)$ in terms of its α-cuts by [11]

$$\mu(x) = \sup_{\alpha\in[0,1]} \alpha I_{\mu_\alpha}(x), \quad (29)$$

where $I_{\mu_\alpha}(x)$ is the indicator function for the α-cuts, i.e.,

$$I_{\mu_\alpha}(x) = \begin{cases} 1, & x\in\left[x(\alpha)_\ell, x(\alpha)_r\right] \\ 0, & otherwise \end{cases}. \quad (30)$$

Note that the α-cuts of a fuzzy membership function exist for $0\leq\alpha\leq\alpha_{\max}\leq 1$, where $\alpha_{\max}$ is the height of $\mu(x)$.

Thus, given the α-cuts

$$\left[x(\alpha)_{c_1\ell}, x(\alpha)_{c_1 r}\right] = \left\{x : \mu_{c_1}(x) \geq \alpha\right\} \quad (31)$$

$$\left[x(\alpha)_{c_2\ell}, x(\alpha)_{c_2 r}\right] = \left\{x : \mu_{c_2}(x) \geq \alpha\right\} \quad (32)$$

$$\left[x(\alpha)_{p_1\ell}, x(\alpha)_{p_1 r}\right] = \left\{x : \mu_{p_1}(x) \geq \alpha\right\} \quad (33)$$

$$\left[x(\alpha)_{p_2\ell}, x(\alpha)_{p_2 r}\right] = \left\{x : \mu_{p_2}(x) \geq \alpha\right\} \quad (34)$$

we can calculate the corresponding α-cuts $\left[x(\alpha)_{c_3\ell}, x(\alpha)_{c_3 r}\right]$ and $\left[x(\alpha)_{p_3\ell}, x(\alpha)_{p_3 r}\right]$ of the fuzzy membership functions $\mu_{c_3}(x)$ and $\mu_{p_3}(x)$ for $0\leq\alpha\leq\min_i \alpha_{\max i}$ using (9)-(10) and (22)-(23), with the input α-cut intervals specified by (31)-(34). Then the Type-1 memberships for $\mu_{c_3}(x)$ and $\mu_{p_3}(x)$ are given by:

$$\mu_{c_3}(x) = \sup_{\forall\alpha\in[0,\alpha_{\max}]} \begin{cases} \alpha, & x(\alpha)_{c_3\ell} \leq x \leq x(\alpha)_{c_3 r} \\ 0 & elsewhere \end{cases}, \quad (35)$$

$$\mu_{p_3}(x) = \sup_{\forall\alpha\in[0,\alpha_{\max}]} \begin{cases} \alpha, & x(\alpha)_{p_3\ell} \leq x \leq x(\alpha)_{p_3 r} \\ 0 & elsewhere \end{cases}, \quad (36)$$

where $\alpha_{\max} = \min_i \alpha_{\max i}$. In a similar vein, we can compute the fuzzy membership function $\mu_t(x)$ of trust from the α-cut intervals for $c\in[c_\ell,c_r]$ and $p\in[p_\ell,p_r]$ using (26)-(27) to compute the corresponding α-cut intervals $\left[x(\alpha)_{t\ell}, x(\alpha)_{tr}\right]$ for t:

$$\mu_t(x) = \sup_{\forall\alpha\in[0,\alpha_{\max}]} \begin{cases} \alpha, & x(\alpha)_{t\ell} \leq x \leq x(\alpha)_{tr} \\ 0 & elsewhere \end{cases}. \quad (37)$$

These membership functions can be de-fuzzified to a representative scalar value by computing their centroids, where the centroid ς_μ of a fuzzy membership function $\mu(x)$ is defined by

ISBN: 978-989-95079-6-8

$$\varsigma_\mu = \frac{\int x\mu(x)dx}{\int \mu(x)dx} \tag{38}$$

In practice, the integrals in (38) are approximated by summations.

5 Interval Type-2 fuzzy inputs

Now suppose that the fuzzy membership functions for c_1, c_2, p_1 and p_2 in equations (1) and (2) are interval Type-2 [13] fuzzy membership functions $\tilde{\mu}_{c_1}(x)$, $\tilde{\mu}_{c_2}(x)$, $\tilde{\mu}_{p_1}(x)$ and $\tilde{\mu}_{p_2}(x)$, respectively. Each of these functions exhibits a "footprint of uncertainty" (FOU) that represents the union of an uncountable number of Type-1 primary fuzzy membership functions lying within the FOU, each having a secondary membership equal to unity [14]. Thus we wish to propagate this imprecision in the input variables through to the corresponding imprecision in the output aggregation of two inputs.

In this case, we use an approach analogous to that of Wu and Mendel [15,16] for linguistic weighted averages, which relies upon the observation that an interval Type-2 fuzzy set is completely specified by the upper and lower convex membership functions (UMF and LMF) that bound its FOU. We denote the Type-2 membership functions for the input confidences and plausibilities by $\mu_{c_1}^{UMF}(x)$, $\mu_{c_1}^{LMF}(x)$, $\mu_{c_2}^{UMF}(x)$, $\mu_{c_2}^{LMF}(x)$, $\mu_{p_1}^{UMF}(x)$, $\mu_{p_1}^{LMF}(x)$, $\mu_{p_2}^{UMF}(x)$ and $\mu_{p_2}^{LMF}(x)$, respectively. To determine the corresponding interval Type-2 membership functions $\tilde{\mu}_{c_3}(x)$ and $\tilde{\mu}_{p_3}(x)$ for the aggregate confidence and plausibility, we employ the α-cut technique of the previous section applied separately to the UMF's and LMF's to compute the output UMF's and LMF's for confidence and plausibility, respectively. The area between these bounding functions represents the primary Type-2 fuzzy membership functions for these variables, with a secondary membership equal to unity over this FOU. In similar fashion, we can find the UMF and LMF for the trust Type-2 membership function $\tilde{\mu}_t(x)$.

Once any of the Type-2 memberships have been calculated by the above method, we have a very rich description of the imprecision of the corresponding fuzzy membership function. These Type-2 membership functions can be *type-reduced* to a Type-1 membership function by calculating the *centroid* of the Type-2 membership function via the Karnik-Mendel algorithm [12]. For interval Type-2 membership functions, the centroid membership function is itself an interval $[z_\ell, z_r]$. This is of course a reduced representation of the imprecision that is fully described by the Type-2 membership function. The interval Type-1 membership function can then be *defuzzified* by calculating the midpoint of this interval to obtain a scalar representation of the aggregate membership function.

6 Example

We now consider an example that approximates the first case depicted in equation (3), but incorporates more general interval Type-2 fuzzy membership functions to describe the distributions of confidence and plausibility. In all cases below, we use $\alpha = 2$ and $\beta = 1/4$.

Fig. 1 shows two Type-2 fuzzy membership functions for the confidence of two different observations, where the FOU of each membership is the area between the solid and dotted curves. In this example, both confidences have positive support intervals, centered upon the values 0.8 (with + shading) and 0.3 (with x shading) respectively, analogous to the point values of confidence in equation (3).

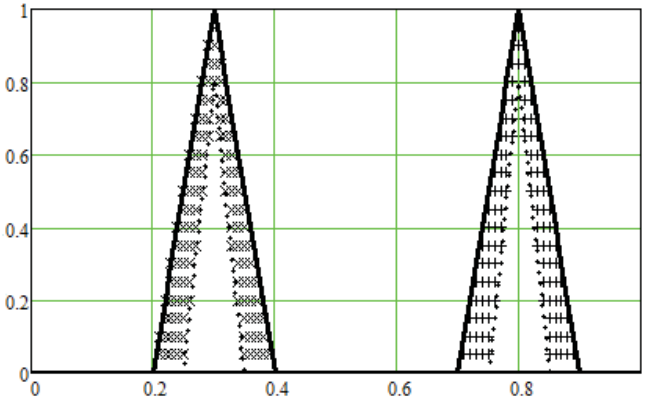


Figure 1 Interval Type-2 fuzzy membership functions for the confidence of two different observations.

In Fig. 2, we show the corresponding interval Type-2 membership functions of the plausibility of these two observations. Note the plausibility of the first observation (+ shading in Fig. 1) is concentrated near zero, while that of the second observation is centered on 0.5, but both cases exhibit a FOU of values.

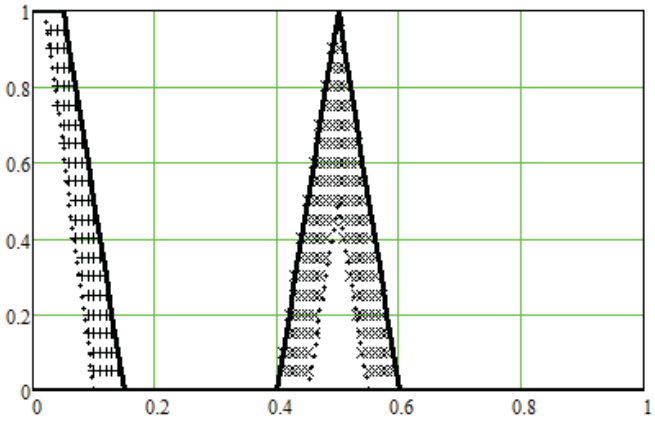


Figure 2 Interval Type-2 fuzzy membership functions for the plausibility of the observations whose confidence distributions are depicted in Fig. 1.

The mathematical apparatus described in previous sections for propagating the imprecision of these values when aggregating these two observations' confidences and plausibilities results in the aggregate confidence (+ shading) and plausibility (x shading) interval Type-2 membership functions shown in Fig. 3. The vertical dashed lines straddling the confidence FOU depict the centroid interval that results from type-reduction of the confidence membership function, while the vertical solid lines depict the midpoint (i.e., the de-fuzzified value) of the centroid intervals for both confidence and plausibility. These have

ISBN: 978-989-95079-6-8

values of 0.315 and 0.51, respectively, which roughly compares with the point wise values in equation (3), confirming that the aggregation of an observation with high confidence but very low plausibility results in essentially the same confidence and plausibility associated with the other observation. Note, however, that there is considerable additional information regarding both distributions that is lost in the type-reduction of a Type-2 membership function to a Type-1 interval membership function, and still more information is lost by representing the entire FOU by its de-fuzzified midpoint value.

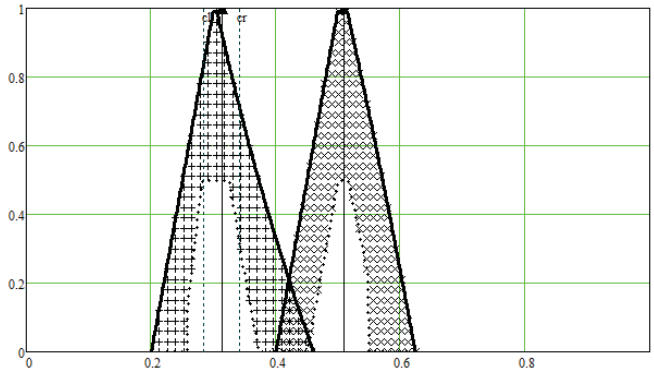


Figure 3 Interval Type-2 fuzzy membership functions for the aggregate confidence (+ shading) and plausibility (x shading) of the case depicted in Figures 4 and 5.

7 Conclusion

In this paper, we have generalized the trust aggregation algebra of Hamilton [8] for point values to admit interval, Type-1 and/or Type-2 input fuzzy membership values, and shown how these more complex inputs can be propagated in a mathematically principled manner to the corresponding output fuzzy membership functions.

References

[1] Grandison, T. and M. Sloman, "A survey of trust in internet applications," *IEEE Communications Surveys and Tutorials,* 3 pp. 2–16, 2000.

[2] Ray, I., S. Chakraborty, and I. Ray, "VTrust: a trust management system based on a vector model of trust," *Proceedings of International Conference on Information Systems Security*, Dec. 2005.

[3] Blaze, M., J. Feigenbaum and J. Lacy, "Decentralized trust management," *Proceedings of 17th IEEE Symposium on Security and Privacy*, Oakland, California, USA, pp. 164-173, IEEE Computer Society Press, 1996.

[4] Blaze, M., J. Feigenbaum and M. Strauss, "Compliance checking in the policymaker trust management system," *Proceedings of the 2nd Financial Crypto Conference*, vol. 1465 of *Lecture Notes in Computer Science*, pp. 254-274, Anguilla, Springer-Verlag, 1998.

[5] Jajodia, S., P. Samarati and V.S. Subramanian, "Logical language for expressing authorizations," *Proceedings IEEE Symposium on Security and Privacy*, pp. 31-42, May 1997.

[6] Damianou, N., N. Dulay, E. Lupu and M. Sloman. "The ponder policy specification language," *Policies for Distributed Systems and Networks*, vols. 1995, 2001.

[7] H. Hosmer, "Using fuzzy logic to represent security policies in the multipolicy paradigm," *ACM SIGSAC Review*, vol. 10, no. 4, pp. 12-21, Fall 1992.

[8] Hamilton, S. N. and Hamilton W. L. Hamilton, "Applying trust models to access control," submitted to *ACM Symposium on Access Control Models and Technologies*, June 2009.

[9] Wu, D. and J. M. Mendel, "Aggregation using the linguistic weighted average and interval Type-2 fuzzy sets," *IEEE Trans. Fuzzy Systems*, vol. 15, no. 6, pp. 1145-1161, Dec. 2007.

[10] Wu, D. and J. M. Mendel, "Corrections to "Aggregation using the linguistic weighted average and interval Type-2 fuzzy sets"," *IEEE Trans. Fuzzy Systems,* to appear.

[11] Klir, G. J. and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[12] Karnik, N. N. and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences*, vol. 132, pp. 195-220, 2001.

[13] Mendel, J. M., *Uncertain Rule-Based Fuzzy Logic Systems*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[14] Mendel, J. M., "Computing with words and its relationship with fuzzistics," *Information Sciences*, vol. 177, pp. 988-1006, 2007.

[15] Mendel, J. M. and H. Wu, "Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 1: forward problem," *IEEE Trans. Fuzzy Systems*, vol. 14, pp. 781-792, 2006.

[16] Mendel, J. M. and H. Wu, "Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 2: inverse problem," *IEEE Trans. Fuzzy Systems*, vol. 15, pp. 301-308, 2007.

[17] Liu, D. and J. M. Mendel, "Encoding words into interval Type-2 fuzzy sets using an *interval* approach," *IEEE Trans. Fuzzy Systems*, vol. 16, pp. 1503-1521, Dec. 2008.

ISBN: 978-989-95079-6-8

On vertices of the k-additive monotone core

Pedro Miranda[1] Michel Grabisch[2]

1. Faculty of Mathematical Sciencies, Universidad Complutense de Madrid
Madrid, Spain
2. Centre d'Économie de la Sorbonne, Université Paris I-Panthéon-Sorbonne
Paris, France
Email: pmiranda@mat.ucm.es, Michel.Grabisch@univ-paris1.fr

Abstract— Given a capacity, the set of dominating k-additive capacities is a convex polytope; thus, it is defined by its vertices. In this paper we deal with the problem of deriving a procedure to obtain such vertices in the line of the results of Shapley and Ichiishi for the additive case. We propose an algorithm to determine the vertices of the k-additive monotone core. Then, we characterize the vertices of the n-additive core and finally, we explore the possible translations for the k-additive case.

Keywords— Capacities, k-additivity, dominance, core.

1 Introduction

One of the main problems of cooperative game theory is to define a solution of a game μ, that is, supposing that all players join the grand coalition X, an *imputation* to each player represents a sharing of the total worth of the game $\mu(X)$. In the case of finite games of n players, an imputation can be written as a n-tuple $(x_1, \ldots, x_n)$ such that $\sum_{i=1}^{n} x_i = \mu(X)$. Of course, some rationality criterion should prevail when defining the sharing.

In this respect, the core is perhaps the most popular solution of a game. It is a well known fact that the core is nonempty if and only if the game is balanced [1]. However, there are games whose core is empty. It is then necessary to give an alternative solution. In this sense, many possibilities have been proposed in the literature, as the dominance core stable sets, Shapley index, the nucleolus, etc. (see e.g. [2]).

On the other hand, Grabisch has defined in [3] the concept of k-additive capacities, for a fixed value $1 \leq k \leq n$. These capacities generalize the concept of probability and they fill the gap between probabilities and general capacities. Moreover, as they are defined in terms of the Möbius transform and this transform can be applied to the characteristic function of any game (not necessarily monotone), the concept of k-additivity can be extended to games as well.

In a previous paper we have defined the so-called k-additive core. The basic idea is to remark that an imputation is nothing other than an additive game, and if the core is empty, we may allow to search for games more general than additive ones, namely k-additive games, dominating the game. We have presented a generalization of balanced games, the k-balanced games, that are those admitting a dominating k-additive game and no dominating $(k-1)$-additive game.

We have seen that for general games, any game is either balanced or 2-balanced. Moreover, the 2-additive core is not a bounded polytope but an unbounded convex polyhedron. However, when dealing with capacities, it makes sense to study the k-additive monotone core and it can be easily seen that in this particular case it is a convex polytope, whence it can be defined through describing its vertices. This paper studies these vertices. In the framework of Game Theory, it has been solved for the (1-additive) core by Shapley and Ichiishi. The vertices of the $(n-1)$-additive core has been obtained in [4].

Moreover, there are other fields in which it is interesting to find the set of probabilities dominating a capacity. For instance, Dempster [5] and Shafer [6] have proposed a representation of uncertainty based on a "lower probability" or "degree of belief", respectively, to every event. Their model needs a lower probability function, usually non-additive but having a weaker property: it is a belief function [6]. This requirement is perfectly justified in some situations (see [5]). The general form of lower probabilities has been studied by several authors (see e.g. [7, 8]). Moreover, in many decision problems, in which we have not enough information, decision makers often feel that they are only able to assign an interval value for the probability of events. In other words, they do not know the real probability distribution but there exists a set of probabilities compatible with the available information. Let us call this set of all compatible probabilities $\mathcal{P}_1$ and let us define $\mu = \inf_{P \in \mathcal{P}_1} P$; then, μ is a capacity (but not necessarily a belief function [9]); μ is called "coherent lower probability", and it is the natural "lower probability function". Of course, if P' is a probability measure dominating μ, it is clear that $E_{P'}(f) \geq \mathcal{C}_\mu(f)$, for any function f, where $\mathcal{C}_\mu$ represents Choquet integral [10]. Chateauneuf and Jaffray use this fact and that $\mu \leq P$, $\forall P \in \mathcal{P}_1$ in [11] to obtain an easy method for computing a lower bound of $\inf_{P \in \mathcal{P}_1} E_P(f)$ whenever μ is 2-monotone. Their method is based on obtaining the set of all probability distributions dominating μ. The same can be done for obtaining an upper bound. In this case, we can find a similar motivation for studying the set of all k-additive capacities dominating a capacity.

The paper is organized as follows: In next section, we give the basic concepts about k-additive capacities and about the set of dominating probabilities. Next, in Section 3 we provide an algorithm for obtaining the set of all k-additive dominating capacities. Section 4 is devoted to characterize the vertices for the n-additive case and, in Section 5, we deal with possible generalizations for the k-additive case.

2 Basic concepts

We will use the following notations: we suppose a finite universal set with n elements, $X = \{1, ..., n\}$. Subsets of X are

ISBN: 978-989-95079-6-8

denoted by capital letters A, B, and so on. The set of subsets of X is denoted by $\mathcal{P}(X)$, and the set of subsets whose cardinality is maximum k is denoted by $\mathcal{P}^k(X)$.

Definition 1 *[12] A* **game** *over X is a mapping $\mu : \mathcal{P}(X) \to \mathbb{R}$ (called* **characteristic function***) satisfying $\mu(\emptyset) = 0$.*
If, in addition,

1. *μ satisfies $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$, the game μ is said to be* **monotone***;*
2. *μ satisfies $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \subseteq X$, $A \cap B = \emptyset$, the game is said to be* **additive***;*
3. *μ satisfies $\mu(A \cup B) + \mu(A \cap B) \geq \mu(A) + \mu(B)$, for all $A, B \subseteq X$, the game is said to be* **convex***.*
4. *μ satisfies*

$$\mu(\bigcup_{i=1}^{k} A_i) \geq \sum_{\substack{K \subseteq \{1,\ldots,k\} \\ K \neq \emptyset}} (-1)^{|K|+1} \mu(\bigcap_{j \in K} A_j) \quad (1)$$

*for any family of k subsets $A_1, \ldots A_k$, $k \geq 2$, the game is said to be k-***monotone***.*

Definition 2 *A* **non-additive measure** *[13] or* **capacity** *[10] or* **fuzzy measure** *[14] μ over X is a monotone game with $\mu(X) = 1$.*

Note that any monotone game can be equivalently defined through a capacity. The set of all capacities on X is a convex polytope, that we will denote $\mathcal{FM}(X)$.

Definition 3 *[15] Let μ be a game on X. The* **Möbius transform (or inverse)** *of μ is a set function on X defined by*

$$m_\mu(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B), \, \forall A \subseteq X. \quad (2)$$

The Möbius transform given, the original characteristic function can be recovered through the *Zeta transform* [11]:

$$\mu(A) = \sum_{B \subseteq A} m(B). \quad (3)$$

Let us turn to the concept of k-additivity. In order to define a capacity, $2^n - 2$ values are necessary. The number of coefficients grows exponentially with n, and so does the complexity of the problem. This drawback reduces considerably the practical use of capacities. Then, some subfamilies of capacities have been defined in an attempt to reduce complexity. In this paper we will use k-additive capacities.

Definition 4 *[16] A game μ is said to be k-***order additive** *or k-***additive** *for some $k \in \{1, \ldots, n\}$ if its Möbius transform vanishes for any $A \subseteq X$ such that $|A| > k$, and there exists at least one subset A of exactly k elements such that $m(A) \neq 0$.*

In this sense, a probability is just a 1-additive capacity [16]. Thus, k-additive capacities generalize probabilities, that are very restrictive in many situations as they do not allow interactions between the elements of X. They fill the gap between probabilities and general non-additive capacities. We will denote by $\mathcal{FM}^k(X)$ (resp. $\mathcal{G}^k(X)$) the set of all k'-additive capacities (resp. games) with $k' \leq k$.

Let us introduce the concept of k-additive monotone core.

Definition 5 *Let $\mu\mu^*$ be two games. We say μ^** **dominates** *μ, and we denote it $\mu^* \geq \mu$, if and only if*

$$\mu^*(A) \geq \mu(A), \, \forall A \subset X, \mu^*(X) = \mu(X). \quad (4)$$

One of the main problems of cooperative game theory is to define a solution of a game ν, that is, supposing that all players join the grand coalition X, an *imputation* to each player represents a sharing of the total worth of the game $\nu(X)$. In the case of finite games of n players, an imputation can be written as a n-tuple $(x_1, \ldots, x_n)$ such that $\sum_{i=1}^n x_i = \nu(X)$. Of course, some rationality criterion should prevail when defining the sharing.

Definition 6 *Let μ be a game. We say that a vector : $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is an* **imputation** *for μ if it satisfies*

$$\sum_{i=1}^{n} x_i = \mu(X). \quad (5)$$

An imputation represents a possible pay-off for the players, i.e. supposing that all players agree to form the grand coalition, it provides a possible sharing of the value $\mu(X)$ among the players.

Remark 1 *For any $\mathbf{x} \in \mathbb{R}^n$, it is convenient to use the notation $x(A) := \sum_{i \in A} x_i$, for all $A \subseteq X$, with the convention $x(\emptyset) = 0$. Thus, x identifies with an additive game for which the values on singletons are x_i.*

The value x_i is the asset player i receives when sharing $\mu(X)$. Suppose that the imputation satisfies $x(A) \geq \mu(A)$, for all $A \subseteq X$. If this is the case, all players should agree with their pay-off, as if they try to form other coalition different of X, the corresponding value for the coalition would be worse than the value the coalition obtains with the additive game x. In other words, any such $(x_1, \ldots, x_n)$ is a possible satisfactory imputation for all players.

Definition 7 *[17] Let μ be a game. The* **core** *of μ, denoted by $\mathcal{C}(\mu)$, is defined by*

$$\mathcal{C}(\mu) := \{x \in \mathbb{R}^n \mid x(A) \geq \mu(A), \, \forall A \subseteq X, x(X) = \mu(X)\}.$$

Since by Remark 1 any $x \in \mathbb{R}^n$ induces an additive game, the core can be equivalently defined as the set of additive games dominating μ. When the core is nonempty, it is usually taken as the solution of the game. Note that for the case of the core, given a dominating additive game, the value x_i coincides with $m(i)$. However, there are games with an empty core. Then, the following definition arises:

Definition 8 *[12] A game μ is* **balanced** *if $\mathcal{C}(\mu) \neq \emptyset$.*

For the special case of μ being a capacity, if $(x_1, \ldots, x_n)$ is in the core, it follows that $(x_1, \ldots, x_n)$ determines a probability distribution on X dominating μ. Thus, in this case, $\mathcal{C}(\mu)$ coincides with the set of all probabilities dominating μ.

When non-empty, the core is a convex polytope and its vertices are known when the game is convex.

Definition 9 *A* **maximal chain** *in 2^X is a sequence of subsets $A_0 := \emptyset, A_1, \ldots, A_{n-1}, A_n := X$ such that $A_i \subset A_{i+1}$, $i = 0, \ldots, n-1$. The set of maximal chains of 2^X is denoted by $\mathcal{M}(2^X)$.*

ISBN: 978-989-95079-6-8

To each maximal chain $C := \{\emptyset, A_1, \ldots, A_n = X\}$ in $\mathcal{M}(2^X)$ corresponds a unique permutation σ on X such that $A_1 = \sigma(1)$, $A_2 \setminus A_1 = \sigma(2)$, $\ldots$, $A_n \setminus A_{n-1} = \sigma(n)$. The set of all permutations over X is denoted by $\mathfrak{S}(X)$. Let μ be a capacity. Each permutation σ (or maximal chain C) induces an additive capacity ϕ^σ (or ϕ^C) on X defined by:

$$\phi^\sigma(\{\sigma(i)\}) := \mu(\{\sigma(1), \ldots, \sigma(i)\}) - \mu(\{\sigma(1), \ldots, \sigma(i-1)\}) \quad (6)$$

or

$$\phi^C(\{\sigma(i)\}) := \mu(A_i) - \mu(A_{i-1}), \quad \forall i \in X, \quad (7)$$

with the above notation.

Theorem 1 *The following propositions are equivalent.*

1. *μ is a convex capacity.*
2. *All additive capacities ϕ^σ, $\sigma \in \mathfrak{S}(X)$, belong to the core of μ.*
3. *$\mathcal{C}(\mu) = \mathrm{co}(\{\phi^\sigma\}_{\sigma \in \mathfrak{S}(X)})$.*
4. *$\mathrm{ext}(\mathcal{C}(\mu)) = \{\phi^\sigma\}_{\sigma \in \mathfrak{S}(X)}$,*

where $\mathrm{co}(\cdot)$ and $\mathrm{ext}(\cdot)$ denote respectively the convex hull of some set, and the extreme points of some convex set.

(i) $\Rightarrow$ (ii) and (i) $\Rightarrow$ (iv) are due to Shapley [17], while (ii) $\Rightarrow$ (i) was proved by Ichiishi [18].

In a previous work [19] we have defined the so-call k-additive monotone core.

Definition 10 *For some integer $1 \leq k \leq n$, the k-**additive monotone core** of a capacity μ is defined by:*

$$\mathcal{MC}^k(\mu) := \{\phi \in \mathcal{FM}^k(X) \mid \phi(A) \geq \mu(A), \quad \forall A \subseteq X\}.$$

If non-empty, it is easy to see that $\mathcal{MC}^k(\mu)$ is a convex polytope. In next sections we will study its vertices, i.e. the capacities such that they cannot be put as a convex combination of two other capacities in the polytope.

3 An algorithm for determining vertices of the k-additive monotone core

Take $\mu \in \mathcal{FM}(X)$. The polytope $\mathcal{MC}^k(\mu)$ can be seen as a subpolytope of $\mathcal{FM}^k(X)$, given by the additional constraints

$$\mu^*(A) \geq \mu(A), \forall A \subseteq X. \quad (8)$$

I.e. we restrict the polytope to the measures $\mu^* \in \mathcal{FM}^k(X)$ dominating μ. Thus, we propose the following algorithm to determine its vertices:

- Initialization: $\mathcal{FM}^k(X)$.
- Take $A \subseteq X$ and add the constraint $\mu^*(A) \geq \mu(A)$.
- Obtain the vertices and the adjacency structure (i.e. whether two vertices are in an edge) of the new polytope.
- Repeat for any $A \subseteq X, A \neq X, \emptyset$.

Let us analyze a step. We will denote by $\mathcal{F}_1$ the polytope before introducing the new constraint $\mu^*(A) \geq \mu(A)$ and by $\mathcal{F}_2$ the resulting polytope. For $\mathcal{F}_1$, we assume that we know its vertices and its adjacency structure. The vertices of $\mathcal{F}_2$ are:

- Vertices of $\mathcal{F}_1$ satisfying the new constraint.
- New vertices, coming from the intersection of the hyperplane defined by $\mu^*(A) = \mu(A)$ and $\mathcal{F}_1$. It can be easily proved that these new vertices are in edges of $\mathcal{F}_1$.

Thus, in order to determine $\mathcal{F}_2$, it suffices to know the vertices and the edges (whether two vertices are adjacent) of $\mathcal{F}_1$.

In order to apply this procedure again, let us determine the adjacency structure of $\mathcal{F}_2$. Consider μ_1, μ_2 two vertices of $\mathcal{F}_2$. We have several cases:

- μ_1, μ_2 are vertices of $\mathcal{F}_1$.
- μ_1 is a vertex of $\mathcal{F}_1$ but μ_2 is not.
- Neither μ_1 nor μ_2 are vertices of $\mathcal{F}_1$.

We will study whether μ_1 and μ_2 are adjacent vertices in each situation.

Consider μ_1, μ_2 two vertices of $\mathcal{F}_2$ that are also vertices of $\mathcal{F}_1$. If they are adjacent vertices in $\mathcal{F}_1$, then they are adjacent in $\mathcal{F}_2$, as $\mathcal{F}_2$ is a subpolytope of $\mathcal{F}_1$ and we are done.

Lemma 1 *Assume μ_1, μ_2 are not adjacent vertices in $\mathcal{F}_1$. Then, if μ_1, μ_2 are adjacent vertices in $\mathcal{F}_2$, they satisfy $\mu_1(A) = \mu(A) = \mu_2(A)$.*

Moreover, the following holds:

Lemma 2 *Consider μ_1, μ_2 two vertices of $\mathcal{F}_1$ and assume they are adjacent vertices in $\mathcal{F}_2$. Then, if μ_1, μ_2 are not adjacent in $\mathcal{F}_1$, necessarily $\mu_1(A) = \mu(A) = \mu_2(A)$.*

Remark 2 *It can be proven that if μ_1, μ_2 are in the conditions of the previous lemma, then they are in a facet of dimension 2 of $\mathcal{F}_1$.*

Moreover, the intersection of this facet with the hyperplane $\mu^*(A) = \mu(A)$ is exactly the segment $[\mu_1, \mu_2]$. Otherwise, if we can find μ_3 outside the segment and in the facet satisfying $\mu_3(A) = \mu(A)$, we can build two linearly independent vectors in the facet and in the hyperplane, whence the facet is contained in the hyperplane and thus, the facet is contained in $\mathcal{F}_2$. But this would imply that, as μ_1, μ_2 are not adjacent in $\mathcal{F}_1$, they are not adjacent in $\mathcal{F}_2$, a contradiction.

Let us now turn to the case in which μ_1 is a vertex of $\mathcal{F}_1$ but μ_2 is not. The conditions for μ_1, μ_2 being adjacent vertices in $\mathcal{F}_2$ are given in next lemma.

Lemma 3 *Consider μ_1, μ_2 two vertices of $\mathcal{F}_2$ and suppose μ_1 is a vertex of $\mathcal{F}_1$ and μ_2 is not. If they are adjacent in $\mathcal{F}_2$, then:*

- *If $\mu_1(A) > \mu(A)$, then μ_2 is in an edge starting in μ_1.*
- *If $\mu_1(A) = \mu(A)$, then μ_1 and μ_2 are in a facet of $\mathcal{F}_1$ of dimension 2. Indeed, the intersection of the facet with the hyperplane $\mu^*(A) = \mu(A)$ is the segment $[\mu_1, \mu_2]$.*

Moreover, the following can be proved:

Lemma 4 *Consider μ_2 a vertex of $\mathcal{F}_2$ such that μ_2 is not a vertex of $\mathcal{F}_1$. Then, there exists exactly one vertex μ_1 of $\mathcal{F}_1$ such that $\mu_1(A) > \mu(A)$ that is adjacent to μ_2 in $\mathcal{F}_2$.*

ISBN: 978-989-95079-6-8

Finally, let us consider the case of μ_1, μ_2 being two vertices of $\mathcal{F}_2$ that are not vertices of $\mathcal{F}_1$.

Lemma 5 *Suppose μ_1, μ_2 are vertices of $\mathcal{F}_2$ that are not vertices of $\mathcal{F}_1$. If they are adjacent vertices of $\mathcal{F}_2$, they are in the same facet of dimension 2 of $\mathcal{F}_1$. Indeed, the intersection of the facet with the hyperplane $\mu^*(A) = \mu(A)$ is the segment $[\mu_1, \mu_2]$.*

As a final remark about this section note that, in order to apply this procedure, it is necessary to know the adjacency structure of the polytope $\mathcal{FM}^k(X)$. As proved in [20, 21], the problem of determining non-adjacency of vertices of a polytope is, in some cases, NP-complete. The vertices of $\mathcal{FM}(X)$ are $\{0,1\}$-valued measures [22]. In [23] a characterization of the adjacency in $\mathcal{FM}(X) = \mathcal{FM}^n(X)$ that allows us to check whether two vertices are adjacent in quadratic time has been obtained. The adjacency structure of $\mathcal{FM}^1(X)$ and $\mathcal{FM}^2(X)$ is also known. However, the structure of $\mathcal{FM}^k(X)$ for other values of k is more complicated [24] and the adjacency structure is not known (indeed, the vertices of the polytope have not been obtained yet). In this last case, we are forced to apply a similar algorithm with the additional constraints $m(A) = 0$ before applying the procedure.

4 The set $\mathcal{MC}^n(\mu)$.

The procedure stated in the previous section can be very time-consuming and thus unfeasible in practice for big values of $|X|$. Thus, it is interesting to look for a characterization of the vertices of the k-additive core; in this line we have the results of Shapley and Ichiishi [17, 18] for $\mathcal{C}^1(\mu)$ and the results in [4] for $\mathcal{C}^{n-1}(\mu)$ (the $(n-1)$-additive core, not restricted to capacities). In this section we provide a characterization of vertices of $\mathcal{MC}^n(\mu)$. We consider the following procedure.

- Let $\prec$ be an order on $\mathcal{P}(X) \setminus \{X, \emptyset\}$. This order allows us to rank the different subsets of X,

$$A_1 \prec A_2 \prec ... \prec A_{2^n-2}. \tag{9}$$

- Next, take a partition $\mathcal{P} = \{\mathcal{U}, \mathcal{L}\}$ on $\mathcal{P}(X) \setminus \{X, \emptyset\}$, where $\mathcal{U}$ or $\mathcal{L}$ could be empty.

- **Initializing step**: For $\prec$ and $\mathcal{P}$ fixed, let us define

$$\overline{\mu}^0(A_i) = 1, \underline{\mu}^0(A_i) = \mu(A_i), \forall A_i. \tag{10}$$

- **Iterating step**: For $i = 1$ until $i = 2^n - 2$, do:

  – If $A_i \in \mathcal{U}$, then assign

$$\mu_{\prec,\mathcal{P}}(A_i) = \overline{\mu}^{i-1}(A_i). \tag{11}$$

  Redefine:
  For $\underline{\mu}^i$, we put

$$\underline{\mu}^i(B) = \max\{\overline{\mu}^{i-1}(A_i), \underline{\mu}^{i-1}(B)\}, \text{ if } A_i \subseteq B \tag{12}$$

$$\underline{\mu}^i(B) = \underline{\mu}^{i-1}(B), \text{ otherwise.} \tag{13}$$

  For $\overline{\mu}^i$, we put

$$\overline{\mu}^i(B) = \overline{\mu}^{i-1}(B), \forall B \subset X. \tag{14}$$

  – If $A_i \in \mathcal{L}$, then assign

$$\mu_{\prec,\mathcal{P}}(A_i) = \underline{\mu}^{i-1}(A_i). \tag{15}$$

  Redefine:
  For $\overline{\mu}^i$, we put

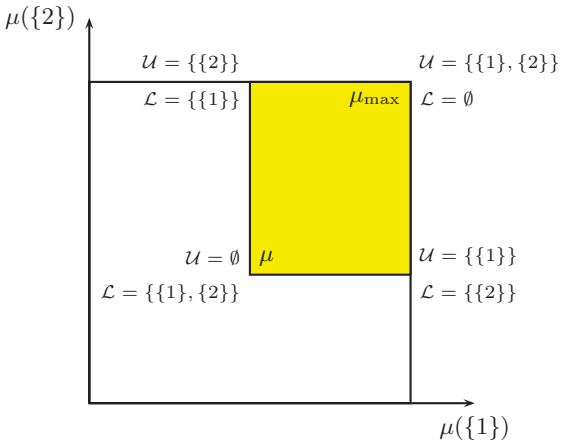
$$\overline{\mu}^i(B) = \min\{\underline{\mu}^{i-1}(A_i), \overline{\mu}^{i-1}(B)\}, \text{ if } B \subseteq A_i \tag{16}$$

$$\overline{\mu}^i(B) = \overline{\mu}^{i-1}(B), \text{ otherwise.} \tag{17}$$

  For $\underline{\mu}^i$, we put

$$\underline{\mu}^i(B) = \underline{\mu}^{i-1}(B), \forall B \subset X. \tag{18}$$

The idea of the procedure is the following: If we are in step i, the values for $A_1, ..., A_{i-1}$ are fixed. For A_i, if $A_i \in \mathcal{U}$, we assign to $\mu_{\prec,\mathcal{P}}(A_i)$ the biggest possible value keeping dominance, which is $\overline{\mu}^i(A_i)$. Similarly, if $A_i \in \mathcal{L}$, we assign to $\mu_{\prec,\mathcal{P}}(A_i)$ the smallest possible value keeping dominance, which is $\underline{\mu}^i(A_i)$. Once the value of $\mu_{\prec,\mathcal{P}}(A_i)$ is fixed, we need to actualize the lower and upper bounds for $A_j, j > i$. These lower and upper bounds are stored in $\underline{\mu}^{i+1}$ and $\overline{\mu}^{i+1}$, respectively. The Figure below explains the performance of the algorithm for the special case of $|X| = 2$. In it, we can see that $\mathcal{MC}^2(\mu)$ has four vertices. Note that μ itself is always a vertex of $\mathcal{MC}^2(\mu)$, corresponding to the case $\mathcal{L} = \mathcal{P}(X) \setminus \{X, \emptyset\}$ and $\mathcal{U} = \emptyset$. Thus, for the n-additive case, $\mu \in \mathcal{MC}^n(\mu)$ and it is the bottom element of this set. Similarly, the measure attaining value 1 for every subset is another vertex of $\mathcal{MC}^2(\mu)$, corresponding to the case $\mathcal{U} = \mathcal{P}(X) \setminus \{X, \emptyset\}$ and $\mathcal{L} = \emptyset$. Thus, for the n-additive case, it is the top element of the set. This also holds when $|X| > 2$.



Note also that this procedure generalizes for the n-additive case the Shapley-Ichiishi theorem for probabilities. In our case, the total order on $\mathcal{P}(X)\setminus\{\emptyset, X\}$ plays the role of chains; the value that is assigned to a singleton in the Shapley-Ichiishi theorem is indeed the upper and lower bound for this value in order to keep dominance. However, in our case we need two "chains" instead of one because lower and upper bounds are not the same for the general case.

Finally, remark that for the general case we do not need to impose any additional condition on μ, while for the set $\mathcal{C}(\mu) = \mathcal{MC}^1(\mu)$ convexity is required.

In next results we will prove that the function $\mu_{\prec,\mathcal{P}}$ obtained through this procedure is a vertex of $\mathcal{MC}^n(\mu)$ and that any vertex can be obtained through a suitable choice of $\prec$ and $\mathcal{P}$.

ISBN: 978-989-95079-6-8

Proposition 1 $\mu_{\prec,\mathcal{P}} \in \mathcal{MC}^n(\mu)$.

The proof of this result is based on the following lemmas, that provide us with additional properties of $\{\overline{\mu}^i(B)\}_{i=0}^{2^n-2}$ and $\{\underline{\mu}^i(B)\}_{i=0}^{2^n-2}$.

Lemma 6 *For any $B \subset X$, the sequence $\{\underline{\mu}^i(B)\}_{i=1}^{2^n-2}$ is nondecreasing. Similarly, the sequence $\{\overline{\mu}^i(B)\}_{i=0}^{2^n-2}$ is nonincreasing.*

Lemma 7 $\underline{\mu}^i, \overline{\mu}^i \in \mathcal{FM}(X), \forall i = 0, ..., 2^n - 2$.

Lemma 8 $\overline{\mu}^i \geq \underline{\mu}^i, \forall i = 0, ..., 2^n - 2$.

Moreover, the following holds.

Proposition 2 *$\mu_{\prec,\mathcal{P}}$ is a vertex of $\mathcal{MC}^n(\mu)$.*

Finally, it can be seen that all the vertices can be derived from this procedure.

Proposition 3 *If μ^* is a vertex of $\mathcal{MC}^n(\mu)$, there exists an order $\prec$ and a partition $\mathcal{P}$ of $\mathcal{P}(X) \setminus \{X, \emptyset\}$ such that $\mu^* = \mu_{\prec,\mathcal{P}}$.*

Note that these results allows us to derive an upper bound for the number of vertices of $\mathcal{MC}^n(\mu)$.

Proposition 4 *The number of vertices of $\mathcal{MC}^n(\mu)$ is bounded by 2^{2^n-2}.*

5 The case of $\mathcal{MC}^k(\mu)$

In this section we treat the general k-additive case. The basic idea is to translate the results of the previous section to this case. In order to translate the results, we will need to solve two new problems:

- For a fixed value of $\mu(A)$, the possible lower and upper bounds of $\mu(B), B \neq A$ are not trivial, as it happened for the n-additive case. Moreover, if we are dealing with $\mathcal{MC}^k(\mu)$, it could be the case that $\mu \notin \mathcal{FM}^k(X)$. Thus, in the k-additive case, it could be the case that no such bottom element for $\mathcal{MC}^k(\mu)$ exists, so we cannot define $\underline{\mu}^0$. Similarly, the measure attaining value 1 for every subset is no longer in $\mathcal{MC}^k(\mu)$, whence we cannot define $\overline{\mu}^0$.
- The structure of the polytope $\mathcal{FM}^k(X)$ is not known for $k \geq 3$. Indeed, it has been proved in [24] that there are vertices of $\mathcal{FM}^k(X), k \geq 3$ that are not $\{0,1\}$-valued measures; moreover, we do not know the vertices of the polytope. However, it could be the case that some of these vertices are in $\mathcal{MC}^k(\mu)$. How can they be characterized?

We will study in this section the particular case in which $\mu \in \mathcal{FM}^k(X)$. Notice that in the n-additive general case, this condition trivially holds. Of course, if $\mu \in \mathcal{FM}^k(X)$, then $\mathcal{MC}^k(\mu) \neq \emptyset$ and has a bottom element (μ itself). Notice again that this was the situation for the n-additive case.

- Let us consider a total order on $\mathcal{P}^k_*(X) := \mathcal{P}^k(X) \setminus \{\emptyset\}$. Notice that in the general case, we have considered an order on $\mathcal{P}(X) \setminus \{\emptyset, X\}$. This total order allows us to range the subsets in $\mathcal{P}^k_*(X)$: $A_1 \prec A_2 \prec ... \prec A_r$, where $r = \sum_{i=1}^k \binom{n}{i}$.
- Next, take a partition $\mathcal{P} = \{\mathcal{U}, \mathcal{L}\}$ on $\mathcal{P}(X) \setminus \{X, \emptyset\}$, where $\mathcal{U}$ or $\mathcal{L}$ could be empty.
- **Initializing step**: Let us define $\mu_0 := \mu$
- **Iterating step**: For $i = 1$ until $i = r$ do:
  - If $A_i \in \mathcal{L}$, then $\mu_i = \mu_{i-1}$.
  - Otherwise $A_i \in \mathcal{U}$. It is easy to see that $\mu_{i-1} \in \mathcal{FM}^k(X)$.
    Let us consider the subset given by

$$\{\mu^* \in \mathcal{FM}^k(X) \,|\, \mu^* \geq \mu_{i-1},$$
$$\mu^*(A_j) = \mu_{i-1}(A_j), j = 1, ..., i-1\}. \quad (19)$$

    As $\mathcal{FM}^k(X)$ is a polytope, so is this set. Thus, for A_i, it follows that $\mu^*(A_i)$ can vary in an interval $[\mu_{i-1}(A_i), s]$. Take the subset of measures in the set satisfying $\mu^*(A_i) = s$. Let us denote this subset by $\mathcal{A}_0$.
    As before, $\mathcal{A}_0$ is a polytope. Then, for $\mu^* \in \mathcal{A}_0$, the value $\mu^*(A_{i+1})$ can vary in an interval $[x_1^-, x_1^+]$. We define

$$\mathcal{A}_1 := \{\mu^* \in \mathcal{A}_0 \,|\, \mu^*(A_{i+1}) = x_1^-\}. \quad (20)$$

    Now, the same can be done for $\mathcal{A}_1$ and considering A_{i+2}. We reiterate until $\mathcal{A}_j$ is just a singleton (notice that necessarily $\mathcal{A}_{r-i}$ is a singleton). This capacity is μ_i.

It can be checked that this algorithm coincides with the algorithm of the previous section for the n-additive case. Let us denote by μ_r the measure obtained in the last iteration.

Note that if $k = 1$ and μ is a probability, then the set $\mathcal{MC}^1(\mu) = \mathcal{C}(\mu) = \{\mu\}$ and the problem is trivial.

Now, the following can be proved:

Lemma 9 *If $\mu \in \mathcal{FM}^k(X)$, the measure μ_r obtained in the procedure is an extreme point of $\mathcal{MC}^k(\mu)$.*

However, this method does not obtain all the vertices of $\mathcal{MC}^k(\mu)$, as next example shows:

Example 1 *Consider $|X| = 4$ and the measure $u_{\{1,4\}}$ given by $u_{\{1,4\}}(A) = 1$ if $\{1,4\} \subseteq A$ and $u_{\{1,4\}}(A) = 0$ otherwise. Then, $u_{\{1,4\}} \in \mathcal{FM}^3(X)$. Consider the measure μ^* given by*

Subset	*1*	*2*	*3*	*4*	*1,2*	*1,3*	*1,4*
μ^*	*0*	*0*	*0*	*0*	*0.5*	*0.5*	*1*

Subset	*2,3*	*2,4*	*3,4*	*1,2,3*	*1,2,4*	*1,3,4*	*2,3,4*
μ^*	*0.5*	*0*	*0*	*0.5*	*1*	*1*	*1*

It has been proved in [24] that μ^* is an extreme point of $\mathcal{FM}^3(X)$. On the other hand, $\mu^* \geq u_{\{1,4\}}$, whence μ^* is an extreme point of $\mathcal{MC}^3(u_{\{1,4\}})$. Let us check that μ^* cannot be obtained through the previous algorithm, no matter the order considered.

Assume $A_1 \in \mathcal{U}$.

ISBN: 978-989-95079-6-8

- *If A_1 is a singleton, straightforward calculus shows that we can obtain $\mu_1(A_1) = 1$, whence μ^* could not be derived. The same holds if A_1 is a pair different from $\{1,4\}$ and $\{2,3\}$.*
- *Suppose $A_1 = \{2,3\}$ and consider the measure whose Möbius transform is given by*

$$m'(\{2,3\}) = 1, m'(\{1,4\}) = 1, m'(\{2,4\}) = 1. \quad (21)$$

In this case, we obtain a 3-additive measure and thus, it is possible to obtain $\mu_1(\{2,3\}) = 1$. Therefore, μ^* cannot be derived in this case.

- *If $A_1 = \{1,2,3\}$, then $m'(\{1,2\}) = 1$ gives $\mu_1(\{1,2,3\}) = 1$, whence μ^* cannot be derived in this case.*
- *For the other possibilities, we have $u_{\{1,4\}}(A_1) = 1$, whence the value is fixed.*

Consequently, it is not possible to obtain μ^* if $A_1 \in \mathcal{U}$. Thus, assume $A_1 \in \mathcal{L}$. This fixes $\mu_1(A_1) = u_{\{1,4\}}(A_1)$.

- *If $A_1 \in \{\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$, then $\mu_1(A_1) = 0$, whence μ^* cannot be recovered.*
- *For other possibilities, we fix a value either 0 or 1.*

But then, we can repeat the process for A_2 with the same results. Thus, μ^* cannot be obtained through the procedure.

6 Conclusions

In this paper we have proposed an algorithm to obtain the vertices of the k-additive monotone core. This procedure can be applied to any value of k. However, it seems that it could be very time-consuming. Next, we have derived the vertices of the n-additive core; these results generalize the Shapley-Ichiishi theorem for the general case. Finally, we have treated the possible extensions for the k-additive case.

An important problem arising in the k-additive case is the number of vertices of the k-additive core. For the general n-additive case, the set of vertices of $\mathcal{FM}(X)$ coincides with the n-th Dedekind number; simulations carried on for the k-additive case seem to show that the number of vertices of $\mathcal{FM}^k(X)$ is even greater, due to vertices that are not $\{0,1\}$-valued. This problem could make unfeasible to store all the vertices of the k-additive core in some cases.

Acknowledgment

This research has been supported in part by grant numbers MTM2007-61193 and CAM-UCM910707.

References

[1] O. Bondareva. Some applications of linear programming to the theory of cooperative games. *Problemy Kibernet*, (10):119–139, 1963.

[2] T. Driessen. *Cooperative Games*. Kluwer Academic, 1988.

[3] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5:587–607, 1997.

[4] M. Grabisch and P. Miranda. On the vertices of the k-additive core. *Discrete Mathematics*, (308):5204–5217, 2008.

[5] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, (38):325–339, 1967.

[6] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, (USA), 1976.

[7] P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *Ann. of Stat.*, 10:741–761, 1982.

[8] M. Wolfenson and T. L. Fine. Bayes-like decision making with upper and lower probabilities. *J. Amer. Statis. Assoc.*, (77):80–88, 1982.

[9] P. Walley. Coherent lower (and upper) probabilities. Technical Report 22, U. of Warwick, Coventry, (UK), 1981.

[10] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, (5):131–295, 1953.

[11] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, (17):263–283, 1989.

[12] G. Owen. *Game Theory*. Academic Press, 1995.

[13] D. Denneberg. *Non-additive measures and integral*. Kluwer Academic, 1994.

[14] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[15] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, (2):340–368, 1964.

[16] M. Grabisch. k-order additive discrete fuzzy measures. In *Proceedings of 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1345–1350, Granada (Spain), 1996.

[17] L. S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1:11–26, 1971.

[18] T. Ichiishi. Super-modularity: Applications to convex games and to the Greedy algorithm for LP. *Journal of Economic Theory*, (25):283–286, 1981.

[19] P. Miranda and M. Grabisch. k-balanced games and capacities. *European Journal of Operational Research*, (Accepted).

[20] T. Matsui. NP-completeness of non-adjacency relations on some 0-1-polytopes. In *Lecture Notes in Operations Research, 1*, pages 249–258. 1995.

[21] C. Papadimitriou. The adjacency relation on the travelling salesman polytope is NP-complete. *Mathematical Programming*, 14(1), 1978.

[22] D. Radojevic. The logical representation of the discrete Choquet integral. *Belgian Journal of Operations Research, Statistics and Computer Science*, 38(2–3):67–89, 1998.

[23] E. F. Combarro and P. Miranda. On the polytope of non-additve measures. *Fuzzy Sets and Systems*, 159(16):2145–2162, 2008.

[24] P. Miranda, E.F. Combarro, and P. Gil. Extreme points of some families of non-additive measures. *European Journal of Operational Research*, 33(10):3046–3066, 2006.

ISBN: 978-989-95079-6-8

Resolution of fuzzy relation equations with sup-inf composition over complete Brouwerian lattices—a review *

Xue-ping Wang[1] Feng Sun[1]

1.Department of Mathematics, Sichuan Normal University
Chengdu, Sichuan 610066, People's Republic of China
Email: xpwang1@hotmail.com, sunfeng1005@163.com

Abstract— *This paper restates the results on fuzzy relation equations with sup-inf composition from the viewpoint of decomposition, presents a way to describe the solution set of fuzzy relation equations, and shows a necessary and sufficient condition, which partly answers the open problem for existence of minimal solutions over complete Brouwerian lattices.*

Keywords— Complete Brouwerian lattice, fuzzy relation equation, minimal solution, solution set, sup-inf composition.

1 Introduction and preliminaries

Let I, J be index sets and $A = (a_{ij})_{I\times J}$ be a coefficient matrix, $B = (b_i)_{i\in I}^T$ be a constant column vector (the sign "T" denotes the "transpose"). Then $A \odot X = B$ or

$$\bigvee_{j\in J} (a_{ij} \wedge x_j) = b_i, i \in I \tag{1}$$

is called a *fuzzy relation equation* assigned on a lattice L, where $\odot$ denotes the sup-inf composite operation, and all x_j, b_i, a_{ij}'s are in L. An X which satisfies (1) is called a *solution* of (1), the *solution set* of (1) is denoted by $\mathscr{X}_1 = \{X : A \odot X = B\}$. A special case of (1) is as follows: $A \odot X = b$ or

$$\bigvee_{j\in J} (a_j \wedge x_j) = b, \tag{2}$$

where $b \in L$, $A = (a_j)_{j\in J}$ is a row vector. Denote $\mathscr{X}_2 = \{X : A \odot X = b\}$ the solution set of (2).

The solvability of fuzzy relation equations in complete Brouwerian lattices was first proposed in connection with medical diagnosis problems in [24]. Sanchez [24] showed that every solvable fuzzy relation equation assigned on complete Brouwerian lattices has the greatest solution. Since then, the resolution of fuzzy relation equations over lattices has been a theme of continuous interest in fuzzy inference and fuzzy systems theory. A number of works in this area were published (see, e.g. [2,6,8,9,12,17,19,20,23,25,27,28,29,30,32,33,38]). Many results are obtained when L is a linear lattice (see, e.g. [3, 7, 13, 14, 15, 18, 26]). There are also many results considered fuzzy relation equations over semi-linear spaces (see [10, 16, 21, 22] for details), and many works dealt with the more general fuzzy relation equations on complete lattices, such as $sup_{j\in J}\mathscr{T}(a_j, x_j) = b$, where $\mathscr{T}$ is a t-norm (see [5]) or a pseudo-t-norm (see [11,34]) or a conjunctor (see [31,37]).

*Supported by the National Natural Science Foundation of China (No.10671138).

This paper mainly focuses on the fuzzy relation equations over complete Brouwerian lattices.

We first recall some definitions and results of lattice theory which will be used in the sequel. Let $(P, \leq)$ be a partially ordered set and $X \subseteq P$. $p \in X$ is a *minimal* element of X if there is no $x \in X$ such that $x < p$. The *greatest* element of X is an element $g \in X$ such that $x \leq g$ for all $x \in X$. A *lattice* is a poset $L = (L, \leq)$ any two of whose elements have a g.l.b. or "meet" denoted by $x \wedge y$, and a l.u.b. or "join" denoted by $x \vee y$. A lattice L is *complete* when each of its subset T has a l.u.b. and a g.l.b. in L. An element a of a lattice L is *join-irreducible* if $x \vee y = a$ implies $x = a$ or $y = a$. A lattice is *Brouwerian* if for any pair of elements $a, b \in L$, the greatest element $x \in L$, denoted by $a\alpha b$, satisfying the inequality $a \wedge x \leq b$ exists. It is easy to verify the following properties: for any a, b, c in a complete Brouwerian lattice L, $b \leq a\alpha(a \wedge b), a\alpha(b \wedge c) = (a\alpha b) \wedge (a\alpha c), a \wedge (a\alpha b) = a \wedge b$. It is well known that any Brouwerian lattice L is distributive, and a complete lattice L is Brouwerian if and only if (iff)

$$x \wedge (\bigvee_{i\in I} y_i) = \bigvee_{i\in I} (x \wedge y_i)$$

for any $x \in L$ and any family of elements $\{y_i \in L : i \in I\}$. Let p be a join-irreducible element of a distributive lattice L. Then $p \leq \bigvee_{i=1}^{k} x_i$ iff $p \leq x_i$ for some x_i. Some other definitions and results of lattice theory which we do not list here are from [1, 4].

2 The sets of solutions for (1)

Throughout the paper, L is always assumed to be a complete Brouwerian lattice with universal bounds 0 and 1, I and J are infinite index sets, unless otherwise specified. Denote $X^* = (x_j^*)_{j\in J}^T$ a greatest solution, $\underline{k} = \{1, 2, \cdots, k\}$ for any positive integer k and $A - B = \{x \in A : x \notin B\}$, where A and B are two (crisp) sets.

It is trivial that $a \wedge x = b$ (resp. $a \vee x = b$) is solvable iff $a \geq b$ (resp. $a \leq b$). Denote $[a]_b$ (resp. $\overline{[a]}_b$) its solution set. For the solvable case, if L is a distributive lattice then $[a]_b = [b) - \bigcup_{b<d\leq a} [d)$ and $\overline{[a]}_b = (b] - \bigcup_{a\leq d<a} (d]$ (see [36]); if L is a complete Brouwerian lattice then $[a]_b = [b, a\alpha b]$ and $\overline{[a]}_b = (b] - \bigcup_{a\leq d<a} (d]$, where $[a) = \{x \in L : x \geq a\}$ and $(a] = \{x \in L : x \leq a\}$ (see [12]). In what follows, for any $i \in I$, denote

$$\mathscr{R}^i = \{R^i = (q_{ij})_{j\in J}^T : b_i = \bigvee_{j\in J} q_{ij}, q_{ij} \leq a_{ij}, \forall j \in J\},$$

ISBN: 978-989-95079-6-8

$\mathcal{X}^i$ the solution set of $\bigvee_{j\in J}(a_{ij}\wedge x_j)=b_i$ and $\mathcal{Q}=\{Q=(q_{ij})_{I\times J}:\bigvee_{j\in J}q_{ij}=b_i,\forall i\in I;\bigcap_{i\in I}[a_{ij}]_{q_{ij}}\neq\emptyset,\forall j\in J\}$. Then we can verify the following criteria of solvability easily.

Theorem 2.1 (Di Nola [9]) *$\mathcal{X}_1\neq\emptyset$ iff $\mathcal{Q}\neq\emptyset$.*

Though we omit its proof, it should be pointed out that the proof of Theorem 2.1 suggests a way to build other elements, different from the greatest element $X^*=(\bigwedge_{i\in I}(a_{ij}\alpha b_i))^T_{j\in J}$, of $\mathcal{X}_1$. What is more, Theorem 2.1 holds when L is a complete lattice.

Note that for any $i\in I$, the i-th row of any $Q\in\mathcal{Q}$ forms a decomposition of the i-th component of B. Moreover, $b_i=\bigvee_{j\in J}q_{ij}$ and $q_{ij}\leq a_{ij}$, $\forall j\in J$, therefore $\bigvee_{j\in J}(a_{ij}\wedge q_{ij})=\bigvee_{j\in J}q_{ij}=b_i$, $\forall i\in I$, thus the transpose of the i-th row of Q solves the i-th equation in (1), i.e. $\mathcal{R}^i\subseteq\mathcal{X}^i,\forall i\in I$.

Let $\mathcal{Q}_j=\bigcap_{i\in I}[a_{ij}]_{q_{ij}}$ for any $j\in J$. Then the following theorem is true.

Theorem 2.2 (Di Nola [9]) *Let $\mathcal{X}_1\neq\emptyset$. Then for any $j\in J$, $\mathcal{Q}_j=[\bigvee_{i\in I}q_{ij},\bigwedge_{i\in I}(a_{ij}\alpha q_{ij})]$, where $(q_{ij})_{I\times J}\in\mathcal{Q}$. Further, put $R=(r_j)^T_{j\in J}$ with $r_j\in\mathcal{Q}_j$ for all $j\in J$, then $R\in\mathcal{X}_1$.*

Note that Theorem 2.2 gives the solution set of (1) when $|J|=1$, i.e. $[\bigvee_{i\in I}b_i,\bigwedge_{i\in I}(a_i\alpha b_i)]$ is the solution set of the system of equations $a_i\wedge x=b_i$, $i\in I$. Then the following equivalent conditions is straightforward.

Corollary 2.1 (Han [12]) *For the system of equations $a_i\wedge x=b_i$, $i\in\underline{m}$ ($m\geq 2$), the following conditions are equivalent:*
(1) The system of equations $a_i\wedge x=b_i$, $i\in\underline{m}$ is solvable;
(2) $a_i\geq b_i$ and $a_j\wedge b_i=b_j\wedge b_i$ for all $i,j\in\underline{m}$;
(3) $\bigvee_{i=1}^m b_i$ is its solution;
(4) $\bigwedge_{i=1}^m(a_i\alpha b_i)$ is its solution.
Further, the solution set is $[\bigvee_{i=1}^m b_i,\bigwedge_{i=1}^m(a_i\alpha b_i)]$.

Remark 2.1 Conditions (1), (2) and (3) in Corollary 2.1 are also equivalent when the system is assigned on distributive lattices.

By Remark 2.1, the following corollary is obvious when L is a distributive lattice.

Corollary 2.2 (Zhang [35]) *The system of equations $a_i\wedge x=b_i$, $i\in\underline{m}$ is solvable iff $\bigvee_{i=1}^m b_i$ is its solution. And the solution set is $[\bigvee_{i=1}^m a_i]_{\bigvee_{i=1}^m b_i}$.*

Combining Corollaries 2.1 and 2.2, we get a property of α operation as follows:

Proposition 2.1 *If $a_i\geq b_i$ and $a_j\wedge b_i=b_j\wedge b_i$ for all $i,j\in\underline{m}$, then $(\bigvee_{i=1}^m a_i)\alpha(\bigvee_{i=1}^m b_i)=\bigwedge_{i=1}^m(a_i\alpha b_i)$.*

Proof. If $a_i\geq b_i$ and $a_j\wedge b_i=b_j\wedge b_i$ for all $i,j\in\underline{m}$, then $a_i\wedge x=b_i$, $i\in\underline{m}$ is solvable and its solution set is $[\bigvee_{i=1}^m b_i,\bigwedge_{i=1}^m(a_i\alpha b_i)]$ by Corollary 2.1. Therefore from Corollary 2.2, we have $[\bigvee_{i=1}^m b_i,(\bigvee_{i=1}^m a_i)\alpha(\bigvee_{i=1}^m b_i)]=[\bigvee_{i=1}^m a_i]_{\bigvee_{i=1}^m b_i}=[\bigvee_{i=1}^m b_i,\bigwedge_{i=1}^m(a_i\alpha b_i)]$. Thus $(\bigvee_{i=1}^m a_i)\alpha(\bigvee_{i=1}^m b_i)=\bigwedge_{i=1}^m(a_i\alpha b_i)$. □

Notice that, in general, $(\bigvee_{i=1}^m a_i)\alpha(\bigvee_{i=1}^m b_i)=\bigwedge_{i=1}^m(a_i\alpha b_i)$ does not always hold.

Example 2.1 Let N be the set of nonnegative integers. We define $a\wedge b=$ l.c.m.$\{a,b\}$, $a\vee b=$ g.c.d.$\{a,b\}$, $a\leq b$ iff a is multiple of b, where $a,b\in N$ and l.c.m. (resp. g.c.d.) stands for the smallest (resp. greatest) common multiple (resp. divisor) between a and b. Then $L=(N,\wedge,\vee,\leq)$ is a complete Brouwerian lattice with operator "α" given by $a\alpha b=$ g.c.d.$\{x\in N:a\wedge x\leq b\}$ for any $a,b\in N$. $(2\vee 4)\alpha(3\vee 5)=2\alpha 1=1$, but $(2\alpha 3)\wedge(4\alpha 5)=3\wedge 5=15$.

When L is a distributive lattice, $I=\underline{m}$ and $J=\underline{n}$, the next two corollaries are straightforward by Theorem 2.1 and Corollary 2.2.

Corollary 2.3 (Zhang [35]) *Equation (1) is solvable iff there exists a matrix $Q=(q_{ij})_{m\times n}$ such that $\bigvee_{j=1}^n q_{ij}=b_i$, and $a_{ij}\wedge(\bigvee_{k=1}^m q_{kj})=q_{ij}$, $i\in\underline{m}$ and $j\in\underline{n}$. Denote (Q) the set of such Q. Then $\mathcal{X}_1=\bigcup_{Q\in(Q)}H_Q$, where $H_Q=\{(r_j)^T_{j\in\underline{n}}:r_j\in[\bigvee_{i=1}^m a_{ij}]_{\bigvee_{i=1}^m q_{ij}},\forall j\in\underline{n}\}$.*

Corollary 2.4 (Zhang [36]) *For given $B'=(b'_i)^T_{i\in\underline{m}}$, $A'=(a'_i)_{i\in\underline{n}}$ and B, there exists X such that $B'\odot A'\odot X=B$ iff $\bigvee_{j=1}^n a'_j\geq\bigvee_{i=1}^m b_i$ and $b'_k\wedge(\bigvee_{i=1}^m b_i)=b_k$, $\forall k\in\underline{m}$.*

Proof. If there exists $X=(x_j)^T_{j\in\underline{n}}$ such that $B'\odot A'\odot X=B$, then $\bigvee_{j=1}^n(a'_j\wedge x_j)$ is a solution of $b'_i\wedge x=b_i$, $i\in\underline{m}$, thus $\bigvee_{j=1}^n a'_j\geq\bigvee_{j=1}^n(a'_j\wedge x_j)\geq\bigvee_{i=1}^m b_i$. Further, $b'_k\wedge(\bigvee_{i=1}^m b_i)=b_k$, $\forall k\in\underline{m}$ by Corollary 2.2.

Conversely, if $\bigvee_{j=1}^n a'_j\geq\bigvee_{i=1}^m b_i$, then $\bigvee_{j=1}^n[a'_j\wedge(\bigvee_{i=1}^m b_i)]=(\bigvee_{j=1}^n a'_j)\wedge(\bigvee_{i=1}^m b_i)=\bigvee_{i=1}^m b_i$, therefore $X=(x_j)^T_{j\in\underline{n}}$ with $x_j=\bigvee_{i=1}^m b_i$ satisfies $B'\odot A'\odot X=B$ since $b'_k\wedge(\bigvee_{i=1}^m b_i)=b_k$, $\forall k\in\underline{m}$. □

Note that Corollary 2.3 was generalized in [37] to fuzzy relation equations with sup-conjunctor composition.

From Theorem 2.2 we know that $R=(r_j)^T_{j\in J}\in\mathcal{X}_1$ with $r_j\in[\bigvee_{i\in I}q_{ij},\bigwedge_{i\in I}(a_{ij}\alpha q_{ij})]$ for all $j\in J$. Particularly, take $R=(\bigvee_{i\in I}q_{ij})^T_{j\in J}$ with $(q_{ij})_{I\times J}\in\mathcal{Q}$ and $\mathcal{R}=\{R=(r_j)^T_{j\in J}:r_j=\bigvee_{i\in I}q_{ij},\forall j\in J,(q_{ij})_{I\times J}\in\mathcal{Q}\}$. It is easy

ISBN: 978-989-95079-6-8

to verify $\mathscr{R} \subseteq \mathscr{X}_1$. For any $X = (x_j)_{i\in J}^T \in \mathscr{X}_1$, we have $R^i = (q_{ij})_{j\in J}^T = (a_{ij} \wedge x_j)_{j\in J}^T \in \mathscr{R}^i, \forall i \in I$. Therefore, $\mathscr{R}$ possesses similar properties as $\mathscr{X}_1$. For instance, we have:

Proposition 2.2 *If $R^1 = (r_j^1)_{j\in J}^T$, $R^2 = (r_j^2)_{j\in J}^T \in \mathscr{R}$ and $R = (r_j)_{j\in J}^T$ such that $R^1 \leq R \leq R^2$, then $R \in \mathscr{R}$.*

Proposition 2.3 *If $R^1 = (r_j^1)_{j\in J}^T$, $R^2 = (r_j^2)_{j\in J}^T \in \mathscr{R}$, then $R^1 \vee R^2 \in \mathscr{R}$.*

Proposition 2.4 *$\mathscr{R} \neq \emptyset$ iff $R^* = ((\bigvee_{i\in I} a_{ij}) \wedge [\bigwedge_{i\in I} (a_{ij}\alpha b_i)])_{j\in J}^T \in \mathscr{R}$. Further, R^* is the greatest element of $\mathscr{R}$.*

Note that, from Proposition 2.4 we know if $|\mathscr{X}_1| = 1$ then $\bigwedge_{i\in I} (a_{ij}\alpha b_i) \leq \bigvee_{i\in I} a_{ij}$ for all $j \in J$, since $R^* \in \mathscr{R} \subseteq \mathscr{X}_1$ must be equal to the greatest element of $\mathscr{X}_1$ (otherwise, $|\mathscr{X}_1| > 1$). By the definition of $\mathscr{R}$, we can build an element of $\mathscr{X}_1$ from $\mathscr{R}^i$, $i \in I$, which may be different from the greatest element in $\mathscr{X}_1$.

Example 2.2 Let L be the lattice considered in Example 2.1. Consider the fuzzy relational equation

$$\begin{cases} (3 \wedge x_1) \vee (7 \wedge x_2) = 3, \\ (6 \wedge x_1) \vee (8 \wedge x_2) = 12. \end{cases} \tag{3}$$

We can see $(12, 21)^T \in \mathscr{R}^1$, $(12, 24)^T \in \mathscr{R}^2$, i.e. $12 \vee 21 = 3$, $12 \vee 24 = 12$. And further, $4 \in [3]_{12} \cap [6]_{12}$, $3 \in [7]_{21} \cap [8]_{24}$, then $(12 \vee 12, 21 \vee 24)^T = (12, 3)^T \in \mathscr{R} \subseteq \mathscr{X}_1$, which is different from the the greatest element $(4, 3)^T$.

Let $\mathscr{R} \neq \emptyset$. From the definition of $\mathscr{R}$, we know that R is an element of $\mathscr{R}$ if and only if there exist $R^i = (q_{ij})_{j\in J}^T \in \mathscr{R}^i$, $\forall i \in I$ such that $R = \bigvee_{i\in I} R^i$ and $\bigcap_{i\in I} [a_{ij}]_{q_{ij}} \neq \emptyset$, $\forall j \in J$. This result has a generalized version considered over semi-linear spaces (see [10, 21, 22] for details) as follows:

Theorem 2.3 (Nosková [16]) *Let (1) be solvable. Then X is a solution to (1) iff there exist $X_i \in \mathcal{X}^i$, $i \in \underline{m}$, such that $X = \bigvee_{i=1}^m X_i$ and $\bigvee_{i=1}^m X_i \leq X^*$.*

For every $i \in I$, denote $\mathcal{S}^i = \bigcup_{R^i \in \mathscr{R}^i} S_{R^i}$, where $R^i = (q_{ij})_{j\in J}^T \in \mathscr{R}^i$ and $S_{R^i} = \{S = (s_j)_{j\in J}^T : s_j \in [a_{ij}]_{q_{ij}}, \forall j \in J\}$. Then $\mathscr{R}^i \subseteq \mathcal{S}^i$ since $q_{ij} \in [a_{ij}]_{q_{ij}}$. Further,

Theorem 2.4 *$\mathcal{S}^i = \mathcal{X}^i$, $\forall i \in I$.*

Proof. For any $i \in I$, if $S = (s_j)_{j\in J}^T \in \mathcal{S}^i$, then there exists $R^i = (q_{ij})_{j\in J}^T \in \mathscr{R}^i$ such that $S \in S_{R^i}$, i.e. $s_j \in [a_{ij}]_{q_{ij}}, \forall j \in J$. Thus $\bigvee_{j\in J} (a_{ij} \wedge s_j) = \bigvee_{j\in J} q_{ij} = b_i$. Therefore $S \in \mathcal{X}^i$, i.e. $\mathcal{S}^i \subseteq \mathcal{X}^i$. Vice versa, for any $i \in I$, if $X = (x_j)_{j\in J}^T \in \mathcal{X}^i$, then $\bigvee_{j\in J} (a_{ij} \wedge x_j) = b_i$. Let $R^i = (a_{ij} \wedge x_j)_{j\in J}^T$, i.e. $R^i \in \mathscr{R}^i$. Then $X \in S_{R^i} \subseteq \mathcal{S}^i$ since $x_j \in [a_{ij}]_{a_{ij}\wedge x_j}$, $\forall j \in J$. Thus $\mathcal{X}^i \subseteq \mathcal{S}^i$, together with $\mathcal{S}^i \subseteq \mathcal{X}^i$ we conclude $\mathcal{S}^i = \mathcal{X}^i$. □

Since $\mathscr{X}_1 = \bigcap_{i\in I} \mathcal{X}^i$, then the following theorem is straightforward by Theorem 2.4.

Theorem 2.5 $\mathscr{X}_1 = \bigcap_{i\in I} \bigcup_{R^i \in \mathscr{R}^i} S_{R^i}$.

From Theorems 2.4 and 2.5, we know that it is important to find all elements of $\mathscr{R}^i$, $i \in I$, in order to determine $\mathscr{X}_1$.

In the sequel, we will consider a method to determine all elements of $\mathscr{R}^i$, $i \in I$ then describe the solution set of (1).

Theorem 2.6 *For any $i \in I$, if $\mathscr{R}^i \neq \emptyset$ and $R^* = (q_{ij}^*)_{j\in J}^T = (a_{ij} \wedge b_i)_{j\in J}^T$ is the greatest element of $\mathscr{R}^i$, then every element $R = (q_{ij})_{j\in J}^T$ of $\mathscr{R}^i$ is determined by*

$$q_{ij} \in \overline{[\bigvee_{k\in J, k\neq j} q_{ik}^j]}_{b_i} \bigcap (a_{ij}],$$

where

$$q_{ik}^j = \begin{cases} q_{ik}, & k \in D, \\ q_{ik}^*, & k \in J - D \end{cases}$$

and $D = \{j \in J : q_{ij}$ has been determined $\}$.

Proof. For any $R = (q_{ij})_{j\in J}^T \in \mathscr{R}^i$ and $j \in J$, $b_i = \bigvee_{k\in J} q_{ik}^* \geq q_{ij} \vee (\bigvee_{k\in J, k\neq j} q_{ik}^j) \geq q_{ij} \vee (\bigvee_{k\in J, k\neq j} q_{ik}) = b_i$, then $q_{ij} \in \overline{[\bigvee_{k\in J, k\neq j} q_{ik}^j]}_{b_i} \bigcap (a_{ij}]$ since $q_{ij} \leq a_{ij}, \forall j \in J$ by the definition of $\mathscr{R}^i$. Vice versa is straightforward by the definition of q_{ij} and q_{ik}^j. □

By Theorems 2.5 and 2.6 the following results is obvious.

Corollary 2.5 *For any $i \in I$, if $R^* = (q_{ij}^*)_{j\in J}^T = (a_{ij} \wedge b_i)_{j\in J}^T$ is the greatest element of $\mathscr{R}^i$, then every element $X = (x_j)_{j\in J}^T$ of $\mathscr{X}_1$ is determined by*

$$x_j \in \bigcap_{i\in I} \bigcup_{q_{ij} \in \overline{[\bigvee_{k\in J, k\neq j} q_{ik}^j]}_{b_i} \bigcap (a_{ij}]} [a_{ij}]_{q_{ij}},$$

where

$$q_{ik}^j = \begin{cases} q_{ik}, & k \in D, \\ q_{ik}^*, & k \in J - D \end{cases}$$

and $D = \{j \in J : q_{ij}$ has been determined $\}$.

Note that Corollary 2.5 generalizes Theorem 3.2 in [12].

As for (2), the corresponding $\mathscr{R}$ mentioned before is nothing but

$$\mathscr{R}_2 = \{R = (r_j)_{j\in J}^T : b = \bigvee_{j\in J} r_j, r_j \leq a_j, \forall j \in J\}.$$

When $J = \underline{n}$, let $b = s_1 = r_1 \vee s_2, s_2 = r_2 \vee s_3, \cdots, s_{n-1} = r_{n-1} \vee s_n, s_n = r_n$. Therefore, $r_1 \vee \cdots \vee r_n = b$ and the following theorem when (2) is assigned on distributive lattices is straightforward, since Theorem 2.4 also holds in distributive lattices in finite case.

Theorem 2.7 (Zhang [35]) $\mathscr{R}_2 = \bigcup_{b=s_1\geq s_2\geq\cdots\geq s_n} \{R = (r_j)_{j\in\underline{n}}^T : r_j \in \overline{[s_{j+1}]}_{s_j} \bigcap (a_j], \forall j \in \underline{n-1}, r_n = s_n\}$. *And* $\mathscr{X}_2 = \bigcup_{b=s_1\geq s_2\geq\cdots\geq s_n} \{X = (x_j)_{j\in\underline{n}}^T : x_j \in [a_j]_{r_j}, \forall j \in \underline{n}, r_j \in \overline{[s_{j+1}]}_{s_j} \bigcap (a_j], \forall j \in \underline{n-1}, r_n = s_n\}$.

ISBN: 978-989-95079-6-8

3 Conditions for existence of minimal solutions to (1)

As it was mentioned in [8] that the determination of minimal elements in $\mathscr{X}_1$, i.e. minimal solutions to (1), when (1) is assigned on complete Brouwerian lattices remains open in finite case as well as in infinite case. In this section we will consider the determination of minimal solutions.

First, we have:

Theorem 3.1 (Di Nola [9]) *If $\mathscr{R}$ has minimal elements then these elements are minimal in $\mathscr{X}_1$, and vice versa.*

As a comment to Theorem 3.1 we have to say we do not know if $\mathscr{X}_1$ (or $\mathscr{R}$) has or has not minimal elements. However, if one suppose the existence of minimal elements in $\mathscr{X}_1$, then these elements must be sought among the minimal elements of $\mathscr{R}$.

Minimal solutions do not always exist, here is an example borrowed from [29].

Example 3.1 Let L be the lattice considered in Example 2.1. If $A = (3, 7)$ and $b = 2$, then (2) has a solution $X = (2, 2)^T$ but no minimal solution since for any $X = (x_1, x_2)^T \in \mathscr{X}_2$, $X' = (2x_1, x_2)^T$ is also an element of $\mathscr{X}_2$ and $X' \leq X$ but $X' \neq X$.

Let $\mathscr{X}_1 \neq \emptyset$ and $\gamma_{j_0}(X) = \{l \in L : (a_{ij_0} \wedge l) \vee [\bigvee_{j\in J, j\neq j_0} (a_{ij} \wedge x_j)] = b_i, \forall i \in I\}$ for every $X = (x_j)^T_{j\in J} \in \mathscr{X}_1$ and $j_0 \in J$. Then:

Theorem 3.2 *For any $X = (x_j)^T_{j\in J} \in \mathscr{X}_1$, if $X' = (x'_j)^T_{j\in J} \in \mathscr{X}_1$ and $X' \geq X$ then $x_j \in \gamma_j(X')$ for all $j \in J$.*

Proof. For any $X = (x_j)^T_{j\in J} \in \mathscr{X}_1$ and $j_0 \in J$, put $\overline{X} = (\overline{x}_j)^T_{j\in J}$ with

$$\overline{x}_j = \begin{cases} x_{j_0}, & j = j_0, \\ x'_j, & j \neq j_0. \end{cases}$$

Then $X \leq \overline{X} \leq X'$, and $\overline{X} \in \mathscr{X}_1$. Therefore, $x_{j_0} \in \gamma_{j_0}(X')$. □

The following theorem gives a criterion to determine when a given solution is minimal:

Theorem 3.3 (Di Nola [9]) *$X = (x_j)^T_{j\in J}$ is a minimal element of $\mathscr{X}_1$ iff $x_{j_0} = \min \gamma_{j_0}(X)$ for every $j_0 \in J$.*

For a given element $R = (r_j)^T_{j\in J} = (\bigvee_{i\in I} q_{ij})^T_{j\in J}$ in $\mathscr{R}$ and each $j_0 \in J$, denote $\overline{\gamma}_{j_0}(R) = \{t \in L : t = \bigvee_{i\in I} t_{ij_0}, (\bigvee_{j\in J, j\neq j_0} q_{ij}) \vee t_{ij_0} = b_i, \forall i \in I\}$, then from Theorems 3.2 and 3.3 the following two theorems are obvious.

Theorem 3.4 *$R = (r_j)^T_{j\in J}$ is a minimal element of $\mathscr{R}$ iff $r_{j_0} = \min \overline{\gamma}_{j_0}(R)$ for all $j_0 \in J$.*

Theorem 3.5 *For any $R = (r_j)^T_{j\in J} \in \mathscr{R}$, if $R' = (r'_j)^T_{j\in J} \in \mathscr{R}$ and $R' \geq R$ then $r_j \in \overline{\gamma}_j(R')$ for all $j \in J$.*

As an immediate consequence of Theorem 3.3, the following theorem on the unicity of solution is obvious.

Theorem 3.6 (Di Nola [9]) *$|\mathscr{X}_1| = 1$ iff for every $j_0 \in J$, $x^*_{j_0} = \bigwedge_{i\in I} (a_{ij_0} \alpha b_i) = \min \gamma_{j_0}(X^*)$.*

In fact, if $\mathscr{X}_1 \neq \emptyset$, then from Theorem 2.2 we have that $|\mathscr{X}_1| = 1$ if and only if $\bigvee_{i\in I} q_{ij} = \bigwedge_{i\in I} (a_{ij} \alpha q_{ij})$ for any $j \in J$ and $(q_{ij})_{I\times J} \in \mathscr{Q}$ (see also Theorem 6 in [25]).

By Theorem 3.1, the following theorem is straightforward.

Theorem 3.7 (Wang [30]) *If $X_* = (x_{j_*})^T_{j\in J}$ is a minimal element of $\mathscr{X}_2$, then $a_j \wedge x_{j_*} = x_{j_*}$ for all $j \in J$, i.e. $b = \bigvee_{j\in J} x_{j_*}$ and $x_{j_*} \leq a_j, \forall j \in J$.*

Though minimal solutions do not always exist, when adding some assumptions to B, we get some interesting results. Here is a sufficient condition for existence of a minimal solution to (1) when I is finite and J is infinite.

Theorem 3.8 (Wang [28]) *If every component of B is a compact element with an irredundant finite join-decomposition, then for each $X \in \mathscr{X}_1$ there exists a minimal element $X_* \in \mathscr{X}_1$ such that $X_* \leq X$.*

Thus the following theorem follows from Theorems 3.1 and 3.8 easily.

Theorem 3.9 *Suppose that each component of B is compact and has an irredundant finite join-decomposition. Then R is minimal in $\overline{\mathscr{R}}$ iff it is minimal in $\mathscr{X}_1$, where $\overline{\mathscr{R}} = \{R : R = \bigvee_{i\in I} R^i, R^i = (q_{ij})^T_{j\in J}$ is minimal in $\mathscr{R}^i, \forall i \in I; \bigcap_{i\in I} [a_{ij}]_{q_{ij}} \neq \emptyset, \forall j \in J\}$.*

Note that Theorem 3.8 was generalized in [31] to fuzzy relation equations with sup-conjunctor composition. With Theorem 3.9 in hands, it is easy to see that if I is finite and every component of B is compact and has an irredundant finite join-decomposition, then $\mathscr{X}_1$ is completely determined by the greatest solution and all minimal solutions. Therefore, Theorem 3.9 generalizes Theorem 3.11 in [8] and the corresponding result in [13] considered over $[0, 1]$. Another generalized version of Theorem 3.11 in [8] is Theorem 4 in [16] considered fuzzy relation equations over semi-linear spaces. When J is also finite, the assumption of compact in Theorem 3.8 can be removed (see Theorem 7.1 in [29]). Further, the following theorem gives the number of minimal solutions of (2) when both I and J are finite.

Theorem 3.10 (Wang [28]) *If $\mathscr{X}_2 \neq \emptyset$ and b has an irredundant finite join-decomposition $\bigvee_{i=1}^{k} p_i$, then the number of minimal solutions of (2) is $\prod_{i\in \underline{k}} |G(p_i)|$, where $G(p_i) = \{j \in \underline{n} : a_j \geq p_i\}$. Further, all minimal solutions $X = (x_j)^T_{j\in J}$ are determined by $x_j = \bigvee_{f_X(i)=j} p_i$ for any mapping $f_X \in \prod_{i\in \underline{k}} G(p_i)$.*

Let both I and J be finite. If for any $i \in I$, b_i has an irredundant finite join-decomposition $\bigvee_{t=1}^{k_i} p_{it}$, then for any $i \in I$, $t_i \in \underline{k_i}$ we define $G(p_{it_i}) = \{j \in J : a_{ij} \geq p_{it_i}\}$ and $F_i = \{f_i : f_i \in \prod_{t_i\in \underline{k_i}} G(p_{it_i}), \bigvee_{f_i(t_i)=j} p_{it_i} \leq \bigwedge_{i\in I} (a_{ij} \alpha b_i), \forall j \in J\}$. Then:

Theorem 3.11 *Let $\mathscr{X}_1 \neq \emptyset$. Then $|\mathscr{X}_1| = 1$ iff for any $j \in J$ and $f_i \in F_i$, $\bigvee_{i\in I} \bigvee_{f_i(t_i)=j} p_{it_i} = \bigwedge_{i\in I} (a_{ij} \alpha b_i)$.*

ISBN: 978-989-95079-6-8

Proof. Note that $X^* = (\bigwedge_{i\in I}(a_{ij}\alpha b_i))_{j\in J}^T \in \mathscr{X}_1$ since $\mathscr{X}_1 \neq \emptyset$. For any $i \in I$ and $f_i \in F_i$, from Theorem 3.10 and the definition of F_i we know that $(\bigvee_{f_i(t_i)=j} p_{it_i})_{j\in J}^T \leq X^*$ and $(\bigvee_{f_i(t_i)=j} p_{it_i})_{j\in J}^T$ is minimal in $\mathcal{X}^i$. Thus $\{X = (x_j)_{j\in J}^T : x_j = \bigvee_{i\in I}\bigvee_{f_i(t_i)=j} p_{it_i}, f_i \in F_i\}$ contains all minimal elements of $\mathscr{X}_1$ from Theorem 3.9. Therefore, the thesis is straightforward. □

Note that Theorem 3 in [7] that considers the unique solvability of fuzzy relation equations over linear lattices and it is just a special case of Theorem 3.11.

When $|I| = 1$ and J is infinite, a necessary and sufficient condition for existence of a minimal solution is as follows:

Theorem 3.12 (Wang [30]) *Let $\mathscr{X}_2 \neq \emptyset$. Then for each $X \in \mathscr{X}_2$ there exists a minimal element X_* of $\mathscr{X}_2$ such that $X_* \leq X$ iff there is a subset B of L with B satisfying:*
(i) $\bigvee B = b$;
(ii) For each $p \in B$, if $p \neq 0$ then $b \neq \bigvee(B - \{p\})$;
(iii) For each $X = (x_j)_{j\in J}^T \in \mathscr{X}_2$ and each $p \in B$ there is an index $k \in J$ such that $p \leq a_k \wedge x_k$.

If I is finite, then from Theorem 3.12 we have

Theorem 3.13 (Wang [30]) *If $\mathscr{X}_1 \neq \emptyset$ and every component b_i, $i \in I$, of B is compact and for each b_i, $i \in I$, there exists a subset B_i of L such that:*
(i) $\bigvee B_i = b_i$;
(ii) For each $p_{it} \in B_i$, if $p_{it} \neq 0$ then $b_i \neq \bigvee(B_i - \{p_{it}\})$;
(iii) For each $X = (x_j)_{j\in J}^T \in \mathcal{X}^i$ and each $p_{it} \in B_i$ there is an index $k \in J$ such that $p_{it} \leq a_{ik} \wedge x_k$;
(iv) For each $p \in \bigcup_{i\in I} B_i$, if $p \neq 0$ then there is no subset Q of $\bigcup_{i\in I} B_i$ such that $p \leq \bigvee(Q - \{p\})$.
Then for each $X \in \mathscr{X}_1$, there exists a minimal element X_* of $\mathscr{X}_1$ such that $X_* \leq X$.

Notice that Theorem 3.8 is a special case of Theorem 3.13. In the following, let $G(b) = \{j \in J : a_j \geq b\}$. Then

Theorem 3.14 *Let $\mathscr{X}_2 \neq \emptyset$. Then there exists a minimal element in $\mathscr{X}_2$ iff either $G(b) \neq \emptyset$ or there is a subset B of L such that:*
(i) $b = \bigvee B$;
(ii) For every $p \in B$, there exists an index $j \in J$ such that $p \leq a_j$;
(iii) For every $p \in B, q \in L$, if $q < p$ then $b \neq [\bigvee(B - \{p\})] \bigvee q$.

Proof. Let $X_* = (x_{*j})_{j\in J}^T$ be minimal in $\mathscr{X}_2$. If $G(b) = \emptyset$ then from Theorem 3.7 $b = \bigvee_{j\in J} x_{*j}$, put $B = \{x_{*j} : x_{*j} \neq 0, j \in J\}$, then B satisfies the conditions (i), (ii) and (iii).

Conversely, if $G(b) \neq \emptyset$ then it is easy to see that there exists a minimal element in $\mathscr{X}_2$. Now suppose that there exists a subset B of L satisfying (i), (ii) and (iii), then we can construct a family of subsets $A_j, j \in J$ of B such that $\bigcup_{j\in J} A_j = B$ and $A_i \bigcap A_k = \emptyset$ when $i \neq k$ and $i, k \in J$. Define $X = (x_j)_{j\in J}^T$ with

$$x_j = \begin{cases} \bigvee_{p\in A_j} p, & A_j \neq \emptyset, \\ 0, & A_j = \emptyset. \end{cases}$$

Then $\bigvee_{j\in J}(a_j \wedge x_j) = \bigvee_{j\in J, A_j\neq\emptyset} x_j = \bigvee(\bigcup_{j\in J} A_j) = \bigvee B = b$, i.e. $X \in \mathscr{X}_2$. Let $Y = (y_j)_{j\in J}^T \in \mathscr{X}_2$ be such that $Y \leq X$. If $x_j = 0$ then $y_j = 0$. If $x_j \neq 0$ then $A_j \neq \emptyset$, therefore $y_j \leq x_j = \bigvee_{p\in A_j} p \leq a_j$ and $b = \bigvee_{j\in J}(a_j \wedge y_j) = \bigvee_{j\in J} y_j$. One can verify that for any $j \in J$ if $A_j \neq \emptyset$ then $y_j = x_j$. Indeed, if there exists an index $j_0 \in J$ such that $A_{j_0} \neq \emptyset$ but $y_{j_0} < x_{j_0}$, put $\overline{Y} = (\overline{y}_j)_{j\in J}$ with

$$\overline{y}_j = \begin{cases} y_{j_0}, & j = j_0, \\ x_j, & j \neq j_0. \end{cases}$$

Then $b = \bigvee_{j\in J} y_j \leq \bigvee_{j\in J} \overline{y}_j \leq \bigvee_{j\in J} x_j = b$ which means $(\bigvee_{j\in J, j\neq j_0} x_j) \bigvee y_{j_0} = b$, contradicts to (iii). Thus $Y = X$, i.e. X is minimal in $\mathscr{X}_2$. □

Note that Theorem 3.14 partly answers the open problem for existence of minimal solutions (see Page 46 in [8]).

It should be pointed out that most of the results in last two sections, such as Theorems 2.1, 2.4, 2.5, 3.1, 3.2 and 3.3, are also hold when L is distributive in finite case.

4 Conclusions

In this paper we gave an overview of the known results for fuzzy relation equations over complete Brouwerian lattices. In our opinions, the resolution of fuzzy relation equations over complete Brouwerian lattices can be solved from the viewpoint of decomposition.

References

[1] G. Birkhoff, Lattice theory. Vol. XXV, 3rd Ed., American Mathematical Society Colloquium Publications, Providence, RI, 1979.

[2] M. M. Bourke, D. G. Fisher, Solution algorithms for fuzzy relational equations with max-product composition, Fuzzy Sets and Systems 94 (1998) 61-69.

[3] K. Cechlarova, Unique solvability of max-min fuzzy equations and strong regularity of matrices over fuzzy algebra, Fuzzy Sets and Systems 75 (1995) 165-177.

[4] P. Crawley, R. P. Dilworth, Algebraic theory of lattices, Prentice-Hall, Englewood Cliffs, NJ, 1973.

[5] B. De Baets, Sup-T equations: state of the art, in: O. Kaynak, L. Zadeh, B. Turksen, I. Rudas (Eds.), Computational Intelligence: Soft Computing and Fuzzy-Neural Integration with Applications, NATO ASI Series F, Vol. 162, Springer, Berlin, 1998, pp. 80-93.

[6] B. De Baets, Analytical solution methods for fuzzy relational equations, in: D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, The Handbooks of Fuzzy Sets Series, Vol. 1, Kluwer Academic Publishers, Dordrecht, 2000, pp. 291-340.

[7] A. Di Nola, S. Sessa, Finite fuzzy relation equations with a unique solution in linear lattices, J. Math. Anal. Appl 132 (1988) 39-49.

ISBN: 978-989-95079-6-8

[8] A. Di Nola, S. Sessa, W. Pedrycz, E. Sanchez, Fuzzy relation equations and their applications to knowledge engineering. Kluwer Academic Publishers, Dordrecht, 1989.

[9] A. Di Nola, On solving relational equations in Brouwerian lattices, Fuzzy Sets and Systems 34 (1990) 365-376.

[10] A. Di Nola, A. Lettieri, I. Perfilieva, V. Novák, Algebraic analysis of fuzzy systems, Fuzzy Sets and Systems 158 (2007) 1-22.

[11] S. C. Han, H. X. Li, J. Y. Wang, Resolution of finite fuzzy relation equations based on strong pseudo-t-norms, Appl. Math. Lett. 19 (2006) 752-757.

[12] S. C. Han, H. X. Li, J. Y. Wang, Resolution of matrix equations over arbitrary Brouwerian lattices, Fuzzy Sets and Systems 159 (2008) 40-46.

[13] M. Higashi, G. J. Klir, Resolution of finite fuzzy relation equations, Fuzzy Sets and Systems 13 (1984) 65-82.

[14] H. Imai, M. Miyakoshi and T. Da-te, Some properties of minimal solutions for a fuzzy relation equation, Fuzzy Sets and Systems 90 (1997) 335-340.

[15] H. Imai, K. Kikuchi and M. Miyakoshi, Unattainable solutions of a fuzzy relation equation, Fuzzy Sets and Systems 99 (1998) 193-196.

[16] L. Nosková, I. Perfilieva, System of fuzzy relation equations with sup-* composition in semi-linear spaces: minimal solutions. In Proc. FUZZ-IEEE Conf. on Fuzzy Systems, July 23-36, 2007, London, pp. 1520-1525.

[17] W. Pedrycz, s-t Fuzzy relational equations, Fuzzy Sets and Systems 59 (1993) 189-195.

[18] K. Peeva, Y. Kyosev, Fuzzy relational calculus-theory, applications and software (with CD-ROM), in Series Advances in Fuzzy Systems - Applications and Theory, Vol. 22, World Scientific Publishing Company, 2004.

[19] K. Peeva, Y. Kyosev, Algorithm for solving max-product fuzzy relational equations, Soft Comput 11 (2007) 593-605.

[20] I. Perfilieva, Fuzzy function as an approximate solution to a system of fuzzy relation equations, Fuzzy Sets and Systems 147 (2004) 363-383.

[21] I. Perfilieva, Semi-linear spaces, in: H. Noguchi, H. Ishii, et al.(Eds.), Proc. of Seventh Czech-Japanese Seminar on Data Analysis and Decision Making under Uncertainty, Hyogo, Japan, 2004, pp. 127-130.

[22] I. Perfilieva, Semilinear spaces-basic structures for fuzzy systems, in: Proc. Conf. of IPMU'2006, Paris, France, 2006, pp. 1087-1094.

[23] I. Perfilieva, V. Novák, System of fuzzy relation equations as a continuous model of IF-THEN rules, Information Sciences 177 (2007) 3218-3227.

[24] E. Sanchez, Resolution of composite fuzzy relation equations, Inform. and Control 30 (1976) 38-48.

[25] S. Sessa, Finite fuzzy relation equations with unique solution in complete Brouwerian lattices, Fuzzy Sets and Systems 29 (1989) 103-113.

[26] M. Wagenknecht, K. Hartmann, On the existence of minimal solutions for fuzzy equations with tolerances, Fuzzy Sets and Systems 34 (1990) 237-244.

[27] X. P. Wang, Method of solution to fuzzy relation equations in a complete Brouwerian lattice, Fuzzy Sets and Systems 120 (2001) 409-414.

[28] X. P. Wang, Infinite fuzzy relational equations in a complete Brouwerian lattice, Indian J. Pure Appl. Math. 33 (2002) 87-95.

[29] X. P. Wang, Conditions under which a fuzzy relational equation has minimal solutions in a complete Brouwerian lattice, Advances In Mathematics 31 (2002) 220-228 (in Chinese).

[30] X. P. Wang, Infinite fuzzy relational equations on a complete Brouwerian lattice, Fuzzy Sets and Systems 138 (2003) 657-666.

[31] X. P. Wang, Q. Q. Xiong, The solution set of a fuzzy relational equation with sup-conjunctor composition in a complete lattice, Fuzzy Sets and Systems 153 (2005) 249-260.

[32] X. P. Wang, X. B. Qu, Continuous join-irreducible elements and their applications to describing the solution set of fuzzy relational equations, Acta Mathematica Sinica, Chinese Series 49 (2006) 1171-1180 (in Chinese).

[33] X. B. Qu, X. P. Wang, Some properties of infinite fuzzy relational equations on complete Brouwerian lattices, Fuzzy Sets and Systems 158 (2007) 1327-1339.

[34] Z. D. Wang, Y. D. Yu, Pseudo-t-norms and implication operators on a complete Brouwerian lattice, Fuzzy Sets and Systems 132 (2002) 113-124.

[35] K. L. Zhang, matrix equations in distributive lattices, Acta Scientiarum Naturalium Universitatis Intramongolicas 22 (1991) 311-317.

[36] K. L. Zhang, The inverse problem for distributive lattices, Fuzzy Sets and Systems 59 (1993) 109-113.

[37] K. L. Zhang, D. H. Li, L. X. Song, On finite relation equations with sup-conjunctor composition over a complete lattice, Fuzzy Sets and Systems 160 (2009) 119-128.

[38] C. K. Zhao, On matrix equations in a class of complete and completely distributive lattices, Fuzzy Sets and Systems 22 (1987) 303-320.

ISBN: 978-989-95079-6-8

Aggregation operators for conditional crispness

Doretta Vivona[1] Maria Divari[2]

1.2.Faculty of Engineering, "Sapienza" University of Rome 1
Rome , Italy
Email: vivona@dmmm.uniroma1.it, maria.divari@alice.it,

Abstract— *In axiomatic way, in fuzzy setting, we introduce the measure of crispness and the measure of conditional clearness. For these measures we propose a functional form. Then, for the last measure, we give some aggregation operators.*

Keywords— Aggregation operators, functional equations, fuzzy sets, measures of fuzziness, crispness measures, conditional crispness.

1 Introduction

De Luca and Termini gave the first definition of fuzziness measure [8, 15], based upon a condition of monotonicity, with respect to a antural relation of less or grater fuzziness. In 1975 Mukaidono presented another order, called monotonicity with respect to ambiguity [11]. Later Couso and Gil have modified the definition [7] and Roventa used the cuts of fuzzy sets in order to define this kind of entropy [12, 13, 14]. In [2] the authors, taking into account some consideration due to Wang [18] introduced small variations, leanding to a more exact order relation among fuzzy sets, which represents the idea of less or greater fuzziness.

We know that in $\mathcal{F}$ there are two different partial order: the classical inclusion relation [19] and the so called *fuzziness relation* according to the greater or smaller "proximity" to a crisp set; with respect to the last relation it is possible to define the infimum and the supremum [16]. In [2, 3, 4] we have studied the fuzziness measure and in [16] there are many examples of fuzziness measure also by using Sugeno's integral.

These researches have suggested to introduced, in axiomatic way, the concept of crispness c through a fuzziness measure d.

Then, taking into account the crispness, we give the definition of conditional crispness.

We know that an aggregation operator is a procedure by which a unique value can be associated to the result obtained through different tests or different values of data base. The unique value is kind of aritmetic mean.

In this setting, in this paper, we give the definition of measure of crispness (shortly crispness) through the fuzziness measure and the definition of conditional crispness.

Then, we propose a characterization of classes of aggregation operators for the conditional crispness.

2 Preliminaires

Given an abstract space X and a σ-algebra $\mathcal{C}$ of crisp subsets C of X, let $\mathcal{F}$ be the family of all $\mathcal{C}$-measurable fuzzy sets [18, 19]. With every fuzzy set F a membership function f is associated, $f \to [0,1]$, which is $\mathcal{C}$-measurable.

Now, we recall the order relation called *fuzziness relation* [2, 3]: F_1 is *less fuzzy* then F_2 ($F_1 \preceq F_2$) if the following condition holds $\forall\, x \in X$:

$$\begin{cases} f_1(x) \le \frac{1}{2} \implies f_1(x) \le f_2(x) \le \frac{1}{2} \\ f_1(x) \ge \frac{1}{2} \implies \frac{1}{2} \le f_2(x) \le f_1(x)\,, \end{cases}$$

which is equivalent to

$$\begin{cases} f_1(x) < \frac{1}{2} \implies f_1(x) \le f_2(x) \\ f_1(x) > \frac{1}{2} \implies f_1(x) \ge f_2(x)\,. \end{cases}$$

We remember that the fuzziness relation is different and not comparable with the inclusion relation:
$F_1 \subset F_2 \iff f_1(x) \le f_2(x)\ \forall\, x \in X$.
For crisp sets the relation $C_1 \preceq C_2$ implies $C_1 = C_2$.

We know [2] that $\left[\frac{X}{2}\right]$ and $\left[\frac{C}{2}\right]$ are those fuzzy sets, whose membership function are $f(x) = \frac{1}{2}, \forall x \in X$, and $f(x) = \frac{1}{2}, \forall x \in C$, respectively. Moreover we shall indicate with F^c the complement of F.

Definition 1 *The fuzziness measure is a map $d : \mathcal{F} \to [0,1]$ such that:*

$[A_1]\ d(C) = 0\ ,\ \forall\, C \in \mathcal{C}$;

$[A_2]\ F_1 \preceq F_2 \implies d(F_1) \le d(F_2)\ \forall F_1, F_2 \in \mathcal{F}$;

$[A_3]\ d\left(\left[\frac{X}{2}\right]\right) = 1$;

$[A_4]\ d(F) = d(F^c)\ \forall F \in \mathcal{F}$.

Moreover, we have recognized in [2] that to every fuzziness measure d we can associate a fuzzy measure μ on $\mathcal{C}$. In fact, the function $\mu : \mathcal{C} \to [0,1]$, defined by $\mu(C) = d\left(\left[\frac{C}{2}\right]\right)$, is non-decreasing with respect to the inclusion relation and $\mu(\emptyset) = 0$. This measure μ is called by us *fuzzy measure associated to d*. In [2] it is showed a form of this measure by using the Sugeno integral.

ISBN: 978-989-95079-6-8

3 Crispness and conditional crispness for fuzzy sets

In this paragraph we introduce the definition of crispness by using the fuzziness measure.

Definition 2 *Given a fuzziness measure* d, *the* crispness *of a fuzzy set* F *is a map* $c : \mathcal{F} \to [0,1]$ *defined by*

$$c(F) = 1 - d(F)\ . \tag{1}$$

Using the properties above $[A_1] - [A_4]$ the crispness (1) enjoys the following:

$[A_1^{'}]\ c(C) = 1\ ,\ \forall\ C \in \mathcal{C},\ \text{ as d(C)} = 0;$

$[A_2^{'}]\ F_1 \preceq F_2 \implies c(F_1) \geq c(F_2),$
as $d(F_1) \leq d(F_2)\ \forall F_1, F_2 \in \mathcal{F}$;

$[A_3^{'}]\ c\left(\left[\frac{X}{2}\right]\right) = 0,\ \text{ as } d\left(\left[\frac{X}{2}\right]\right) = 1;$

$[A_4^{'}]\ c(F) = c(F^c)\ \forall F \in \mathcal{F},\ \text{ as d(F)} = \text{d}(\text{F}^{\text{c}})\ .$

As an extension of crisp setting [9], we assume that two fuzzy sets F and H are *algebrically independent* if $F \cap H \neq \emptyset$.

Definition 3 *Fixed a fuzzy set* $H \in \mathcal{F}$ *and a crispness (1), the* conditional crispness *of any variable set* $F \in \mathcal{F}$ *is a map*

$$c_H(\cdot) : \mathcal{F} \to [0,1]$$

defined by the following axioms:

$(i)\ \ F_1 \preceq F_2 \implies c_H(F_1) \geq c_H(F_2), \forall F_1, F_2 \in \mathcal{F};$

$(ii)\ c_H\left(\left[\frac{X}{2}\right]\right) = 0,\ \left[\frac{X}{2}\right] \in \mathcal{C},$

$(iii)\ c_H(F) = c(F), \forall F \in \mathcal{F},$
if F is c-independent by H.

We assume the (iii) as the c-independence i.e.independence with respect to c.

As conseguence, every crisp set C is c-independent from H:
$c_H(C) = c(H),\ \forall C \in \mathcal{C}$.

4 Example

Let F be a fuzzy set composed by a group of old men. Given a fuzziness measure d, it is possible to calculate $d(F)$ and $c(F)$ according with (1). Moreover, we consider another fuzzy set H, whose elements are sick men. For us, the meaning of $c_H(F)$ is that the crispness of F influenced by the sickness of the men in H.

5 A form of measure of conditional crispness

Now, we propose the following form for the conditional crispness measure:

$$c_H(F) = \Phi\Big(c(F), c(H)\Big) \tag{2}$$

where $\Phi : [0,1] \times [0,1] \to [0,1]$. From $[(i), (ii), (iii)]$, we obtain:

$(a) F_1 \preceq F_2 \implies \Phi\Big(c(F_1), c(H)\Big) \geq \Phi\Big(c(F_2), c(H)\Big),$
$(b)\ \Phi(0, c(H)) = 0,$
$(c)\ \Phi(1, c(H)) = c(H).$

If we put $x = c(F_1), x' = c(F_2), y = c(H)$, with $, x, x', y \in [0,1]$, we get:

$$\begin{cases} (a')\ \Phi(1,y) = y \\ \\ (b')\ \Phi(0,y) = 0 \\ \\ (c')\ \Phi(x,y) \geq \Phi(x',y) \text{ if } x \geq x'\ . \end{cases}$$

For the system $[(a') - (c')]$ we have the following:

Proposition 1 *A class of solution of the system* $[(a') - (c')]$ *is*

$$\Phi = h^{-1}\Big(h(x) \cdot h(y)\Big) \tag{3}$$

where $h : [0,1] \times [0,1] \to [0,1]$ *is increasing with* $h(0) = 0$ *and* $h(1) = 1$.

Proof. The proof is immediate.

6 Example

If $h(x) = x^n, n \in (0, +\infty)$, then $\Phi(x,y) = x \cdot y$.

7 Aggregation operators for conditional crispness measures

In [17] we have studied the aggregation operators of the general conditional information. We shall use again the same procedure for the characterization of some forms of aggregation operators of the conditional crispness.

Let $\mathcal{I}$ be the family of the conditional crispness measures $c_H(\cdot)$. The aggregation operator $L : \mathcal{I} \to [0, \text{K}]$, $0 < \text{K} < [0, +\infty]$ of $n \in [0, +\infty)$ conditional crispness measures $c_H(F_1), ..., c_H(F_i), ..., c_H(F_n)$, with $F_i \in \mathcal{F}, \text{i} = 1, ..., \text{n}, \text{H} \in \mathcal{F}$ has the following properies as in [10]:

$(I)\ idempotence : c_H(F_i) = \lambda, \forall i = 1, ..., n \implies$

$$L(\underbrace{\lambda, ..., \lambda}_{n\ times}) = \lambda;$$

$(II)\ monotonicity : c_H(F_1) \leq c_H(F_1^{'}) \implies$
$L(c_H(F_1), ..., c_H(F_i), ..., c_H(F_n)) \leq$
$L(c_H(F_1^{'}), ..., c_H(F_i), ..., c_H(F_n)),$
$F_1^{'}, F_i \in \mathcal{F}, \text{i} = 1, ..., \text{n};$

$(III)\ continuity\ from\ below : c_H(F_{1,m}) \nearrow c_H(F_1) \implies$

ISBN: 978-989-95079-6-8

$c_H(F_{1,m}), ..., c_H(F_i), ..., c_H(F_n)) \quad \nearrow$
$c_H(F_1), ..., c_H(F_i), ..., c_H(F_n))$ $F_1^{'}, F_i \in \mathcal{F}, \mathrm{i} = 1, ..., \mathrm{n}.$

Putting $c_H(F_i) = x_i, i = 1, ...n, c_H(F_1^{'}) = x_1^{'}, c_H(F_{1,m}) = x_{1,m}$, with $x_i, i = 1, ...n, x_1^{'}, x_{1,m} \in [0,1]$, we obtain the following system of functional equations:

$$\begin{cases} (I')\ L(\underbrace{\lambda, ..., \lambda}_{n\ times}) = \lambda, \\ (II')\ x_1 \leq x_1^{'} \Longrightarrow L(x_1, ..., x_n) \leq L(x_1^{'}, ...x_n), \\ (III')\ x_{1,m} \nearrow x_1 \Longrightarrow L(x_{1,m}, ..., x_n) \nearrow L(x_1, ...x_n). \end{cases}$$

For the solution of the system $[(I') - (III')]$, we propose the following:

Proposition 2 *Two natural solutions of the system* $[(I') - (III')]$ *are*

$$L(x_1, ..., x_n) = \bigwedge_{i=1}^{n} x_i$$

and

$$L(x_1, ..., x_n) = \bigvee_{i=1}^{n} x_i.$$

Proof. The proof is immediate.

Proposition 3 *A class of solution of the system* $[(I') - (III')]$ *is*

$$L(x_1, ..., x_n) = h^{-1}\left(\frac{h(x_1) + ... + h(x_n)}{n}\right),$$

where $h : [0,1] \to [0,K](0 < K < +\infty)$ *is a continuous, strictly increasing function with* $h(0) = 0$ *and* $h(1) = K$.

Proof. The proof is immediate.

8 Remark

If the function h is linear, then the aggregation operator L is the aritmetic mean.

9 Conclusion

First, we have given the definitions of the crispness and the conditional crispness for a fuzzy set.

Second, we have proposed some classes of aggregation operators of the conditional crispness, solving a suitable system of functional equations.

References

[1] J. ACZEL: *Lectures on functional equations and their applications*, (1966), New Jork, Academic Press .

[2] P.BENVENUTI - D.VIVONA - M.DIVARI: On fuzziness measures via Sugeno integral, *Fuzzy logic and Soft Computing, Advances in Fuzzy Systems* **v.IV** (1995), 330-336.

[3] P.BENVENUTI - D.VIVONA - M.DIVARI: Equiordered Fuzzy Sets and Fuzziness Measure, *Proc.VII IFSA Congress,* (1997), pp.409-412.

[4] P.BENVENUTI - D.VIVONA: Order relation for fuzzy sets and entropy measures, *New Trend in Fuzzy Systems,* Ed.D.Mancini, M.Squillante and A.Ventre, World Scientific, (1998), pp.224-232.

[5] P.BENVENUTI - D.VIVONA - M.DIVARI: Aggregation operators and associated fuzzy mesures, *Int.Journ.of Uncertainty, Fuzziness and Knownledge-Based Systems,* **2**, (2001), 197-204.

[6] T.CALVO - G.MAYOR - R.MESIAR: *Aggregation operators,* New Trend and Applications, Physisc-Verlag, Heidelberg, (2002).

[7] I COUSO - P.GIL: Measures of fuzziness of type 2 fuzzy sets. *Proc.IPMU 96,* **2**, (1996), pp.581-584.

[8] A.DE LUCA - S.TERMINI: A definition of non probabilistic entropy in the setting of fuzzy sets theory. *Infor.and Contr.,* **20**, (1972), pp.301-312.

[9] B.FORTE: Measure of information. The general axiomatic theory. *R.I.R.O.* (1969), pp.63-99.

[10] E.P.KLEMENT: Binary aggregation operators which are bounded by the minimum (a mathematical overview), *Proc.AGOP05,* (2005), 13-16.

[11] M.MUKAIDONO: On some properties of fuzzy logic. *Syst.-Comput.-Control,* **6**, n.2, (1975), pp.36-43.

[12] E.ROVENTA: On the Degree of Fuzziness of a Fuzzy Sets, *FSS,* **36** (1990) pp.259-264.

[13] E.ROVENTA: On a Class of Measures of Fuzziness, *J.Math.Anal.Appl.,* **175** (1993) pp.337-341.

[14] E.ROVENTA - D.VIVONA: Representation Theorem for transom based measures of fuzziness, *Rend.Mat.Roma ,* **16** *rm (1996)* .

[15] S.TERMINI: Remarks on the measures of fuzziness. *Scuola Estiva "Logica Borosa"* El Escorial, Madrid (1992).

[16] D.VIVONA: Mathematical Aspect of the Theory of Measures of Fuzziness. *Mathware and Soft Computing,* **3** (1996), pp.211-224.

[17] D.VIVONA - M.DIVARI: Aggregation operators for conditional information without probability, *Proc.IPMU08,* (2008), pp.258-260.

ISBN: 978-989-95079-6-8

[18] Z.WANG – ZI-XIAO: Fuzzy measures and measures of fuzziness. *J.Math.An.Appl.,* **104** (1984), 581-601.

[19] L.A.ZADEH: Fuzzy sets, *Inf. and Control,* **8** (1965), pp.338-353.

ISBN: 978-989-95079-6-8

A New Conditioning Rule, Its Generalization and Evidential Reasoning

Koichi YAMADA[1] Vilany KIMALA[2] Muneyuki UNEHARA[1]

1 Faculty of Engineering, Nagaoka University of Technology
2 Graduate School of Engineering, Nagaoka University of Technology
1603-1, Kami-tomioka, Nagaoka, Niigata, Japan, 940-2188
Email: yamada@kjs.nagaokaut.ac.jp, kimala@mis.nagaokaut.ac.jp, unehara@kjs.nagaokaut.ac.jp

Abstract *- In Evidence theory, several conditioning rules for updating belief have been proposed, including Dempster's rule of conditioning. The paper views the conditioning rules proposed so far and proposes a new rule of conditioning based on three requirements. Then, it generalizes the rule to be applied to the case where condition is given by an uncertain belief. The paper also discusses a few interpretations of an equation used for evidential reasoning, one of which is interpreted as conditioning with an uncertain condition.*

Keywords— Dempster-Shafer theory of evidence, evidence theory, conditioning rule, evidential reasoning, uncertainty.

1 Introduction

The paper discusses a new conditioning rule of Evidence theory that we proposed recently [1,2] and generalizes the rule to be applied to the case where condition is given by an uncertain belief. Then, it discusses a few interpretations of an equation used for evidential reasoning, one of which is interpreted as conditioning with an uncertain condition.

The first and well-known conditioning rule in the theory is Dempster's rule of conditioning [3]. However, it has been pointed out that the rule has a few problems, and several alternatives have been proposed so far [4-13]. A well-known problem is that a small change of prior belief, which is represented as basic belief assignment (*bba*) or mass function in the paper, may produce a large change of posterior belief [4]. Another is a practical problem that conditional belief is undefined, when a given condition *totally conflicts* with the prior belief in the sense of $Pl(B) = 0$, where B is the condition represented by a subset of a frame of discernment Θ and Pl is the plausibility measure equivalent to the prior belief.

Dempster's rule of *conditioning* is derived as a special case of Dempster's rule of *combination*. Suppose there are two distinct beliefs represented by *bba*'s m, m', and the total mass of m' is assigned to a set B, *i.e.* $m'(B) = 1$. In the case, the belief combined from m and m' is the same as the conditional belief $m(\bullet \mid B)$. This means that both information of the prior belief m and the condition B are treated equivalently or symmetrically.

The paper deals with the two pieces of information asymmetrically. In the basic conditioning rule, which is one with a sure condition without any uncertainty in the paper, it is assumed that the condition is a piece of information given *a posteriori* and always certain, while the prior belief is given *a priori* and may not be correct. When they *partially conflict* with each other (*i.e.* $X \cap B = \varnothing$ for a X, where X is a focal element of the prior belief m, and $'$ is the empty set), the proposed conditioning trusts in B rather than X, and reassigns the mass $m(X)$ to B. The simple idea produces a conditioning rule different from conventional ones proposed so far. Then, the paper expands the basic conditioning to the one with an uncertain condition, and discusses the relation between the conditioning rule and evidential reasoning.

The paper is composed as follows; section 2 reviews Dempster-Shafer theory of evidence and major conditioning rules proposed so far including those with an uncertain condition represented by a belief. Section 3 discusses requirements that the basic conditioning rule should satisfy, and proposes a new basic conditioning. Section 4 expands the conditioning rule and proposes another generalized rule by removing a requirement. Section 5 discusses interpretations of an equation for evidential reasoning, one of which is a conditioning rule with an uncertain condition.

2 Conditioning Rules in Evidence Theory

2.1 Dempster's Theory of Evidence [3]

Let Θ be a frame of discernment. A function $m : 2^{\Theta} \rightarrow [0,1]$ satisfying the following equations represents a *bba*;

$$m(\varnothing) = 0, \tag{1}$$

$$\sum_{A \in 2^{\Theta}} m(A) = 1, \tag{2}$$

where 2^{Θ} is the power set of Θ. $m(A)$ is a degree of belief that a variable whose domain is Θ has the value in A, but has no information about elements and subsets of A. Subsets A satisfying $m(A) > 0$ are called focal elements.

The belief could be represented in other forms, such as Belief and Plausibility measures defined below.

$$Bel(A) = \sum_{X \subseteq A} m(X) \ . \tag{3}$$

$$Pl(A) = \sum_{X \cap A \neq \varnothing} m(X) \ . \tag{4}$$

Both measures are functions from 2^{Θ} to [0,1], and satisfy $Bel(\varnothing) = Pl(\varnothing) = 0$ and $Bel(\Theta) = Pl(\Theta) = 1$. Functions $m(\bullet)$, $Bel(\bullet)$ and $Pl(\bullet)$ are all transformable one another, and represent the same belief state.

ISBN: 978-989-95079-6-8

When two distinct beliefs m_1 and m_2 are derived from two independent information sources respectively, they could be combined by the famous rule, Dempster's rule of combination;

$$m_D(A) = \frac{\sum_{A=X\cap Y} m_1(X) \bullet m_2(Y)}{1 - \sum_{X\cap Y=\varnothing} m_1(X) \bullet m_2(Y)}. \quad (5)$$

where $A \neq \varnothing$ and $m_D(\varnothing) = 0$. If the denominator is zero (*i.e.* m_1 and m_2 totally conflict with each other) the combination is not defined.

When $m_1(\bullet) = m(\bullet)$ and $m_2(B) = 1$, the combination gives the next equation called Dempster's rule of conditioning.

$$m_D(A \mid B) = \frac{\sum_{A=X\cap B} m(X)}{1 - \sum_{X\cap B=\varnothing} m(X)}, \quad (6)$$

where $A' \in$ and $m_D(' \mid B) = 0$. When $Pl(B) = 0$ or m and B totally conflict with each other, the denominator becomes zero and $m_D(A \mid B)$ is undefined.

2.2 Alternative Conditioning Rules

As discussed in Introduction, many alternative conditioning rules have been proposed so far. Major rules are shown below ($B \neq \varnothing$ is assumed).

(i) Focusing rule

The rule is proposed in [7, 8] separately and is called focusing rule in [10].

$$Bel_F(A \mid B) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\overline{A} \cap B)}. \quad (7a)$$

$$Pl_F(A \mid B) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\overline{A} \cap B)}. \quad (7b)$$

Representation with *bba* is studied in [13].

(ii) Zhang's rule

Dempster's rule reassigns all mass of $m(X)$ to $X \cap B$. Zhang asserts that it is "radical" and proposes the next rule [11].

$$m_Z(A \mid B) = k \sum_{X\cap B=A} \frac{|X \cap B|}{X} m(X). \quad (8)$$

where k is a constant for normalization.

(iii) Geometric rule [5]

This is a strong rule in the sense that all mass of focal elements that are not included in B is abandoned.

$$m_G(A \mid B) = \begin{cases} m(A) \Big/ \sum_{X\in B} m(X), & \text{if } A \in B, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

(iv) Planchet's Conditional rule [6]

This is a weak rule in the sense that only mass of focal elements that have no intersection with B is abandoned.

$$m_P(A \mid B) = \begin{cases} m(A) \Big/ \sum_{X\in B'} m(X), & \text{if } A \in B', \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

(v) Ito&Inagaki's rule

Ito and Inagaki proposed a rule assuming that the condition is just an assertion of an information source [12].

a) In the case of $Bel(\overline{B}) < 1$,

$$Bel_I(A \mid B) = \begin{cases} \{1 - k_B Bel(\overline{B})\}\{Bel(A \cup \overline{B}) - Bel(\overline{B})\}, & \text{if } A \neq \Theta, \\ 1, & \text{if } A = \Theta, \end{cases} \quad (11a)$$

where k_B is a constant that satisfies $0 \le k_B \le 1/(1 - Bel(\overline{B}))$.

b) In the case of $Bel(\overline{B}) = 1$,

$$Bel_I(A \mid B) = \begin{cases} 1, & \text{if } A = ', \\ 0, & \text{otherwise.} \end{cases} \quad (11b)$$

(vi) Rules with an uncertain condition

Ichihashi and Tanaka proposed three Jeffrey-like conditioning rules with an uncertain condition [14].

a) Plausible conditioning

$$m_{pl}(A \mid m_b) = \sum_{A=X\in Y} \frac{m(X) \bullet m_b(Y)}{1 - \sum_{X\in Y=\varnothing} m(X)}, \quad (12)$$

where $m_b(\bullet)$ is a *bba* on Θ and $m_{pl}(\varnothing \mid m_b) = 0$. It is assumed $\sum_{X\cap Y=\varnothing} m(X) < 1$ for any focal element Y of m_b.

b) Credible conditioning

$$m_{cr}(A \mid m_b) = \sum_{A\subseteq Y} \frac{m(A) \bullet m_b(Y)}{1 - \sum_{X\not\subset Y, X\neq Y} m(X)}, \quad (13)$$

where $\sum_{X\in Y, X'\ Y} m(X) < 1$ for any focal element Y of m_b. $X \subset Y$ means that X is a proper subset of Y in the paper.

c) Possible conditioning

$$m_{po}(A \mid m_b) = \sum_{A\cap Y\neq\varnothing} \frac{m(A) \bullet m_b(Y)}{1 - \sum_{X\cap Y=\varnothing} m(X)}, \quad (14)$$

where $\sum_{X\cap Y=\varnothing} m(X) < 1$ for any focal element Y of m_b.

When m is Bayesian, and focal elements of m_b partition Θ, all of m_{pl}, m_{cr} and m_{po} equal to Jeffrey's rule. When $m_b(B) = 1$ for a specific B, m_{pl}, m_{cr} and m_{po} equal to

ISBN: 978-989-95079-6-8

Dempster's rule, Geometric rule and Planchet's rule, respectively.

3 Basic Conditioning Rule

In the paper, the basic conditioning rule means the one with a sure condition without any uncertainty. It assumes the following:

The given condition is exactly true information given temporally after the prior belief. It puts a strict restriction on focal elements of the posterior belief.

In short, the condition B is an established fact, and the true value actually exists in B. Thus, we trust in B rather than the prior belief m, when they partially conflict with each other, *i.e.* $B \cap X = \varnothing$ for an X, where X is a focal element of m. Based on the above idea, the requirements that the posterior belief must satisfy are discussed below.

(1) Requirement-1: $m(A \mid B) = 0$ when $A \cap \overline{B} \neq \varnothing$, where $\overline{B}$ is the complement of B. Assumed the condition B is an exact fact, the requirement might be self-evident. When the requirement holds, the flowing equations also hold.

$$m(A \mid B) > 0 \Leftarrow A' B, \tag{15}$$

$$B' A \Leftarrow Bel(A \mid B) = 1, \tag{16}$$

$$Bel(A \mid B) = Bel(A \cap B \mid B), \tag{17}$$

where $'$ is implication.

(2) Requirement-2: $m(A \mid B)$ must be defined except the case where $B = '$. This requirement comes from the assertion by Shafer [15] and Smets [9] that Evidence theory must be interpreted as a theory of belief, not as a theory of frequency. If *bba*, *Bel* and *Pl* represent degrees of a human belief, they should not be recognized as completely and precisely correct scales. Even if B is not plausible, *i.e.* $Pl(B) = 0$, a fact B may happen in the actual world. If the posterior belief conditioned by B is undefined in those cases, it might be a useless theory in real applications.

(3) Requirement 3: $m(\bullet \mid \Theta) = m(\bullet)$. The condition Θ represents the true value is in $'$. The condition is clearly unnecessary, unless we deal with an open world problem like Transferable Belief Model [16], which is an expanded evidence theory in the sense that $m(') > 0$ is allowed.

No conditioning rule in the previous section satisfies all the above requirements. All rules satisfy the 3rd one. Dempster's, Focusing, Zhang's and geometric rules satisfy the 1st one. However, nothing satisfies the 2nd requirement [1,2].

Conditioning rules could be derived from various combination rules in the same way as how Dempster's conditioning rule is derived as shown in section 2.1. However, such rules derived from combination rules proposed in [20-23] do not satisfy all the requirements, either. They do not satisfy the 1st requirement [1,2].

In this section, a new conditioning rule, which satisfies all the requirements, is proposed. In the proposed rule, the mass of focal elements of the prior belief are reassigned to another subset in the posterior belief as follows:

(*a***)** When a focal element X of the prior belief and the condition B are consistent (*i.e.* $X \cap B \neq \varnothing$), the condition B imposes restriction on the focal element, the possible area where the true value may exist. That is, the mass $m(X)$ is re-assigned to $X' B$.

(*b***)** When X and B partially conflict with each other (*i.e.* $X \cap B = \varnothing$), the grounds for the mass given to X are lost. Thus, the mass is temporarily given to $'$ representing total ignorance. Then, the restriction by B is imposed. As a result, the mass $m(X)$ is re-assigned to $\Theta \cap B = B$.

From the above (*a*) and (*b*), we get the next equations.

$$m_K(A \mid B) = \begin{cases} \Leftarrow m(X), \text{ if } A' \mid \ , A \{ B, \\ A = X \Leftarrow B \\ \Leftarrow m(X) + \Leftarrow m(X), \text{ if } A = B, \\ A = B \lfloor X \qquad X \Leftarrow B = \mid \\ 0, \text{ otherwise,} \end{cases} \tag{18a}$$

where $B \neq \varnothing$. This equation could be written as

$$m_K(A \mid B) = \sum_{A = X \cap B} m(X) + \sum_{\substack{X \cap B = \varnothing \\ A = B}} m(X), \tag{18b}$$

where $A \neq \varnothing, B \neq \varnothing$ and $m_K(\varnothing \mid B) = 0$.

The idea in (*b*) that the mass is given to $'$ in case of partial conflict is the same as one used in Yager's combination rule [21]. However, the conditioning rule derived from Yager's rule is different from the above, and does not satisfy the 1st requirement [1,2].

It is proved easily that $m_K(A \mid B)$ is a *bba* and that it satisfies the three requirements. It is also clear that $m_K(A \mid B) = m_D(A \mid B)$ when m and B are consistent (= all focal elements of m and B are consistent).

Then, the next inequality could be proved [2].

$$Pl_K(A \mid B)' Pl_D(A \mid B)' Bel_D(A \mid B)' Bel_K(A \mid B), \tag{19}$$

where $Pl_K(A \mid B)$ and $Pl_D(A \mid B)$ ($Bel_K(A \mid B)$ and $Bel_D(A \mid B)$) are plausibility (belief) functions calculated using eq.s (18) and (6), respectively.

$m_K(A \mid B)$ is not a generalization of Bayesian conditioning. When m is Bayesian, $Pl(B) = P(B)$ holds. In the case of $Pl(B) = P(B) = 0$ where $B \neq \varnothing$, $P(A \mid B)$ is undefined, while $m_K(A \mid B)$ must be defined due to Requirement-2. In addition, $m_K(A \mid B)$ is not Bayesian in general, when $|B| > 1$.

Associativity does not hold in this conditioning, namely, $m_K(A \mid B \mid C) \neq m_K(A \mid C \mid B) \neq m_K(A \mid B \cap C)$ for $B \cap C \neq \varnothing$, where $m_K(A \mid B \mid C) = m_K^B(A \mid C)$ and $m_K^B(A) =$

ISBN: 978-989-95079-6-8

$m_K(A \mid B)$. Non-associativity is caused by asymmetric nature of the conditioning between the two pieces of information, the prior belief and the condition.

Suppose the case where $X \cap B \cap C = \emptyset$, $X \cap B \neq \emptyset$, $X \cap C \neq \emptyset$ and $B \in C'\in$, where X is a focal element of the prior belief m. When calculating $m_K(\bullet \mid B)$, $m(X)$ is reassigned to $X \cap B$. Then, when calculating $m_K(\bullet \mid B \mid C)$, the mass $m(X)$ given to X' B is reassigned again to C, because $X \cap B \cap C = \emptyset$. On the other hand, the same mass $m(X)$ is reassigned to B when calculating $m_K(\bullet \mid C \mid B)$, and to $B \cap C$ when calculating $m_K(\bullet \mid B \cap C)$.

4 Generalized Rule of Conditioning

4.1 Conditioning Rule with Unreliable Condition

The basic conditioning rule is generalized by moderating the 1st requirement. First, we introduce an index λ $(0 \le \lambda \le 1)$ to represent reliability of the condition. When $\lambda < 1$, the condition is not completely correct, therefore requirement-1 is abandoned.

The mass reassignment in this case is an expansion of the one in the basic conditional rule, and is conducted as described below.

(a') When $X \in B'\in$, reassign mass of $'m(X)$ to X' B, and $(1-\lambda)m(X)$ to $X \cap \Theta = X$.

(b') When $X \cap B = \emptyset$, reassign mass of $\lambda m(X)$ to $\in'$ $B = B$, and $(1-\lambda)m(X)$ to X' $\in = X$.

In (a'), the degree that B is trusted is λ. Thus, only $\lambda m(X)$ is reassigned to $X \cap B$. The rest $(1-\lambda)m(X)$ is reassigned to $X \cap \Theta = E$, with replacing B by Θ, because B cannot be trusted at the degree of $1-\lambda$.

In the case of (b'), $\lambda m(X)$ is reassigned to $\Theta \cap B = B$, because the mass $\lambda m(X)$ is temporarily given to total ignorance Θ due to $X \cap B = \emptyset$, then restriction B is imposed. The rest $(1'\in)m(X)$ is reassigned to $X \cap \Theta = X$, because B cannot be trusted at the degree of $1-\lambda$.

From the above re-assignment, we get the next equation.

$$m_\lambda(A \mid B) = \begin{cases} \sum_{A = X \cap B} \lambda m(X) + (1-\lambda)m(A), \\ \qquad\qquad \text{if } B \supset A \neq \emptyset, \\ \sum_{A = B \subseteq X} \lambda m(X) + (1-\lambda)m(B) \\ \qquad + \sum_{X \cap B = \emptyset} \lambda m(X), \ \text{if } A = B, \\ (1-\lambda)m(A), \ \text{if } A \not\subset B, A \neq B, A \neq \emptyset, \\ 0, \ \text{otherwise}, \end{cases} \tag{19}$$

where $B'\in$. It is proved easily that $m_\lambda(A \mid B)$ is a *bba*. In addition, it is also proved that $m_\lambda(A \mid B)$ satisfies both the requirements 2 and 3, and that the next equation holds [2].

$$m_\lambda(A \mid B) = \lambda m_K(A \mid B) + (1-\lambda)m(A). \tag{20}$$

The equation could be understood as discounting of condition in conditioning, while discounting of information source in combination rule was proposed in [3].

4.2 Conditioning Rule with Uncertain Condition

The basic conditional rule is also generalized so that an uncertain condition given by *bba* m_b would be accepted in the same way as plausible, credible and possible conditioning rules in Section 2.

The mass reassignment in the case is similar to (a'') and (b''). The difference is i) reliability λ of condition B is replaced by *bba* $m_b(Y)$ of a focal element Y of m_b, and that ii) re-assignment of unreliable mass corresponding to $(1-\lambda)m(X)$ is unnecessary. The mass reassignment is done as follows:

(a'') When a focal element X of the prior belief m and a focal element Y of the uncertain condition m_b are consistent (*i.e.* $X \cap Y \neq \emptyset$ for any X and Y), the condition Y imposes restriction on the focal element X with uncertainty $m_b(Y)$. As a result, mass $m(X) \bullet m_b(Y)$ is re-assigned to $X \cap Y$.

(b'') When X and Y partially conflict with each other (*i.e.* $X \cap Y = \emptyset$), the grounds for the mass given to X are lost. Thus, the mass is temporarily given to $'$ representing total ignorance. Then, the restriction by Y is imposed with uncertainty $m_b(Y)$. As a result, the mass $m(X) \bullet m_b(Y)$ is re-assigned to $\in'$ $Y = Y$.

From (a'') and (b''), we get the following conditioning rule, if $A \neq \emptyset$.

$$\begin{aligned} m_{KYU}(A \mid m_b) &= \sum_{A = X \cap Y} m(X) \bullet m_b(Y) + \sum_{\substack{X \cap Y = \emptyset \\ A = Y}} m(X) \bullet m_b(Y) \\ &= \sum_Y m_b(Y) \left(\sum_{A = X \cap Y} m(X) + \sum_{\substack{X \cap Y = \emptyset \\ A = Y}} m(X) \right) \\ &= \sum_{Y \subseteq \Theta} m_K(A \mid Y) \bullet m_b(Y). \end{aligned} \tag{21}$$

It is proved easily that $m_{KYU}(A \mid m_b)$ is a *bba* and that it satisfies both the requirements 2 and 3.

The above equation shows that posterior belief with uncertain condition can be obtained by weighted sum of the basic conditioning rule $m_K(A \mid B)$ with different conditions B, where weight is given by the mass of the conditional set B of the belief representing the uncertain condition.

The following is obtained easily from eq. (20) and (21).

$$\begin{aligned} m_{\in}(A \mid B) &= \in n_K(A \mid B) + (1'\in)m(A) \\ &= \in n_K(A \mid B) + (1'\in)m_K(A \mid \in) \\ &= m_{KYU}(A \mid m_{(B,\in)}), \end{aligned} \tag{22}$$

ISBN: 978-989-95079-6-8

where $m_{(B,')}(B) = '$ and $m_{(B,\lambda)}(\Theta) = 1 - \lambda$.

In addition, it is proved easily that $m_{KYU}(A \mid m_b) = m_{pl}(A \mid m_b)$, when m and m_b are consistent each other, (*i.e.* $X \cap Y \neq \varnothing$ for all combination of X and Y, where X is a focal element of m and Y is of m_b). However, $m_{KYU}(A \mid m_b)$ is not an extension of Jeffrey's rule unlike $m_{pl}(A \mid m_b)$, because $m_K(A \mid B)$ is not Bayesian in general, even if the prior belief m is Bayesian.

5 Interpretations of Evidential Reasoning

The section shows that the conditioning represented by eq. (21) could be used as an interpretation of evidential reasoning. First, we review a few conventional discussions about evidential reasoning.

5.1 Representation of Rules

There are two ways to represent rules in evidential reasoning: joint form and conditional form [17]. The former represents a rule by a joint belief function on a product $\Theta_y \times \Theta_x$, where Θ_y, Θ_x are frames of discernment for variables y, x in antecedent and consequent of rules, respectively. The latter represents a rule by a conditional belief such as $m(A \mid B)$, where $B \subseteq \Theta_y$, $A \subseteq \Theta_x$. Though papers adopting joint form seem published more than those with conditional form, conditional form is more natural and easier for users to understand rules than joint form [17,18]. In addition, conditional form needs fewer values as knowledge than joint form; number of values conditional form needs is $2^{|'_x|} \bullet 2^{|'_y|} = 2^{|'_x|+|'_y|}$, while joint form needs $2^{|\Theta_x|\bullet|\Theta_y|}$, where $|\bullet|$ represents cardinality [17, 18].

5.2 Conventional Interpretations with Conditional Form

We have two kinds of knowledge. One is a set of rules described as

$$if\ y \in B\ then\ x \in A\ with\ belief\ m^{(1)}(A \mid B),$$

and the other is a prior belief $m^{(2)}(B)$ about y. We know all values of $m^{(1)}(A \mid B)$ and $m^{(2)}(B)$ for all combinations of $B \in '_y$ and $A \in '_x$ except for $m^{(1)}(A \mid \varnothing)$.

There are a few theories for evidential reasoning with conditional form of rules. Liu, *et al*. proposes the next equation for evidential reasoning [19].

$$m(A) = \sum_{B \subseteq \Theta_y} m^{(1)}(A \mid B) \bullet m^{(2)}(B). \tag{23}$$

The theoretical support of the equation is Bayesian formula of probabilistic reasoning with cylindrical extension.

$$P(A_E) = \sum_{B_E \in 2^{\Theta_y} \times 2^{\Theta_x}} P^{(1)}(A_E \mid B_E) \bullet P^{(2)}(B_E), \tag{24}$$

where P, $P^{(1)}$, $P^{(2)}$ are probability measures on $2^{\Theta_y} \times 2^{\Theta_x}$, and $B_E = \{B\}'\ 2^{\in_x}$, $A_E = 2^{\Theta_y} \times \{A\}$. It is assumed that $P(A_E) = m(A)$, $P^{(1)}(A_E \mid B_E) = m^{(1)}(A \mid B)$ and $P^{(2)}(B_E) = m^{(2)}(B)$.

Smets proposed evidential reasoning for Transferable Belief Model [16] based on conjunctive rule of combination in [18]. Conjunctive combination is equivalent to Dempster's rule combination without normalization and is defined by

$$m_\cap(A) = \sum_{A = X \cap Y} m'(X) \bullet m''(Y), \tag{25}$$

where m' and m'' are distinct *beliefs* defined on the same frame of discernment Θ.

The equation is proved to be equivalent to the next equation in [5].

$$m_\cap(A) = \sum_{Y \subseteq \Theta} m'_\cap(A \mid Y) \bullet m''(Y), \tag{26}$$

where $m_\cap(A \mid B)$ is conjunctive rule of conditioning, which is equal to $m_D(A \mid B)$ without normalization.

What should be noted is that reasoning with eq. (26) deviates from the standard evidential reasoning, if m' and m'' partially conflict with each other (*i.e.* $X \cap Y = \varnothing$ for a pair of X and Y, which are focal elements of $m'(\bullet)$ and $m''(\bullet)$, respectively), because $m_\cap(\in) > 0$ is derived.

Fortunately, the deviation from the standard evidential reasoning never happens, if m', m'' are *bba*'s on $\Theta = \Theta_y \times \Theta_x$, and *A, X and Y* are cylindrical extensions of $A' \subseteq \Theta_x$, $X \in '_x$ and $Y' \subseteq \Theta_y$, respectively (*i.e.* $A = \in_y' A \in$, $X = \Theta_y \times X'$, $Y = Y' \times \Theta_x$), because $X \cap Y = Y' \times X' \neq \varnothing$ for any non-empty X' and Y'. Therefore, $m_\cap$ obtained from eq. (26) is surely a *bba* on $\Theta_y \times \Theta_x$. Then, we can derive the following reasoning equation for the standard evidence theory.

$$m(A') = \sum_{Y' \subseteq \Theta_y} m^{(1)}(A' \mid Y') \bullet m^{(2)}(Y'), \tag{27}$$

where it is assumed that $m^{(1)}(A' \mid Y') = m'_\cap(A \mid Y)$, $m^{(2)}(Y') = m''(Y)$, $m(A') = m_\cap(A)$.

The two reasoning equations (23) and (27) are actually the same. That is, the evidential reasoning can be interpreted from two viewpoints: one is from probabilistic reasoning, and the other is from conjunctive combination. Note that conjunctive combination and conjunctive conditioning could be replaced by Dempster's combination and Dempster's conditioning, respectively in this case, because m' and m'' defined on $\Theta_y \times \Theta_x$ are consistent and normalization of Dempster's rule is unnecessary.

5.3 New Interpretation as Uncertain Conditioning

ISBN: 978-989-95079-6-8

Let us recall the conditioning rule with an uncertain condition given by eq. (21). We re-write the equation here.

$$m_{KYU}(A \mid m_b) = \sum_{Y \subseteq \Theta} m_K(A \mid Y) \bullet m_b(Y).$$

The equation is apparently similar to eq. (26). However, it is different, because $m_{\cap}(\varnothing) > 0$ may occur in eq. (26), while $m_{KYU}(\varnothing) = 0$ always holds.

However, if $m_K(A \mid Y)$ and $m_b(Y)$ are *bba*'s on $\in = \in_y{}' \in_x$, and $A = \in_y{}' A\in$, $Y = Y' \times \Theta_x$, $A\in\in{}'_x$, $Y' \subseteq \Theta_y$ similar to the case of eq. (26), $m_K(A \mid Y) = m_D(A \mid Y) = m_{\cap}(A \mid Y)$ holds as discussed in Section 3. Thus, equations (21) and (26) are actually the same in this case.

As a result, the equation (23) or (27) of evidential reasoning could be interpreted from the viewpoint of conditioning given a belief including uncertainty (eq. (21)), as well as from the viewpoint of probabilistic reasoning (eq. 24)) and conjunctive combination (eq. (25)).

6 Conclusions

The paper discussed and proposed a new basic conditioning rule of Evidence theory and generalizes the rule to be applied to the case where condition is given by an uncertain belief. Then, it discusses a few interpretations of an equation used for evidential reasoning, one of which is interpretation by conditioning with an uncertain condition.

The paper first discussed three requirements that should be satisfied by a basic conditioning rule. Then, it proposed a new basic rule that satisfies all of them, however, none of major conventional rules does.

The basic conditioning rule was generalized by abandoning one of the three requirements in order to cope with the case where condition is given as an uncertain belief. The generalized rule is obtained by weighted sum of the basic posterior beliefs with different conditions, where weight is the mass of conditional set in the belief given as an uncertain condition.

Then the paper discussed theoretical backgrounds of an equation used for evidential reasoning, which could be supported by probabilistic reasoning or by conjunctive combination of evidence. The paper showed that it could also be supported by a viewpoint of a conditioning rule with an uncertain condition.

References

[1] V. Kimala, Y. Koichi: Another conditioning rule in Evidence Theory, *11th Inter. Conf. on Humans and Computers*, Nagaoka, 323-328 (2008)

[2] V. Kimala, Y. Koichi, M. Unehara: Requirements of conditioning rules and a new conditioning in Evidence theory (submitted, in Japanese)

[3] G.A.Shafer: *Mathematical Theory of Evidence*, Princeton Univ. Press (1976)

[4] L.A.Zadeh: Review of books, a mathematical theory of evidence, *AI Magazine,* **5**-3, 81-83 (1984)

[5] D.Dubois, H.prade: On the unicity of Dempster rule of combination, *Int. J. of Intelligent Systems* **1**, 133-142 (1986)

[6] B. Planchet: Credibility and conditioning, *J. of Theoretical Probability* **2**, 289-299 (1989)

[7] L. M. De Campos, M.T. Lamata, S. Moral: The concept of conditional fuzzy measure, *Int. J. of Intelligent Systems* **5**, 237-246 (1990)

[8] J. Y. Halpern, R. Fagin : A new approach to updating beliefs, in *Uncertainty in Artificial Intelligence* **6** (eds. P. P. Bonissone, M. Henrion, L. N. Kanal, J. F. Lemmer), 347-373 (1991)

[9] P. Smets: What is Dempster-Shafer's model?, in *Advances in the Dempster-Shafer Theory of Evidence* (eds. R.R.Yager, M. Fedrizzi and J. Kacprzyk), John Wiley & Sons, 5-34 (1994)

[10] D.Dubois, H.Prade: Focusing versus updating in belief function theory, the same book as [9], 71-95.

[11] L. Zhang: Representation, independence, and combination of evidence in the Dempster-Shafer theory, the same as [9], 51-61.

[12]M.Itoh, T.Inagaki: A new conditioning rule for Belief Updating in the Dempster-Shafer theory of evidence, *J. of the society of instrument and control engineers,* **31**-12, 2011-2017 (1995) (in Japanese)

[13] E.C.Kulasekere, *et al.* : Conditioning and updating evidence, *Int.J of Approximate Reasoning* **36**, 75-108 (2004)

[14] H. Ichihashi, H. Tanaka: Jeffrey-like rules of conditioning for the Dempster-Shafer theory of evidence, *Int. J. of Approximate Reasoning* **3**, 143-156 (1989)

[15] G. Shafer: Perspective on the theory and practice of belief function, *Int. J of Approximate Reasoning* **4**, 323-362 (1990)

[16] P. Smets, R. Kennes: The transferable belief model, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**-5, 447-458 (1990)

[17] B. B. Yaghlane, K. Mellouli: Inference in directed evidential networks based on the transferable belief model, *Int. J. of Approximate Reasoning* **48**, 399-418 (2008).

[18] P. Smets: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, *Int. J. of Approximate Reasoning* **9**, 1-35 (1993)

[19] W. Liu, J.G.Hughes, M.F.Mctear: Representing Heuristic knowledge and propagating beliefs in the Dempster-Shafer theory of evidence, in *the same book as [9],* 441-471.

[20] R.R.Yager: On the relationships of methods of aggregation of evidence in expert systems, *Cybernet. Systems* 16, 1/21 (1985)

[21] R.R. Yager: On the Dempster-Shafer framework and new combination rules, *Inform. Sci*. 41, 93/137 (1987)

[22] D.Dubois, H.Prade: Representation and combination of uncertainty with belief functions and possibility measures, *Comput. Intel.* 1(1),244/264(1988)

[23] H.Y. Hau, *et al*.: Belief combination and propagation in a lattice-structured inference network, *IEEE SMC* 20 (1), 45/58 (1990)

ISBN: 978-989-95079-6-8

Consistency of pair-wise comparison matrix with fuzzy elements

Jaroslav Ramík

School of Business Administration in Karviná, Silesian University in Opava
733 40 Karviná, Czech Republic

Email: ramik@opf.slu.cz

Abstract—*I*n this paper, consistency of pair-*w*ise comparison matri*x* *w*ith fuzzy elements based on the method of *g*eometric mean is investigated. A consistency inde*x* of reciprocal matri*x* *w*ith fuzzy elements is introduced based on a ne*w*ly designed method of lo*g*arithmic least s*q*uares for eliciting associated *w*eights. Some basic properties of the inde*x* are presented and t*w*o simple illustrating examples are supplied.

Key*w*ords—decision making, uncertainty, pair-*w*ise comparison, reciprocal matri*x*, consistency, triangular fuzzy numbers.

1 Introduction

The classical pair-*w*ise comparison method re*q*uires the decision-maker (DM) to e*x*press his/her preferences in the form of a precise ratio matri*x* encodi*n*g a valued preference relation. Ho*w*ever, it can often be difficult for the DM to e*x*press e*x*act estimates of the ratios of importance. Therefore many kinds of methods employing intervals or fuzzy numbers as elements of a pair-*w*ise comparison matri*x* have been proposed to cope *w*ith this problem. This allo*w*s for a more fle*x*ible specification of pair-*w*ise preference intensities accounti*n*g for the incomplete kno*w*ledge of the DM, see [4].

*I*n practice, *w*hen interval-valued matrices are employed, the DM often *g*ives ranges narro*w*er than his/her actual perception *w*ould authorize, because he/she mi*g*ht be afraid of e*x*pressi*n*g information *w*hich is too imprecise. On the other hand, a fuzzy number or fuzzy interval e*x*presses rich information because the DM can provide the core of the fuzzy interval as a rough estimate of his perceived preference and also the support set of the fuzzy interval as the ra*n*ge that the DM believes to surely contain the unkno*w*n ratio of relative importance.

*I*n this paper, consistency of pair-*w*ise comparison matri*x* *w*ith fuzzy numbers based on the method of *g*eometric mean is investigated. A consistency inde*x* of reciprocal matri*x* *w*ith triangular fuzzy elements is introduced based on a ne*w*ly desi*g*ned method of lo*g*arithmic least s*q*uares for eliciting associated *w*eights *w*ith the minimal measure of fuzziness. Here, *w*e accept the reasons discussed in [3] (and in other papers cited there) for using *g*eometric mean instead of Saaty's procedures based on the principal eigenvector of the pair-*w*ise comparison matri*x*. Our approach is ho*w*ever different to the method used by Van Laarhoven and Pedrycz, see [7]. Defini*n*g the consistency inde*x*, *w*e use the different approach to Buckley's at al. [2], the consistency inde*x* proposed there is based on the Saaty's definition. Moreover, the calculation procedure of the inde*x* is rather comple*x*.

2 Consistency of pair-wise comparison matrix

Consider an n×n pair-*w*ise comparison matri*x* **A** such that

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{bmatrix}. \qquad (1)$$

This matri*x* is reciprocal, if $a_{ij} = \frac{1}{a_{ji}}$ for each $1 \le i, j \le n$.

We say that **A** is consistent if

$$a_{ij}.a_{jk} = a_{ik}, \qquad (2)$$

for each $1 \le i, j, k \le n$.

*I*f for some integer i, *j*, k, (2) does not hold, than **A** is said to be inconsistent. *I*n the popular Analytic hierarchy process (AHP), it is assumed that $\frac{1}{9} \le a_{ij} \le 9$, for all $1 \le i, j \le n$, see [6]. The consistency of **A** is measured by the consistency inde*x* C*I* as

$$CI = \frac{\lambda_{max} - n}{n-1}, \qquad (3)$$

*w*here λ_{max} is the ma*x*imal eigenvalue of **A**. *I*t can be sho*w*n that C*I* ≥ 0, see [6].

*I*f **A** is an n×n reciprocal pair-*w*ise comparison matri*x*, then **A** is consistent iff C*I* = 0. To provide a measure independent of the order of the matri*x*, n, Saaty proposed the consistency ratio (CR). This is obtained by taki*n*g the ratio bet*w*een λ_{max} - n to its e*x*pected value over a large number of positive reciprocal matrices of order n, *w*hose entries are randomly chosen in the set of values $\left\{\frac{1}{9},...,9\right\}$. For this consistency measure, he proposed a 10% threshold for the CR to accept the estimation. *I*n practical decision situations inconsistency is "acceptable" if CR < 0,1. For the priorization procedure based on *g*eometric mean, the *g*eometric consistency ratio *G*CR *w*as proposed in [1], *w*ith an interpretation analogous to that considered for Saaty´s CR.

ISBN: 978-989-95079-6-8

3 Pair-wise comparison matrix with triangular fuzzy elements

Since using fuzzy numbers as elements of a pair-wise matrix is more expressive than using crisp values or intervals, we hope that the fuzzy approach allows a more accurate description of the decision making process. Rather than forcing the DM to provide precise representations of imprecise perceptions, we suggest using an imprecise representation instead.

We define a new consistency measure for an n×n reciprocal pair-wise comparison matrix with triangular fuzzy numbers being the elements of the matrix. The membership function of a triangular fuzzy number is piece-wise linear. There exist at least three reasons for using triangular shape of fuzzy elements:

1. The membership function of triangular fuzzy elements is piece-wise linear, i.e. relatively simple.
2. Triangular fuzzy numbers can be easily manipulated, e.g. added, multiplied etc.
3. Crisp (non-fuzzy) numbers, i.e. the most practical values, can be represented as triangular ones.

A triangular fuzzy number a can be equivalently expressed by a tripple of real numbers, i.e. $a = (a^L; a^M; a^U)$, where a^L is the Lower number, a^M is the Middle number, and a^U is the Upper number, $a^L \le a^M \le a^U$. If $a^L = a^M = a^U$, then a is said to be the crisp number (non-fuzzy number). Evidently, the set of all crisp numbers is isomorphic to the set of real numbers. In order to distinguish fuzzy and non-fuzzy numbers we shall denote fuzzy numbers, vectors and matrices by the tilde above the symbol, e.g. $\tilde{a} = (a^L; a^M; a^U)$. It is well known that the arithmetic operations +,-, * and / can be extended to fuzzy numbers by the Extension principle.

If all elements of an m×n matrix $\mathbf{A}$ are triangular fuzzy numbers we call $\mathbf{A}$ the matrix with triangular fuzzy elements and this matrix is composed of triples as follows

$$\tilde{\mathbf{A}} = \begin{bmatrix} (a_{11}^L; a_{11}^M; a_{11}^U) & \cdots & (a_{1n}^L; a_{1n}^M; a_{1n}^U) \\ \vdots & \ddots & \vdots \\ (a_{m1}^L; a_{m1}^M; a_{m1}^U) & \cdots & (a_{mn}^L; a_{mn}^M; a_{mn}^U) \end{bmatrix}.$$

Particularly, let $\tilde{\mathbf{A}}$ be an n×n matrix with triangular fuzzy elements. We say that $\tilde{\mathbf{A}}$ is reciprocal, if the following condition is satisfied: $\tilde{a}_{ij} = (a_{ij}^L; a_{ij}^M; a_{ij}^U)$ implies $\tilde{a}_{ji} = (\frac{1}{a_{ij}^U}; \frac{1}{a_{ij}^M}; \frac{1}{a_{ij}^L})$ for all i,j = 1,2,...,n, i.e.

$$\tilde{\mathbf{A}} = \begin{bmatrix} (1;1;1) & (a_{12}^L; a_{12}^M; a_{12}^U) & \cdots & (a_{1n}^L; a_{1n}^M; a_{1n}^U) \\ (\frac{1}{a_{12}^U}; \frac{1}{a_{12}^M}; \frac{1}{a_{12}^L}) & (1;1;1) & \cdots & (a_{2n}^L; a_{2n}^M; a_{2n}^U) \\ \vdots & \vdots & \ddots & \vdots \\ (\frac{1}{a_{1n}^U}; \frac{1}{a_{1n}^M}; \frac{1}{a_{1n}^L}) & (\frac{1}{a_{2n}^U}; \frac{1}{a_{2n}^M}; \frac{1}{a_{2n}^L}) & \cdots & (1;1;1) \end{bmatrix}, \quad (4)$$

where $1 \le a_{ij}^L \le a_{ij}^M \le a_{ij}^M$, i,j=1,2,...,n. Without loss of generality we assume that $1 \le a_{ij}^M \le a_{ik}^M$ whenever $i \le j \le k$. In some sense, our new inconsistency measure will become an extension of the consistency index CR. For this purpose let us consider n elements (e.g. some criteria or variants) being evaluated by the DM. The result of the pair-wise comparison evaluations of the DM is a reciprocal matrix with triangular fuzzy elements $\tilde{\mathbf{A}} = \{\tilde{a}_{ij}\}$, i.e. (4), where

$$\tilde{a}_{ji} = (\frac{1}{a_{ij}^U}; \frac{1}{a_{ij}^M}; \frac{1}{a_{ij}^L}) \quad \text{for all } i,j = 1,2,...,n.$$

Here, $\frac{1}{\sigma} \le a_{ij}^L \le a_{ij}^M \le a_{ij}^U \le \sigma$ and $S = [\frac{1}{\sigma}, \sigma]$, $\sigma > 1$, is an interval of real numbers called the scale. DMs use the scale for evaluating pairs of elements. In classical AHP, $\sigma = 9$, see [6].

Now, consider deviations

$$\frac{\tilde{w}_i}{\tilde{w}_j} \simeq \tilde{a}_{ij}, \quad (5)$$

where $\tilde{w}_k = (w_k^L; w_k^M; w_k^U)$ are triangular fuzzy weights associated with evaluated elements and extended arithmetic operations on fuzzy numbers, see [2]. The question is how these weights should be calculated so that the deviations are minimized in some (fuzzy) sense.

Here, we propose a new method for calculating w_k^L, w_k^M, w_k^U by solving the following optimization problem:

$$\sum_{i,j} \left(\log \frac{w_i^L}{w_j^L} - \log a_{ij}^L \right)^2 + \left(\log \frac{w_i^M}{w_j^M} - \log a_{ij}^M \right)^2 + \left(\log \frac{w_i^U}{w_j^U} - \log a_{ij}^U \right)^2 \longrightarrow \min, \quad (6)$$

subject to

$$w_k^U \ge w_k^M \ge w_k^L \ge 0, \; k = 1, 2,...,n. \quad (7)$$

By the well known method of making the derivatives zero (a necessary condition for optimality) it can be easily shown, that the solution of problem (3), (4) shall satisfy the following relations

$$w_k^L = C_L \cdot \left(\prod_{j=1}^n a_{kj}^L \right)^{1/n}, \; w_k^M = C_M \cdot \left(\prod_{j=1}^n a_{kj}^M \right)^{1/n}, \; w_k^U = C_U \cdot \left(\prod_{j=1}^n a_{kj}^U \right)^{1/n} \quad (8)$$

for each k = 1, 2,...,n, where coefficients C_L, C_M, C_U are arbitrary positive constants. We shall find the values of the coefficients with the "minimal measure of fuzziness" of the associated weights.

ISBN: 978-989-95079-6-8

Moreover, we require that the middle values w_k^M of the fuzzy weights $\tilde{w}_k = (w_k^L; w_k^M; w_k^U)$ satisfy the normalization condition, i.e. $\sum_{k=1}^{n} w_k^M = 1$. Hence, from normalization condition and from (7) we obtain

$$C_M = \frac{1}{\sum_{i=1}^{n} (\prod_j a_{ij}^M)^{1/n}}. \quad (9)$$

Hence, from (6), (7) we obtain

$$C_L \le C_M \min_{i=1,\dots,n} \left\{ \frac{\left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}{\left(\prod_{j=1}^{n} a_{ij}^L\right)^{1/n}} \right\},$$

$$C_U \ge C_M \max_{i=1,\dots,n} \left\{ \frac{\left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}{\left(\prod_{j=1}^{n} a_{ij}^U\right)^{1/n}} \right\}. \quad (10)$$

Moreover, we want to find weights $\tilde{w}_k = (w_k^L; w_k^M; w_k^U)$, k = 1, 2,...,n, with the minimal spread, i.e. the minimal measure of fuzziness $s_k = w_k^U - w_k^L$. Hence, by (8), (9) we obtain the weights as follows:

$$w_k^L = C_{min} \cdot \frac{\left(\prod_{j=1}^{n} a_{kj}^L\right)^{1/n}}{\sum_{i=1}^{n} \left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}, C_{min} = \min_{i=1,\dots,n} \left\{ \frac{\left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}{\left(\prod_{j=1}^{n} a_{ij}^L\right)^{1/n}} \right\}, \quad (11)$$

$$w_k^M = \frac{\left(\prod_{j=1}^{n} a_{kj}^M\right)^{1/n}}{\sum_{i=1}^{n} \left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}, \quad (12)$$

$$w_k^U = C_{max} \cdot \frac{\left(\prod_{j=1}^{n} a_{kj}^U\right)^{1/n}}{\sum_{i=1}^{n} \left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}, C_{max} = \max_{i=1,\dots,n} \left\{ \frac{\left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}{\left(\prod_{j=1}^{n} a_{ij}^U\right)^{1/n}} \right\}. \quad (13)$$

Particularly, if $\tilde{\mathbf{A}}$ is a crisp (i.e. nonfuzzy) matrix, i.e. $a_{ij}^L = a_{ij}^M = a_{ij}^U$ for all i, j, then $C_{min} = C_{max} = 1$, therefore, $w_k^L = w_k^M = w_k^U$ for all k. Consequently, the solution - weights are crisp, too.

4 New consistency index

Supposing the maximum of all deviations (5) for all i,j is equal to "zero", then intuitively $\tilde{\mathbf{A}}$ should be consistent. Hence, a consistency measure shall be introduced as the maximal deviation relative to the measure which could be achieved by a matrix of the same dimension.

For a given scale $S = [\frac{1}{\sigma}, \sigma]$ we define a consistency index $I_n^\sigma(\tilde{\mathbf{A}})$ of n×n reciprocal matrix $\tilde{\mathbf{A}}$ with fuzzy triangular elements as follows

$$I_n^\sigma(\tilde{\mathbf{A}}) = C_n^\sigma . \max_{i,j} \left\{ \max\{ \left| \frac{w_i^L}{w_j^U} - a_{ij}^L \right|, \left| \frac{w_i^M}{w_j^M} - a_{ij}^M \right|, \left| \frac{w_i^U}{w_j^L} - a_{ij}^U \right| \} \right\} \quad (14)$$

where w_k^L, w_k^M, w_k^U are given by (11) – (13) for all k= 1,2,…,n, and

$$C_n^\sigma = \frac{1}{\max\left\{ \sigma - \sigma^{\frac{2-2n}{n}}, \sigma^2 \left(\left(\frac{2}{n}\right)^{\frac{2}{n-2}} - \left(\frac{2}{n}\right)^{\frac{n}{n-2}} \right) \right\}}, \text{ if } \sigma < \left(\frac{n}{2}\right)^{\frac{n}{n-2}},$$

$$C_n^\sigma = \frac{1}{\max\left\{ \sigma - \sigma^{\frac{2-2n}{n}}, \sigma^{\frac{2n-2}{n}} - \sigma \right\}}, \text{ if } \sigma \ge \left(\frac{n}{2}\right)^{\frac{n}{n-2}}. \quad (15)$$

We say that $\tilde{\mathbf{A}}$ is F-consistent if $I_n^\sigma(\tilde{\mathbf{A}}) = 0$, otherwise, if $I_n^\sigma(\tilde{\mathbf{A}}) > 0$, then $\tilde{\mathbf{A}}$ is F-inconsistent.

The following theorem shows, that C_n^σ in (15) is a "normalizing" constant. The proof is easy, see [5].

Theorem 1. *If* $\tilde{\mathbf{A}}$ is an n×n reciprocal matrix with triangular fuzzy elements evaluated from the scale $[\frac{1}{\sigma}, \sigma]$, then

$$0 \le I_n^\sigma(\tilde{\mathbf{A}}) \le 1. \quad (16)$$

The following theorem states that the new concept of F-consistency is compatible with the old concept of consistency (2).

Theorem 2. Let $\tilde{\mathbf{A}} = \mathbf{A}$ be a non-fuzzy reciprocal pair-wise comparison matrix. *If* **A** is consistent then **A** is F-consistent.

5 Examples

In this section we present two examples showing that the newly defined consistency index is a convenient tool not only for measuring inconsistency of pair-wise comparison matrices with fuzzy elements, but also for measuring inconsistency of crisp pair-wise comparison matrices.

The first example gives the (crisp) pair-wise comparison matrix $\mathbf{A}^*$ with inconsistency index $I_n^\sigma(\mathbf{A}^*) = 1$.

Example 1. Consider 7×7 reciprocal matrix with crisp elements from the scale $[\frac{1}{9}, 9]$:

ISBN: 978-989-95079-6-8

$$\mathbf{A}^{*}=\begin{bmatrix} 1 & 9 & 9 & 9 & 9 & 9 & 9 \\ \frac{1}{9} & 1 & 9 & 9 & 9 & 9 & 9 \\ \frac{1}{9} & \frac{1}{9} & 1 & 9 & 9 & 9 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 & 9 & 9 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 & 9 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 \end{bmatrix}.$$

Here, as $9>\left(\frac{7}{2}\right)^{\frac{7}{7-2}}=5.78$, we obtain

$$\frac{w_1}{w_n}-a_{1n}=\frac{\left(\sigma^{n-1}\right)^{\frac{1}{n}}}{\left(\frac{1}{\sigma^{n-1}}\right)^{\frac{1}{n}}}-\sigma=\sigma^{\frac{2n-2}{n}}-\sigma=9^{\frac{12}{7}}-9=34.24.$$

$$a_{n1}-\frac{w_n}{w_1}=\sigma-\frac{\left(\frac{1}{\sigma^{n-1}}\right)^{\frac{1}{n}}}{\left(\sigma^{n-1}\right)^{\frac{1}{n}}}-=\sigma-\sigma^{\frac{2-2n}{n}}-=9-9^{-\frac{12}{7}}=8.98.$$

Hence,

$$\max_{i,j}\left\{\left|\frac{w_i}{w_j}-a_{ij}\right|\right\}=\max\{34.24,8.98\}=34.24.$$

By (14) and (15) we get $I_7^9(\mathbf{A}^{*})=1$.

Example 2. Consider 3×3 reciprocal matrix $\widetilde{\mathbf{A}}$ with triangular fuzzy elements

$$\widetilde{\mathbf{A}}=\begin{bmatrix} (1 & 1 & 1) & (2 & 3 & 4) & (4 & 5 & 6) \\ (\frac{1}{4} & \frac{1}{3} & \frac{1}{3}) & (1 & 1 & 1) & (3 & 4 & 5) \\ (\frac{1}{6} & \frac{1}{5} & \frac{1}{4}) & (\frac{1}{5} & \frac{1}{4} & \frac{1}{3}) & (1 & 1 & 1) \end{bmatrix},$$

S = [1/9, 9] is an evaluation scale.

As $9>\left(\frac{3}{2}\right)^{\frac{3}{3-2}}=3.375$, we calculate

$$C_3^9=\frac{1}{\max\left\{9-9^{-\frac{4}{3}},9^{\frac{4}{3}}-9\right\}}=\frac{1}{\max\{8.95,9.72\}}=0.103.$$

By (11) – (13), we obtain $C_{max}=0.855$, $C_{min}=1.145$ and

$$\widetilde{w}_1=(w_1^{L};w_1^{M};w_1^{U})=(0.458\,;\,0.627\,;\,0.627),$$
$$\widetilde{w}_2=(w_2^{L};w_2^{M};w_2^{U})=(0.264\,;\,0.280\,;\,0.295),$$
$$\widetilde{w}_3=(w_3^{L};w_3^{M};w_3^{U})=(0.094\,;\,0.094\,;\,0.094).$$

Finally, we calculate the inconsistency index $I_3^9(\widetilde{\mathbf{A}})=0.219$, hence $\widetilde{\mathbf{A}}$ is F-inconsistent.

For comparison with the well known AHP method, consider the corresponding crisp matrix $\mathbf{A}^{M}$ taking into account only the middle values of $\widetilde{\mathbf{A}}$, i.e.

$$\mathbf{A}^{M}=\begin{bmatrix} 1 & 3 & 5 \\ \frac{1}{3} & 1 & 4 \\ \frac{1}{5} & \frac{1}{4} & 1 \end{bmatrix}.$$

Consistency index $I_3^9(\mathbf{A}^{M})=0.174$, hence $\mathbf{A}^{M}$ is also F-inconsistent, the consistency index is, however, smaller than the previous one, as there is no fuzziness in $\mathbf{A}^{M}$. Applying consistency index by T. Saaty to matrix $\mathbf{A}^{M}$, we obtain CR = 0.082.

6 Conclusions

In this paper, we investigated consistency of pair-wise comparison matrix with triangular fuzzy elements based on a newly designed method of logarithmic least squares for eliciting associated weights. A new consistency index of reciprocal matrix with fuzzy triangular elements was defined. The weights we derived have the minimal measure of fuzziness among all weights based on the solution of a specific optimization problem. Some properties of the consistency index were derived and two illustrating examples were presented. The newly defined inconsistency index seems to be a convenient tool not only for measuring inconsistency of pair-wise comparison matrices with fuzzy elements, but also for measuring inconsistency of crisp pair-wise comparison matrices.

Acknowledgment

Supported by GACR project No. 402090405

References

[1] Aguarón, J., Moreno-Jimenéz, J. M., The geometric consistency index: Approximated thresholds. European Journal of Operational Research 147, 2003, pp. 137–145, ISSN 0377-2217.

[2] Buckley, J. J., Feuring, T., Hayashi, Y., Fuzzy hierarchical analysis revisited. Fuzzy Sets and Systems 129, 2001, pp. 48-64, ISSN 0165-0114.

[3] Lootsma, F. A., A model for the relative importance of the criteria in the multiplicative AHP and SMART. European Journal of Operational Research 94, 1996, pp. 467-476, ISSN 0377-2217.

[4] Ohnishi, S. et al., A study on the Analytic Hierarchy Process using a fuzzy reciprocal matrix. Proceeding of the conference IPMU 2006, irit.fr, http://www.irit.fr/~Didier.Dubois/Papers0804/ODPY_IPMU06.pdf.

[5] Ramík, J., Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean. To be published in Fuzzy Sets and Systems.

[6] Saaty, T. L., Multicriteria decision making - the Analytical Hierarchy Process. Vol. I., RWS Publications, Pittsburgh, 1991.

[7] Van Laarhoven, P. J. M. and Pedrycz, W., A fuzzy extension of Saaty's priority theory. Fuzzy Sets and Systems 11, 1983, 4, pp. 229-241, ISSN 0165-0114.

ISBN: 978-989-95079-6-8

Fuzzy Continuous Evaluation in Training Systems Based on Virtual Reality

Ronei M. Moraes[1] and Liliane S. Machado[2]

1 Department of Statistics, Federal University of Paraíba
João Pessoa/PB, Brazil
2 Department of Informatics, Federal University of Paraíba
João Pessoa/PB, Brazil
Email: ronei@de.ufpb.br, liliane@di.ufpb.br

Abstract*—The approach of continuous evaluation is an important methodology in educational learning process. However, only recently it was applied in training based on virtual reality. This paper presents a methodology of evaluation that uses a fuzzy continuous evaluation approach to provide a user profile from his several training. This information can be used to improve the user performance in the real execution of a task. The methodology proposed is given by the union of fuzzy statistical measures, fuzzy statistics and fuzzy parameters (fuzzy testing of hypothesis and fuzzy regression models) as input for a fuzzy rule based expert system (FRBES). The FRBES is able to construct an individual profile for each trainee. This new approach is a diagnostic tool that enables a trainee to understand the areas in which he presents difficulties, allowing him to concentrate on improving skills related to them.*

***Keywords*—** Continuous evaluation, fuzzy rule based expert system, fuzzy statistics, fuzzy measures, fuzzy regression models.

1 Introduction

The researches in training evaluation based on virtual reality (VR) [4] are recent. The firsts evaluation systems were offline [3], in which a training based on VR was recorded in videotapes for post-analysis by experts. Recently, online methods were proposed by [13], in which the evaluation is performed during the training process and the user receives that evaluation immediately after the end of training. Since then, several papers were produced in that subject [9, 12, 15, 16, 18, 19, 20, 22, 24, 25, 26]. However, all those methodologies did not use any technique of continuous evaluation to improve trainee performance. Continuous evaluation is a educational methodology used in present and distance learning to help the construction of the knowledge and the cognitive training [1, 10].

In the present work, the goal is to construct a diagnostic to help trainees to understand their difficulties. The first methodology of this kind was proposed only in 2005 by [23] where the goal was to construct a profile to help trainees to understand their difficulties and to improve their performance. That methodology, based on a classical based rule expert system, was able to provide an Evaluation Report and a Continuous Evaluation Report, showing the performance of trainee in the last training and in all trainings performed by him/her, respectively. Morris et al. (2006) suggested the use of statistical linear regression to evaluate user's progress in a bone surgery. After that, in 2009 [17] had proposed another approach based on fuzzy based rule expert system (FRBES), using statistical measures and statistical models, statistical testing of hypothesis, as well as fuzzy measures as input for a FRBES. However, some problems had not been solved yet. The problems are related to assumptions of the statistical models and statistical testing of hypothesis, mainly the Gaussian distributions of data.

In this paper, we propose a new conception of fuzzy continuous evaluation to construct a trainee profile from his/her several trainings and to help him/her to improve his/her performance [2, 6]. In this approach we use fuzzy statistical measures, fuzzy statistics and fuzzy parameters (fuzzy testing of hypothesis and fuzzy regression models) to solve that problems with statistical assumptions. Those fuzzy parameters compose inputs for an fuzzy based rule expert system (FRBES) [30]. The FRBES combines logically all information about fuzzy statistical variables and parameters to making decisions about complex conjectures [7] and is able to construct a trainee profile.

2 Theoretical Aspects

For the reader's better understanding, we first present a short review about fuzzy sets, fuzzy statistics and fuzzy rule based expert system.

2.1 Fuzzy Sets and Fuzzy Numbers

As some variables in the training system do not present an exactly correspondence to the real world, some measures cannot be exact. Then, fuzzy sets are used to measure those variables [7]. In classical set theory a set A of a universe X can be expressed by means of a membership function $\mu_A(x)$, with $\mu_A: X \rightarrow \{0,1\}$, where for a given $a \in A$, $\mu_A(a)=1$ and $\mu_A(a)=0$ respectively express the presence and absence of a in A. Mathematically:

$$\mu_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases} \quad (1)$$

Zadeh [29] introduced the fuzzy set theory in 1965. A fuzzy set or fuzzy subset is used to model an ill-known quantity. A fuzzy set A on X is characterized by its membership function $\mu_A: X \rightarrow [0,1]$. We say that a fuzzy set A of X is "precise" when $\exists\ c^* \in X$ such that $\mu_A(c^*)=1$ and $\forall c \neq c^*$, $\mu_A(c)=0$. A fuzzy set A will be said to be "crisp", when $\forall\, c \in X$, $\mu_A(c) \in \{0,1\}$.

The intersection and union of two fuzzy sets are performed trough the use of *t-norm* and *t-conorm* operators respectively, which are commutative, associative and

ISBN: 978-989-95079-6-8

monotonic mappings from [0,1]→[0,1]. Moreover, a t-norm Γ (respec. t-conorm ⊥) has 1 (respec. 0) as neutral element (e. g.: Γ=*min*, ⊥=*max*) [8]. Thus, we can define intersection and union of two fuzzy sets as:
The intersection of two fuzzy sets *A* and *B*, with membership functions $\mu_A(x)$ e $\mu_B(x)$ is a fuzzy set *C* with membership function given by:

$$C=A \cap B \Leftrightarrow \mu_C(x)= \Gamma\{\mu_A(x),\mu_B(x)\}, \ \forall x\in X. \quad (2)$$

The union of two fuzzy sets *A* and *B*, with membership functions $\mu_A(x)$ e $\mu_B(x)$ is a fuzzy set *C* with membership function given by:

$$C=A \cup B \Leftrightarrow \mu_C(x)= \perp\{\mu_A(x),\mu_B(x)\}, \ \forall x\in X. \quad (3)$$

The complement of a fuzzy set *A* in *X*, denoted by $\neg A$ is defined by:

$$\mu_{\neg A}(x) = n(\mu_A(x)), \ \forall x\in X. \quad (4)$$

where: *n:* [0,1]→[0,1] is a negation operator which satisfies the following properties:

- $n(0)=1$ and $n(1)=0$
- $n(a) \leq n(b)$ if $a>b$
- $n(n(a))=a$, $\forall x\in$ [0,1]

and a negation is a strict negation if it is continuous and satisfies

- $n(a)<n(b)$ if $a>b$

The main negation operator which satisfies these four conditions is $n(a) = 1-a$.
The implication function between two fuzzy sets *A* and *B*, with membership functions $\mu_A(x)$ e $\mu_B(x)$ is a fuzzy set *C* with membership function given by:

$$C=A \Rightarrow B \Leftrightarrow \mu_C(x,y)= \nabla\{\mu_A(x),\mu_B(y)\}, \ \forall x\in X, \ \forall y\in Y \quad (5)$$

where ∇: $[0,1]^2\rightarrow[0,1]$ is an implication operator which obeys the following properties: $\forall a, a', b, b' \in [0,1]$:

- If $b \leq b'$ then $\nabla(a,b) \leq \nabla(a,b')$;
- $\nabla(0,b)=1$;
- $\nabla(1,b)=b$.

The pure implications obeys too:

- If $a \leq a'$ then $\nabla(a,b) \geq \nabla(a',b)$;
- $\nabla(a, \nabla(b,c))= \nabla(b, \nabla(a,c))$.

Beyond concept of fuzzy sets and their operations, it is important to the concept of fuzzy numbers [3]. In this paper we used triangular or triangular shaped fuzzy numbers, which are defined by three real numbers $d<e<f$ where the interval $[d,f]$ is the base of the triangle and *e* is a vertex.

2.2 Fuzzy Statistics

In this paper the statistical methods used were:

1. fuzzy statistical measures;
2. fuzzy statistical models (time dependent or not) and
3. fuzzy statistical testing of hypotheses.

A set of fuzzy statistical measures, commonly used for general purposes as fuzzy mean, fuzzy median, fuzzy standard deviation, etc. [3], can be used to describe user interactions during the training. Besides, fuzzy statistical models based on fuzzy regression can be used to construct fuzzy models for the way followed by user in task execution [10]. This model can be done by:

$$Y = Ax_1 + Bx_2 + Cx_3 \quad (6)$$

where *Y, A, B* and *C* are fuzzy numbers and x_1, x_2 and x_3 are real numbers.
Fuzzy statistical measures and fuzzy statistical parameters of fuzzy models (as *A, B* and *C* above) can be compared with fuzzy parameters of ideal fuzzy models using appropriate fuzzy statistical testing of hypothesis from its α-cuts [3]. As results of these comparisons, we can make fuzzy statistical decisions about equality or difference between parameters using measure of fuzzy probability of significance. The information synthesized by fuzzy statistical measures and fuzzy parameters helps to construct a profile for user, his/her Evaluation Report and his/her Continuous Evaluation Report.

2.3 Fuzzy Rule Based Expert System

Expert systems [27] use the knowledge of an expert in a given specific domain to answer non-trivial questions about that domain. For example, an expert system for image classification would use knowledge about the characteristics of the classes present in a given region to classify a pixel in an image of that region. This knowledge also includes the "how to do" methods used by the human expert. Usually, the knowledge in an expert system is represented by rules as:

```
IF <condition> THEN <conclusion>
```

Most rule-based expert systems allows the use of connectives AND or OR in the premise of a rule, and of connective AND in the conclusion. From rules and facts, new facts will be obtained through an inference process.
In several cases, we do not have precise information about conditions or conclusions. Then, the knowledge in the rules cannot be expressed in a precise manner. Thus, it could be interesting to use a fuzzy rule-based expert system [30]. An example of simple fuzzy rule could be:

```
IF <access to the help is persistent>
   AND <Global Users Performance is bad>
THEN <user is Novice>.
```

where "persistent" and "bad" can be characterized by fuzzy sets.

ISBN: 978-989-95079-6-8

3 Methodology

According to [23], a tool for continuous evaluation must be interconnected with an online evaluation system and must receive information from it about all variables of interest. The evaluation system works near a virtual reality simulator. In general, an online evaluation system should be capable to monitor user interactions while he/she operates the simulation system. In order to achieve that, it is necessary to collect the information about positions in the space, forces, torque, resistance, speeds, accelerations, temperatures, visualization and/or visualization angle, sounds, smells and etc. This information will be used to feed the evaluation system. In the Figure 1 [21], we can observe that the virtual reality simulator and the system of evaluation are independent systems, however they act simultaneously.

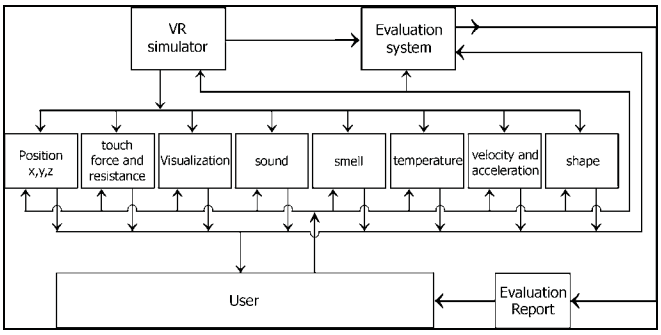


Figure 1: Diagram of A Virtual Reality Simulator With an Evaluation System.

User interactions are monitored and the information are sent to the evaluation system that analyzes the data and emits, at the end of the training, an Evaluation Report about the user performance, according *M* pre-defined classes of performance. A set of rules of the fuzzy based rule expert system (FRBES) [30] defines each one of the possible classes of performance, which are defined from specialists' knowledge. The interaction variables will be monitored according to their relevance to the training. Thus, each application will have their own set of relevant variables that will be monitored [21].

If the same user had performed other trainings, the Continuous Evaluation Tool uses the data collected from user interaction in his/her several training to create an User Profile and construct a Continuous Evaluation Report about all set of training. That information is used to evaluate the trainee and can improve his/her performance in real tasks [28]. The Figure 2 [23] shows a diagram of an Evaluation System able to perform continuous evaluation.

A Fuzzy Continuous Evaluation Tool makes a union of fuzzy statistical measures, models and testing of hypothesis, and an FRBES to construct an individual profile for trainee. Fuzzy statistical tools are programmed to make an automatic analysis of the database and construct fuzzy statistical measures, graphics fuzzy statistical models and results of fuzzy statistical hypothesis testing. FRBES uses this information to create a user profile and a continuous evaluation report. The continuous evaluation report presents the trainee profile and shows, with fuzzy statistical measures, graphics and models, the execution performance of specific tasks. They are used as input for FRBES [30]. Figure 3 shows the new methodology presented.

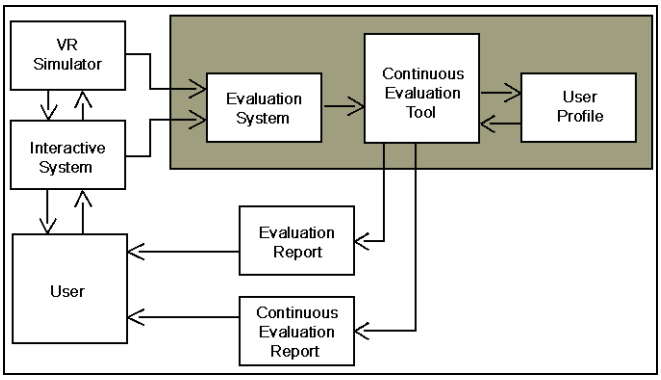


Figure 2: Diagram of an Evaluation System With Approach of Continuous Evaluation. Adapted from [23].

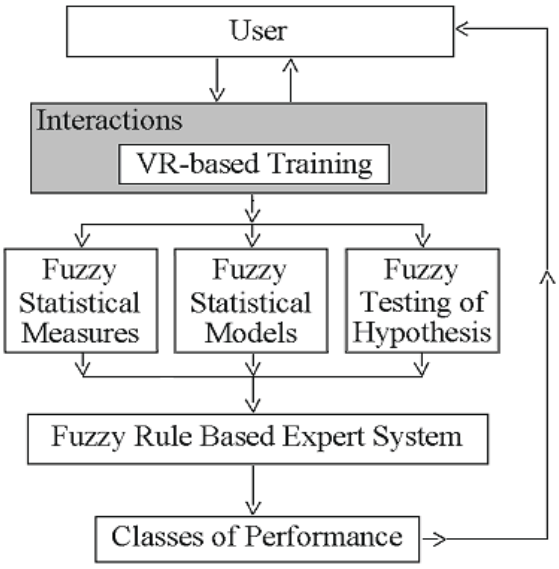


Figure 3: Diagram of The Continuous Evaluation System using FRBES.

In the first time that user executes his training, the Evaluation Report emits information about the user performance, at the end of the training, according to classes of performance previously defined. This information is stored in a User Profile for posterior evaluations with approach of continuous evaluation. In a second time when user execute his/her training, the Fuzzy Continuous Evaluation is able to construct a Continuous Evaluation Report, which presents information about user performance over specific tasks, using statistical measures, tables, graphics and models. Both reports present information from the last training. But, additionally, the Continuous Evaluation Report will show accumulated information about the sequence of trainings for this user.

ISBN: 978-989-95079-6-8

4 Application

This methodology can be applied for any activity, particularly those who offer risks to the user/trainee or to people who depends of him/her. In this context, fuzzy continuous evaluation is an interesting tool to improve knowledge construction. For example in medical area, where invasive procedures can be simulated by VR, it is necessary some kind of evaluation tool with properties of continuous evaluation. In those simulators, users perform realistic virtual procedures and can acquire dexterity and/or improve skills. If a continuous evaluation tool is attached to those applications, user interactions can be collected in real time to be used to evaluation since the tool can show to the user his/her qualities and his/her deficiencies in the execution of the medical procedure.

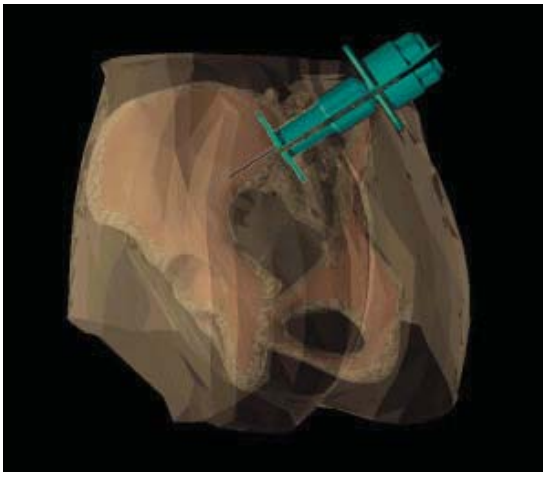


Figure 4: Screenshot of the Bone Marrow Simulator with a Semi-Transparent View of the Pelvic Region.

An example is the bone marrow simulator (Figure 4), a virtual reality simulator to training the extraction of bone marrow in children [13]. In this application the user is a novice surgeon that must acquire dexterity to insert a needle in the pelvic region of a child and find the bone marrow, located inside the iliac bone. The user manipulates a haptic device, a 6DOF interaction device, represented in the system by a needle. This device is responsible to provide tactile sensations and force feedback related to the manipulation of the needle in the system. With this device, the user can touch and feel the tissue properties of anatomy models, use his perception to identify bone location under the skin and penetrate the tissues to extract bone marrow. Figure 5 shows the haptic device used for interaction. In spite of Figure 4 presents a semi-transparent model, it is only for demonstration purposes. In the real bone marrow harvest procedure user cannot have internal visualization of the body and the touch sense is the only information perceived during needle penetration over 4 different properties tissue layers. Table 1 presents an example of description of each tissue layer present in the bone barrow harvest simulator. The hardness refers how much force must be applied to penetrate the tissue, the viscosity refers to how hard are the movements when inside the layer and surface friction indicates how slippery is the surface of the layer. The percentages are only an estimation of values and the calibration of the properties is done by a physician. The haptic device used in the simulation can read 1000 samples per second, including spatial position and applied forces. This data is an example of information used by the evaluation tool. After calibration and system setup, the same physician performs the simulation several times and in different ways to provide labels for each one of M classes of performance.



Figure 5: Haptic Device for Interaction with Tactile Sensation and Force Feedback.

Table 1: Subjective description of tissue properties for a bone marrow simulator.

Tissue	*hardness*	*viscosity*	Surface friction
Epidermis	20%	70%	60%
Subcutaneous	20%	70%	60%
Periosteum	80%	90%	20%
Hard bone	80%	90%	20%
Bone marrow	1%	5%	1%

The FRBES contains different rules to manipulate fuzzy information collected from users interaction using haptic device. The user must introduce the needle in pelvic region and go to the bone to harvest the marrow. This path can be modeled by a fuzzy regression model and the parameters of this regression can be compared with the M models previously stored by the physician. In the procedure, forces and torques applied through the needle are approximately constants between skin and bone. However, it is necessary to apply a different torque to introduce the needle into the bone. Fuzzy statistical parameters, as mean and standard deviation, can be used to compare this procedure with the parameters of procedure performed by physician. All those fuzzy statistical parameters can be compared with parameters of classes of performance using a fuzzy testing of hypothesis. The fuzzy probability of these comparisons and their correspondent hypothesis are used as input in the FRBES. If everything was performed correctly, an Evaluation Report is created to inform the user about it. In opposition, if something was performed wrong, the FRBES use their rule databases to find problems in execution and inform the user. Examples are presented in the following.

ISBN: 978-989-95079-6-8

4.1 Example 1

Let be five classes of performance done by fuzzy sets in Ω={very good, good, reasonable, bad, novice}. The fuzzy regression model for the way followed by user has parameters (for coordinates of tri-dimensional space), done by the fuzzy sets *A, B* and *C*.

Vectors of data, whose fuzzy mean and fuzzy standard deviation are given by D_{f1}, E_{f1}, G_{t1} and H_{t1}, respectively, store the force and torque applied between skin and bone. In the introduction of the needle into the bone, force and torque new vectors of data are stored, whose fuzzy mean and fuzzy standard deviation are given by D_{f2}, E_{f2}, G_{t2} and H_{t2}, respectively.

For each one class of performance, there are reference parameters for models, forces and torques. Those parameters are denoted by *A*, B** and *C** for fuzzy models and D_{fi}*, E_{fi}*, G_{ti}*, H_{ti}* , with *i=1,2* for forces and torques respectively. To compare those parameters with these reference parameters for each class, fuzzy statistical hypothesis testing are used and fuzzy probabilities for null hypothesis are obtained from them. These probabilities are treated by FRBES using rules as:

```
IF <Fuzzy Probability of Applied Forces in
    Phase 1 is Acceptable>
   AND <Fuzzy Probability of Applied Forces in
        Phase 2 is Unacceptable>
THEN <User Performance is bad>.
```

or

```
IF <Fuzzy Probability of A parameter in Fuzzy
    Model is Acceptable>
   AND <Fuzzy Probability of B parameter in
        Fuzzy Model is Acceptable>
   AND <Fuzzy Probability of C parameter in
        Fuzzy Model is Acceptable >
THEN <User Performance is very good>.
```

When the system classifies a user as good, reasonable, bad or novice, the FRBES can detect the user mistakes by the search for unacceptable fuzzy probabilities. From them, is possible to find where user made mistake. For example, it can be noticed the user made mistake in the application of forces in the Phase 2, but everything was performed correctly in the Phase 1.

```
IF <Global User Performance is bad >
   AND <Fuzzy Probability of Applied Forces in
        Phase 2 is Unacceptable>
THEN <Applied Forces in Phase 2 has
      Mistake>.
```

By the rule above, the user will know about his/her mistake. Also it is possible to know what was performed wrong. In this case, only three cases are possible:

1. Applied force was excessive and it was not possible to harvest the marrow.
2. Applied force was normal and it produced a good procedure.
3. Applied force was lower than normal, and the needle cannot be introduced into the bone.

The FRBES has rules to verify which situation really happened. For example:

```
IF <Global User Performance is bad >
   AND <User Applied Forces in Phase 2 has
        Mistake>
   AND <Applied Forces in Phase 2 is Excessive>
THEN <Applied Forces in Phase 2 is Upper
      than normal>.
```

4.2 Example 2

In the Section 4.1 it was illustrated how the Evaluation Report works. Now, we present the Continuous Evaluation Report. This report must show the evolution of user according to the sequence of training he/she performed, including the last one. The goal of this report is to help user to improve his/her performance.

Let be a training sequence for a user, denoted by $S=\{s_1,..., s_n\}$. In each s_j ($j=1,...,n$) is stored a set of vectors with: variables of training, performance by variable of training and global user performance. From this information the FRBES can produce a Continuous Evaluation Report, showing to the user his/her evolution. For example, mistakes are less frequent now than before; serious mistakes are not made there is a long time; average of the time of procedure execution is coming to ideal with small standard deviation. The rules below show how it is processed:

```
IF (<Fuzzy Correlation (serious mistakes) is
     Negative OR Approximate Zero)
    AND <Fuzzy Correlation (number of mistakes)
         is Negative OR Approximate Zero>
THEN <User's Continuous Evaluation is Good>.
```

or

```
IF (<Fuzzy Probability of Execution Time is
     Acceptable)
    AND <Fuzzy Standard Deviation (Execution
         Time) is Small>
THEN <User's Continuous Evaluation is Very
      Good>.
```

FRBES can evaluate the global users performance along the sequence of performed procedures. It is done by rules like this one:

```
IF <Fuzzy Correlation (Global User Performance)
    is Positive OR Approximate Zero>
   AND <The Last One (Global User Performance)
        is Very Good>
THEN <User's Continuous Evaluation is Very
      Good>.
```

For better user comprehension of his/her situation, some graphics are presented showing all history of evaluations. For example, a graphic illustrate the number of serious mistakes or his/her Global Users Performance along the sequence of performed procedures.

5 Conclusions

In this paper we introduced a new methodology for evaluation training using a fuzzy continuous evaluation approach. This methodology uses fuzzy statistical measures, models and results of fuzzy statistical hypothesis testing, as inputs of a FRBES. This system is able to construct an

ISBN: 978-989-95079-6-8

individual profile for trainee and emit to him/her information about his/her performance at the end of the training, according to classes of performance previously defined, as proposed in others methodologies. Moreover, this methodology can provide to user information about his performance in specific tasks in the training and show where the user made mistakes. The methodology was illustrated by examples to show its functionalities and how the Evaluation Report and Continuous Evaluation Report are made.

A system developed using the proposed methodology is a diagnostic tool, which helps a trainee to understand his/her difficulties. From information presented by a Fuzzy Continuous Evaluation system, the trainee can understand his/her difficulties and improve his performance.

Acknowledgment

This work is partially supported by Brazilian Council for Scientific and Technological Development, CNPq (Process 303444/2006-1) and Brazilian Research and Projects Financing, FINEP (Grant 01-04-1054-000).

References

[1] G.M. Aalsvoort, W.C.M. Resing and A.J.J.M. Ruijssenaars, *Learning potential assessment and cognitive training: Actual research and perspectives in theory building and methodology*. JAI/Elsevier Science, 2002.

[2] J.A. Arter and J. Mctighe, *Scoring Rubrics In The Classroom: Using Performance Criteria For Assessing And Improving Student Performance*. Corwin Press, 2000.

[3] J.J. Buckley. *Fuzzy Probability and Statistics*. Springer, 2006.

[4] G. Burdea, G. Patounakis, V. Popescu and R.E. Weiss, Virtual Reality Training for the Diagnosis of Prostate Cancer. Virtual Reality Annual International Symposium Proceedings. IEEE, pp. 190-197, 1998.

[5] G. Burdea and P. Coiffet, *Virtual Reality Technology*. 2nd ed. New Jersey, John Wiley & Sons, 2003.

[6] M.O. Carlson, G.E. Humphrey and K.S. Reinhardt, *Weaving Science Inquiry and Continuous Assessment: Using Formative Assessment to Improve Learning*. Corwin Press, 2003.

[7] D. Dubois and H. Prade, *Fuzzy sets and systems - theory and applications*. New York: Academic Press, 1980.

[8] D. Dubois and H. Prade, *Possibility theory: An approach to computerized processing of uncertainty*. Plenum Press, 1988.

[9] M. Färber et al. "Heinz Handels Training and evaluation of lumbar punctures in a VR-environment using a 6DOF haptic device", Studies in Health Tech. and Informatics, 132:112-114, 2008.

[10] Galef Institute; *Assessment: Continuous Learning*. Stenhouse Publishers, 1995.

[11] J. Kacprzyk, M. Fedrizzu (eds): Fuzzy Regression Analysis. Physica Verlag. Heidelberg, 1992.

[12] K. Kumagai, J. Yamashita, O. Morikawa, and K. Yokoyama, "A New Force-Based Objective Assessment of Technical Skills in Endoscopic Sinus Surgery". Studies in Health Technology and Informatics, 125:235-237, 2007.

[13] L.S. Machado, Virtual Reality to Modeling and Simulate Invasive Procedures for Pediatrics Oncology: A Case Study in Bone Marrow Transplant. PhD Thesis (in Portuguese). Polytechnic School - University of São Paulo, 2003.

[14] L.S. Machado, R.M. Moraes and M.K. Zuffo, A Fuzzy Rule-Based Evaluation for a Haptic and Stereo Simulator for Bone Marrow Harvest for Transplant. In: Annals of Phantom Users Group 2000. Aspen, 2000.

[15] L.S. Machado, M.C.O. Valdek and R.M. Moraes, On-Line Assessment System for a Training System with Multiple Stages Based on Virtual Reality. *Journal of Multiple-Valued Logic and Soft Computing*, 14(3-5):511-524, 2008.

[16] L.S. Machado and R.M. Moraes, Qualitative and Quantitative Assessment for a VR-Based Simulator. *Studies in Health Technology and Informatics*, 2009. In press.

[17] L.S. Machado and R.M. Moraes, Continuous Evaluation in Training Systems Based on Virtual Reality Using Fuzzy Rule Based Expert Systems. Proceedings of VI International Conference on Engineering and Computer Education (ICECE'2009). Buenos Aires, 2009. In press.

[18] P.B. McBeth et al. Quantitative Methodology of Evaluating Surgeon Performance in Laparoscopic Surgery. *Studies in Health Technology and Informatics*. 85:280-286, 2002.

[19] L. Moody et al., Objective Surgical Performance Evaluation Based on Haptic Feedback. *Studies in Health Technology and Informatics* 85:304-310, 2002.

[20] R.M. Moraes and L.S. Machado, Hidden Markov Models For Learning Evaluation In Virtual Reality Simulators. *International Journal of Computers & Applications*, 25(3):212-215, 2003.

[21] R.M. Moraes and L.S. Machado, Fuzzy Gaussian Mixture Models for Online Training Evaluation in Virtual Reality Simulators. In: Proceedings of International Conference on Fuzzy Information Processing. China, v. 2, p. 733-740, 2003.

[22] Moraes, R. M.; Machado, L. S.; Using Fuzzy Hidden Markov Models for Online Training Evaluation and Classification in Virtual Reality Simulators. *International Journal of General Systems*, 33(2-3):281-288, 2004.

[23] R.M. Moraes and L.S. Machado, Continuous Evaluation in Training Systems Based on Virtual Reality Proceedings of Global Congress on Engineering and Technology Education (GCETE'2005). Brazil, p.1048-1051, 2005.

[24] R.M. Moraes and L.S. Machado, Fuzzy Bayes Rule for On-Line Training Assessment in Virtual Reality Simulators. *Journal of Multiple-Valued Logic and Soft Computing*, 14(3-5):325-338, 2008.

[25] Morris, et al.; Visuohaptic simulation of bone surgery for training and evaluation. *IEEE Computer Graphics & Applications*, 26(6): 48-57, 2006.

[26] D.J. Pucel and L.D. Anderson, Developing Computer Simulation Performance Tests: Challenges and Criteria. Proc. 2003 IASTED International Conference on Computers and Advanced Technology in Education (CATE'2003). Greece, pp 170-174,2003.

[27] E. Rich and K. Knight, *Artificial Intelligence*. McGraw Hill, 1993.

[28] R.J. Sternberg and E. Grigorenko, *Dynamic Testing: The Nature and Measurement of Learning Potential*. Cambridge University Press, 2001.

[29] L.A. Zadeh. Fuzzy Sets. *Information Control*. (8):338-353, 1965.

[30] L.A. Zadeh. Fuzzy Logic. *Computer*, (8):83-93, 1988.

ISBN: 978-989-95079-6-8

A Comparative Analysis of Symbolic Linguistic Computational Models

R.M. Rodríguez[1] L. Martínez[1] M. Espinilla[1]
1.Computer Sciences Department, University of Jaén
Jaén, Campus Las Lagunillas s/n 23071 - Jaén. Spain.
Email: rmrodrig@ujaen.es, martin@ujaen.es, mestevez@ujaen.es

Abstract*— *There are many situations in which problems deal with vague and imprecise information. In such cases, the information could be modelled by means of numbers, however, it doesn't seem logical to model imprecise information in a precise way. Therefore, the use of linguistic modelling have been used with successful results in these problems. The use of linguistic information involves the need of carrying out processes which operate with words, so called Computing with Words (CW). In the literature exists different linguistic approaches and different computational models. We focus in this contribution on the use of the fuzzy linguistic approach (FLA) to model vague and imprecise information, but more specifically we focus on their computational models paying more attention on different symbolic computational models that have been defined to deal with linguistic information. We are going to review their main features and make a comparative analysis among them.

***Keywords*—** Computing with words (CW), fuzzy linguistic approach, linguistic 2-tuple, linguistic variable.

1 Introduction

Many problems in the real world deal with vague and imprecise information. There exist different kinds of tools to manage this type of information. The probability theory can be a powerful tool in order to treat the uncertainty and can be applied in different areas, like decision making, evaluation, planning, scheduling and so on. However, it is easy to see that many aspects of uncertainties have a non-probabilistic character since they are related to imprecision and vagueness of meanings. The use of the fuzzy linguistic approach [18] to model this kind of information provided successful results because the experts involved in such situations provide linguistic values rather than numbers. The linguistic modelling implies processes of CW, in the specialized literature can be found two classical linguistic computational models that provide linguistic operators for CW:

i) Model based on the extension principle (semantic model) [2, 4].

ii) Symbolic model [6, 8, 16].

The former provides accuracy but their results cannot be expressed by linguistic terms without an approximation process. The latter also needs an approximation process to express the results in a linguistic way, but the computational process is simpler and easier to understand by the experts involved in the problems.

Due to the previous facts, different symbolic approaches based on the fuzzy linguistic approach have been defined to improve the classical computational model. These approaches have modified the representation of the linguistic information from different points of view in order to improve the computational results. The symbolic models in which we are interested are: the 2-tuple linguistic representation model [9], the virtual linguistic model [15] and the proportional 2-tuple model [13]. These models improve the classical symbolic model by avoiding the approximation in processes of CW in order to improve the precision in the final results.

The aim of this contribution is to make a comparative analysis of the three aforementioned symbolic models to show their features and discuss their correctness regarding the fuzzy linguistic approach as their basis.

This paper is structured as follows: In Section 2, we introduce in short the fuzzy linguistic approach and its classical computational models. In Section 3, we shall review the different symbolic computational models, such as, the 2-tuple linguistic representation model, the virtual linguistic model and the proportional 2-tuple representation model. In Section 4, we shall make a comparative analysis among the different symbolic computational models, and finally we shall point out some concluding remarks.

2 Fuzzy Linguistic Approach

Many aspects of different activities in the real world cannot be assessed in a quantitative form, but rather in a qualitative one, i.e., with vague or imprecise knowledge. In that case, a better approach might be to use linguistic assessments instead of numerical values. The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables [18]. We have to choose the appropriate linguistic descriptors for the linguistic term set and their semantics. To do so, an important aspect to be analyzed is the *granularity of uncertainty* i.e., the level of discrimination among different degrees of uncertainty. Typical values of cardinality used in the linguistic models are odd ones, such as 7 or 9, where the mid term represents an assessment of "roughly 0.5" and the rest of the terms being placed symmetrically around it [2]. Once the cardinality of the linguistic term set has been established, we must provide the linguistic terms and their semantics. There exist different possibilities to accomplish this task [1, 8, 17]. One of them consists in supplying directly the term set by considering all the terms distributed on a scale on which a total order is defined [8, 17]. For example, a set of seven terms S, could be:

$$S = \{s_0 : N, s_1 : VL, s_2 : L, s_3 : M, s_4 : H, s_5 : VH, s_6 : P\}$$

Usually, in these cases, it is required that in the linguistic term set there exists:

1. A negation operator $\text{Neg}(s_i) = s_j$ such that $j = g - i$ ($g + 1$ is the cardinality)

ISBN: 978-989-95079-6-8

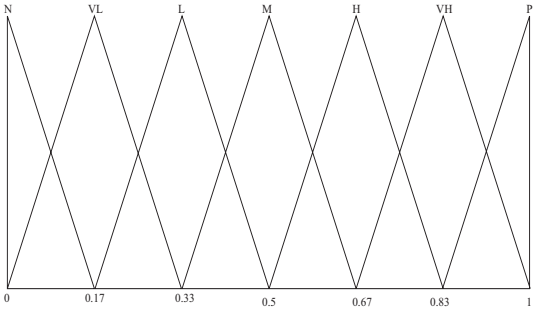


Figure 1: A Set of 7 Terms with its Semantic

2. A max operator: $\max(s_i, s_j) = s_i \; if \; s_i \geq s_j$

3. A min operator: $\min(s_i, s_j) = s_i \; if \; s_i \leq s_j$

The semantics of the terms are represented by fuzzy numbers, defined in the interval $[0, 1]$, described by membership functions. A way to characterize a fuzzy number is to use a representation based on parameters of its membership function [2]. The linguistic assessments given by the users are just approximate ones, then linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments [5]. This representation is achieved by the 4-tuple (a, b, c, d), where b and d indicate the interval in which the membership value is 1, with a and c indicating the left and right limits of the definition domain of the trapezoidal membership function [2]. A particular case of this kind of representation are the linguistic assessments whose membership functions are triangular, i.e., $b = d$, so we represent this kind of membership function by 3-tuples (a, b, c). An example can be:

$$\begin{array}{ll} P = (.83, 1, 1) & VH = (.67, .83, 1) \\ H = (.5, .67, .83) & M = (.33, .5, .67) \\ L = (.17, .33, .5) & VL = (0, .17, .33) \\ N = (0, 0, .17). & \end{array}$$

which is graphically shown in Figure 1.

2.1 Classical Computational Models

The use of linguistic variables implies processes of computing with words such as their fusion, aggregation, comparison, etc. To perform these computations in the fuzzy linguistic approach appeared two classical computational models:

1. **Model based on the Extension Principle (Semantic Model):** This model carries out operations with linguistic terms by means of operations associated to their membership functions based on the Extension Principle. The Extension Principle is a basic concept in the fuzzy sets theory [7] which is used to generalize crisp mathematical concepts to fuzzy sets. The use of extended arithmetic based on the Extension Principle [7] increases the vagueness of the results. Therefore, the results obtained by the fuzzy linguistic operators based on the Extension Principle are fuzzy numbers that usually do not match with any linguistic term in the initial term set. For this reason, it is necessary to carry out a linguistic approach in order to express the results in the original expression domain. In the literature, we can find different linguistic approximation operators [2, 4]. A linguistic aggregation operator based on the extension principle acts according to:

$$S^n \xrightarrow{\tilde{F}} F(\mathcal{R}) \xrightarrow{app_1(\cdot)} S \tag{1}$$

where S^n symbolizes the n Cartesian product of S. $\tilde{F}$ is an aggregation operator based on the extension principle, $F(R)$ the set of fuzzy sets over the set of real numbers R, $app_1 : F(R) \rightarrow S$ is a linguistic approximation function that returns a label from the linguistic term set S, being S the initial term set.

2. **Symbolic Model:** This model uses the ordered structure of the linguistic terms set, $S = \{s_0, s_1, \ldots, s_g\}$ where $s_i < s_j$ if $i < j$, to operate [6, 8, 16]. The intermediate results of these operations are numeric values, $\alpha \in [0, g]$, which must be approximated in each step of the process by means of an approximation function $app_2 : [0, g] \rightarrow \{0, \ldots, g\}$ that obtains a numeric value, such that, it indicates the index of the associated linguistic term, $s_{app_2(\alpha)} \in S$. Formally, it can be expressed as:

$$S^n \xrightarrow{C} [0, g] \xrightarrow{app_2(\cdot)} \{0, \ldots, g\} \longrightarrow S \tag{2}$$

where C is a symbolic linguistic aggregation operator, $app_2(\cdot)$ is an approximation function used to obtain an index $\{0, \ldots, g\}$ associated to a term in $S = \{s_0, \ldots, s_g\}$ from a value in $[0, g]$

Both models when operate with linguistic information produce loss of information due to the approximation processes and hence a lack of precision in the results. This loss of information is produced because the information representation model of the fuzzy linguistic approach is discrete in a continuous domain.

3 New Symbolic Computational Models

The necessity of dealing with linguistic information in many real world problems has driven the researchers to develop models in order to improve the processes of CW. Different models have been presented in the literature recently. In this section we focus our attention on different symbolic approaches that have developed new representation and computational models for the linguistic information in order to improve the accuracy of the results of processes of CW. Such approaches are the 2-tuple linguistic model [9], the virtual linguistic model [15] and the proportional 2-tuple linguistic model [13], which improve the limitations that present the classical symbolic computational model, regulating the loss of information and imprecision in its computations. These approaches have been widely used in problems dealing with linguistic information such as, Decision making [12, 14], Evaluation [3], Recommender Systems [11] and so on.

Due to the fact that, these symbolic approaches based on the fuzzy linguistic approach have modified the linguistic representation in order to improve the processes of CW, their review implies the study of their linguistic representation models and their computational models to accomplish the processes of CW in a symbolic and precise way.

ISBN: 978-989-95079-6-8

3.1 2-Tuple linguistic representation model

This model was presented in [9] to avoid the loss of information and to express symbolically any counting of information in the universe of the discourse.

(a) Representation model:

This representation is based on the concept of *symbolic translation* and uses it for representing the linguistic information by means of a pair of values, called *2-tuples*, (s_i, α) where s is a linguistic term and α is a numerical value representing the symbolic translation. Let $S = \{s_0, \ldots, s_g\}$ be a term set, and $\beta \in [0, g]$ a numerical value in its interval of granularity (e.g.: let β be a value obtained from a symbolic aggregation operation).

Definition 1 *The symbolic translation is a numerical value assessed in* $[-0.5, 0.5)$ *that supports the "difference of information" between a counting of information* β *assessed in the interval of granularity* $[0, g]$ *of the term set* S *and the closest value in* $\{0, \ldots, g\}$ *which indicates the index of the closest linguistic term in* S.

From this concept, is developed a linguistic representation model which represents the linguistic information by means of 2-tuples (s_i, α), $s_i \in S$ and $\alpha_i \in [-0.5, 0.5)$

- s_i represents the linguistic label of the information
- α is a numerical value expressing the value of the translation

This representation model defines a set of functions to facilitate computational processes with 2-tuples [9].

Definition 2 *Let* $S = \{s_0, \ldots, s_g\}$ *be a set of linguistic terms. The 2-tuple set associated with* S *is defined as* $\langle S \rangle = S \times [-0.5, 0.5)$. *We define the function* $\Delta : [0, g] \longrightarrow \langle S \rangle$ *given by*

$$\Delta(\beta) = (s_i, \alpha), \text{ with } \begin{cases} i = \text{round}\,(\beta), \\ \alpha = \beta - i, \end{cases} \tag{3}$$

where round *assigns to* β *the integer number* $i \in \{0, 1, \ldots, g\}$ *closest to* β.

We note that Δ is bijective [9, 10] and $\Delta^{-1} : \langle S \rangle \longrightarrow [0, g]$ is defined by $\Delta^{-1}(s_i, \alpha) = i + \alpha$. In this way, the 2-tuples of $\langle S \rangle$ will be identified with the numerical values in the interval $[0, g]$.

Remark 1 *We can consider the injective mapping* $S \longrightarrow \langle S \rangle$ *that allows us to transform a linguistic term* s_i *into a 2-tuple:* $(s_i, 0)$. *On the other hand,* $\Delta_S(i) = (s_i, 0)$ *and* $\Delta_S^{-1}(s_i, 0) = i$, *for every* $i \in \{0, 1, \ldots, g\}$.

Let's suppose a symbolic aggregation operation over labels assessed in $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ that obtains as its result $\beta = 2.8$, then the representation of this counting of information by means of a 2-tuple will be:

$$\Delta(2.8) = (s_3, -0.2)$$

b) Computational model:

Together the representation model, a linguistic computational approach based on the functions Δ and Δ^{-1} was also defined in [9] with the following computations and operators:

1. Comparison of 2-tuples

The comparison of linguistic information represented by 2-tuples is carried out according to an ordinary lexicographic order.

Let (s_k, α_1) and (s_l, α_2) be two 2-tuples, with each one representing a counting of information:

- if $k < l$ then $(s_k, \alpha_1) < (s_l, \alpha_2)$
- if $k = l$ then
  1. if $\alpha_1 = \alpha_2$ then (s_k, α_1), (s_l, α_2) represents the same information
  2. if $\alpha_1 < \alpha_2$ then $(s_k, \alpha_1) < (s_l, \alpha_2)$
  3. if $\alpha_1 > \alpha_2$ then $(s_k, \alpha_1) > (s_l, \alpha_2)$

2. Negation operator of a 2-tuple

The negation operator over 2-tuples was defined as:

$$Neg((s_i, \alpha)) = \Delta(g - (\Delta^{-1}(s_i, \alpha))) \tag{4}$$

where $g + 1$ is the cardinality of S, $S = \{s_0, ..., s_g\}$.

3. Aggregation of 2-tuples

The aggregation of information consists of obtaining a value that summarizes a set of values, therefore, the result of the aggregation of a set of 2-tuples must be a 2-tuple. There exists several 2-tuple aggregation operators [9]. For instance, the 2-tuple arithmetic mean is defined as:

Definition 3 *Let* $x = \{(s_1, \alpha_1), \ldots, (s_n, \alpha_n)\}$ *be a set of 2-tuples, the 2-tuple arithmetic mean* $\overline{x}^e$ *is computed as,*

$$\overline{x}^e = \Delta(\sum_{i=1}^{n} \frac{1}{n} \Delta^{-1}(s_i, \alpha_i)) = \Delta(\frac{1}{n} \sum_{i=1}^{n} \beta_i) \tag{5}$$

The arithmetic mean for 2-tuples allows to compute the mean of a set of linguistic values in a precise way without any approximation process.

Example

Let's suppose an example where we have the following linguistic preference vector:

s_2	s_3	s_3	s_2

and $S = \{s_0 : N, s_1 : VL, s_2 : L, s_3 : M, s_4 : H, s_5 : VH, s_6 : P\}$ is the term set shown in the Figure 1.

We want to aggregate these values by using the arithmetic mean as aggregation operator. We follow the process below:

- The preference vector is transformed into 2-tuples as follows:

$(s_2, 0)$	$(s_3, 0)$	$(s_3, 0)$	$(s_2, 0)$

- The linguistic aggregated value obtained by the arithmetic 2-tuple is:
  $\overline{x} = \Delta(\frac{1}{4}(\Delta^{-1}(s_2, 0) + \Delta^{-1}(s_3, 0) + \Delta^{-1}(s_3, 0) + \Delta^{-1}(s_2, 0))) = \Delta(2.5) = (s_3, -0.5)$

ISBN: 978-989-95079-6-8

3.2 Virtual linguistic model

This model was presented by Xu in [15] to avoid the loss of information in processes of CW and increase the operators in processes of CW.

a) Representation model:

In this symbolic model, Xu extended the discrete term set S to a continuous linguistic term set $\bar{S} = \{s_\alpha | s_l < s_\alpha \leq s_t, \alpha \in [1,t]\}$, where, if $s_\alpha \in S$, s_α is called *the original linguistic term*, otherwise, s_α is called *virtual linguistic term* which **does not have assigned any semantics**.

In general, experts use the original linguistic terms to assess the linguistic variables, and the virtual linguistic terms appear in operations.

b) Computational model:

To accomplish processes of CW with this representation model, Xu introduced the following operational laws:

Let $s_\alpha, s_\beta \in \bar{S}$, be any two linguistic terms and $\mu, \mu_1, \mu_2 \in [0,1]$.

1. $(s_\alpha)^\mu = s_{\alpha^\mu}$
2. $(s_\alpha)^{\mu_1} \otimes (s_\alpha)^{\mu_2} = (s_\alpha)^{\mu_1+\mu_2}$
3. $(s_\alpha \otimes s_\beta)^\mu = (s_\alpha)^\mu \otimes (s_\beta)^\mu$
4. $s_\alpha \otimes s_\beta = s_\beta \otimes s_\alpha = s_{\alpha\beta}$
5. $s_\alpha \oplus s_\beta = s_{\alpha+\beta}$
6. $s_\alpha \oplus s_\beta = s_\beta \oplus s_\alpha$
7. $\mu s_\alpha = s_{\mu\alpha}$
8. $(\mu_1 + \mu_2)s_\alpha = \mu_1 s_\alpha \oplus \mu_2 s_\alpha$
9. $\mu(s_\alpha \oplus s_\beta) = \mu s_\alpha \oplus \mu s_\beta$

Example

Let's suppose the example presented previously. In order to compute the arithmetic mean with this model, we have to apply the above operational rules to the linguistic terms:

- The arithmetic mean according to Xu is defined as:

$$\overline{x}^e = \frac{\sum_{i=1}^n s_i}{n} = \frac{1}{n} s_{\sum_{i=1}^n i} \tag{6}$$

- We aggregate the preference vector and we obtain the following collective preference value

$$\overline{x} = \frac{1}{4} s_{(2+3+3+2)} = \frac{1}{4} s_{10} = s_{2.5}$$

3.3 Proportional 2-Tuples representation model

This model presented by Wang and Hao in [13] develops a new way to represent the linguistic information that is a generalization and extension of 2-tuple linguistic representation model [9]. This model deals with linguistic labels in a precise way, but it does not require that the labels are symmetrically distributed around a medium label and either having "equal distance" between them. Besides, it describes the initial linguistic information by members of a "continuous" linguistic scale domain which does not necessarily require the ordered linguistic terms of a linguistic variable being equidistant.

a) Representation model:

This model represents the linguistic information by means of proportional 2-tuples, such as $(0.2A, 0.8B)$ for the case when someone's grades in the answer scripts of a whole course are distributed as $20\%A$ and $80\%B$. The authors point out that if B were used as the approximative grade then some performance information would be lost. This approach, proportional 2-tuples, is based on the concept of *symbolic proportion* [13].

Definition 4 *Let $S = \{s_0, s_1, ..., s_g\}$ be an ordinal term set, $I = [0,1]$ and*

$$IS \equiv I \times S = \{(\alpha, s_i) : \alpha \in [0,1] \; and \; i = 0, 1, ..., g\} \tag{7}$$

where S is the ordered set of $g+1$ ordinal terms $\{s_0, ..., s_g\}$. Given a pair (s_i, s_{i+1}) of two successive ordinal terms of S, any two elements (α, s_i), (β, s_{i+1}) of IS is called a symbolic proportion pair and α, β are called a pair of symbolic proportions of the pair (s_i, s_{i+1}) if $\alpha+\beta = 1$. A symbolic proportion pair (α, s_i), $(1-\alpha, s_{i+1})$ is denoted by $(\alpha s_i, (1-\alpha)s_{i+1})$ and the set of all the symbolic proportion pairs is denoted by $\overline{S}$, i.e., $\overline{S} = \{(\alpha s_i, (1-\alpha)s_{i+1}) : \alpha \in [0,1] \; and \; i = 0, 1, \ldots, g-1\}$.

Remark 2 *Since for $i = \{2, \ldots, g-1\}$, ordinal term s_i can use either $(0s_{i-1}, 1s_i)$ or $(1s_i, 0s_{i+1})$ as its representative in $\overline{S}$, by abuse of notation.*

$\overline{S}$ is called the *ordinal proportional 2-tuple set* generated by S and the members of $\overline{S}$, *ordinal proportional 2-tuples*, which is used to represent the ordinal information for CW.

The notion of proportional 2-tuple allows experts to express their opinions using two adjacent ordinals.

In a similar way to the symbolic 2-tuple Wang and Hao introduced functions in order to facilitate the computations with this type of representation.

Definition 5 *Let $S = \{s_0, s_1, \ldots, s_g\}$ be an ordinal term set and $\overline{S}$ be the ordinal proportional 2-tuple set generated by S. The function $\pi : \overline{S} \rightarrow [0, g]$ was defined by*

$$\pi((\alpha s_i, (1-\alpha)s_{i+1})) = i + (1-\alpha), \tag{8}$$

where $i = \{0, 1, \ldots, g-1\}, \alpha \in [0,1]$ and π is called the position index function of ordinal 2-tuples.

Note that, under the identification convention which was remarked after the definition 4, the position index function π becomes a bijection from $\overline{S}$ to $[0, g]$ and its inverse $\pi^{-1} : [0, g] \rightarrow \overline{S}$ is defined by

$$\pi^{-1}(x) = ((1-\beta)s_i, \beta s_{i+1}) \tag{9}$$

where $i = E(x)$, E is the integer part function, $\beta = x - i$.

b) Computational model:

To operate with linguistic information under proportional 2-tuple contexts, Wang and Hao expanded the computational

ISBN: 978-989-95079-6-8

techniques for symbolic to proportional 2-tuples and defined the following operators:

1. Comparison of Proportional 2-tuples

The comparison of linguistic information represented by proportional 2-tuples is carried out as follows: Let $S = \{s_0, \ldots, s_g\}$ be an ordinal term set and $\bar{S}$ be the ordinal proportional 2-tuple set generated by S. For any $(\alpha s_i, (1-\alpha)s_{i+1}), (\beta s_j, (1-\beta)s_{j+1}) \in \bar{S}$, define $(\alpha s_i, (1-\alpha)s_{i+1}) < (\beta s_j, (1-\beta)s_{j+1}) \Leftrightarrow \alpha i + (1-\alpha)(i+1) < \beta j + (1-\beta)(j+1) \Leftrightarrow i + (1-\alpha) < j + (1-\beta)$.

Thus, for any two proportional 2-tuples $(\alpha s_i, (1-\alpha)s_{i+1})$ and $(\beta s_j, (1-\beta)s_{j+1})$:

- if $i < j$, then
  1. $(\alpha s_i, (1-\alpha)s_{i+1}), (\beta s_j, (1-\beta)s_{j+1})$ represents the same information when $i = j-1$ and $\alpha = 0, \beta = 1$
  2. $(\alpha s_i, (1-\alpha)s_{i+1}) < (\beta s_j, (1-\beta)s_{j+1})$ otherwise
- if $i = j$, then
  1. if $\alpha = \beta$ then $(\alpha s_i, (1-\alpha)s_{i+1}), (\beta s_j, (1-\beta)s_{j+1})$ represents the same information
  2. if $\alpha < \beta$ then $(\alpha s_i, (1-\alpha)s_{i+1}) < (\beta s_j, (1-\beta)s_{j+1})$
  3. if $\alpha > \beta$ then $(\alpha s_i, (1-\alpha)s_{i+1}) > (\beta s_j, (1-\beta)s_{j+1})$

2. Negation operator of a Proportional 2-Tuple

The negation for proportional 2-tuples is defined as:

$$Neg((\alpha s_i, (1-\alpha)s_{i+1})) = ((1-\alpha)s_{g-i-1}, \alpha s_{g-i}), \quad (10)$$

where $g+1$ is the cardinality of S, $S = \{s_0, s_1, \ldots, s_g\}$

3. Aggregation of Proportional 2-Tuple

Wang and Hao defined many aggregation operators to handle processes of CW. The definitions of these aggregation operators are based on canonical characteristic values of linguistic labels. To do so, they used the similar corresponding aggregation operators developed in [9] in order to aggregate ordinal 2-tuples through their position indexes [13].

Example

By using the example presented previously, if we apply the proportional 2-tuples we obtain the following results:

- The arithmetic mean according to Wang and Hao is defined as:

$$\begin{aligned} \overline{x} &= \pi^{-1}(\textstyle\sum_{i=1}^{n} \frac{1}{n}\pi(\alpha s_i, (1-\alpha)s_{i+1})) = \\ &= \pi^{-1}(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(i + (1-\alpha)) \end{aligned} \quad (11)$$

- The preference vector is transformed into proportional 2-tuple as follows:

$(1s_2, 0s_3)$	$(1s_3, 0s_4)$	$(1s_3, 0s_4)$	$(1s_2, 0s_3)$

- The collective preference value obtained under ordinal proportional 2-tuple contexts is

  $\overline{x} = \pi^{-1}(\frac{1}{4}(\pi(1s_2, 0s_3) + \pi(1s_3, 0s_4) + \pi(1s_3, 0s_4) + \pi(1s_2, 0s_3))) = \pi^{-1}(\frac{1}{4}(2+3+3+2)) = \pi^{-1}(2.5) = ((1-0.5)s_2, 0.5s_3) = (0.5s_2, 0.5s_3)$

4 Comparative Analysis

The aim of this contribution is to make a comparative analysis among the symbolic approaches presented in section 3. This analysis will consist of studying all the approaches from different points of view such as: representation model, computations, accuracy, comprehension and so on.

- *The Representation*

  Here, we want to point out that although the authors of the three approaches said that they are based on the fuzzy linguistic approach. It is clear that the *Virtual model* does not follow this approach because its representation does not have any semantics to interpret the linguistic information. Nevertheless, the 2-tuple and the proportional 2-tuple in spite of using additional information to the representation of the linguistic information both, keep the basis of the definition of linguistic variable provided in the fuzzy linguistic approach by using fuzzy numbers to represent the semantics of the linguistic terms.

- *The accuracy*

  By comparing the results provided by the three different approaches in the example showed we can see that the results are similar although the representation of the information would be different. The reason of similar results in this case is because we have considered a linguistic term set symmetrically distributed. However, if we go in deep through the three approaches we can see that the 2-tuple approach guarantees the accuracy when the labels are symmetrically distributed and there is only point with maximum height. The proportional 2-tuple guarantees the accuracy when the terms have the same width in their support. Finally regarding the virtual model is difficult to say anything about accuracy because it does not have semantics to represent the information.

- *The computations*

  By comparing the computational models provided by the three approaches we can observe an important difference. The Virtual model proposes many symbolic operations directly on the linguistic terms obtaining results without any meaning. However the 2-tuple and proportional 2-tuple approaches propose their symbolic computational models with symbolic operators and additionally provide transformation functions to facilitate the computations.

- *The comprehension*

  An important point that should be taken into account regarding the linguistic information is, that this type of information is not only used to represent the information but also to facilitate the comprehension of the results to the users in many different problems.

  Therefore, we can say that the virtual model is only valid for ranking issues because due to the fact the virtual terms do not have semantics are hard to understand apart of a simple order. The proportional 2-tuple provides a clear representation but comparing with the 2-tuple is a little bit more complex due to the fact of the use of four values to represent a single one.

In the table 1 we show the main differences among models.

ISBN: 978-989-95079-6-8

Table 1: Comparative analysis among models

	2-Tuple	Virtual Linguistic	Proportional 2-Tuple
Representation	symbolic and semantics	No semantics	symbolic and semantics
Accuracy	equidistant labels	always because no semantics	same width
Computation	symbolic	no symbolic	symbolic
Comprehension	easy to understand	only ranking issues	understandable

5 Concluding Remarks

In this paper we have reviewed the classical computational models and we have made a brief review of recent symbolic computational models, like the 2-tuple model, the proportional 2-tuple model and the virtual linguistic model.

We have also made a comparative analysis among them in which the most remarkable finding obtained is that the virtual model can not be considered a linguistic model in the sense of the fuzzy linguistic approach due to the fact that it does not require semantics to their linguistic terms. In fact, it could be considered a crisp approach because it does not use any real element to represent the vagueness of the qualitative information.

Acknowledgements

This work is partially supported by the Research Project TIN-2006-02121 and FEDER funds.

References

[1] P.P. Bonissone. *A fuzzy sets based linguistic approach: Theory and applications*, pages 329–339. Approximate Reasoning in Decision Analysis, North-Holland, 1982.

[2] P.P. Bonissone and K.S. Decker. *Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-Off Precision and Complexity*. In L.H. Kanal and J.F. Lemmer, Editors., Uncertainty in Artificial Intelligence. North-Holland, 1986.

[3] G. Buyukozkan, D. Ruan, and O. Feyzioglu. Evaluating e-learning web site quality in a fuzzy environment. *International Journal of Intelligent System*, 22:567–586, 2007.

[4] R. Degani and G. Bortolan. The problem of linguistic approximation in clinical decision making. *International Journal of Approximate Reasoning*, 2:143–162, 1988.

[5] M. Delgado, J.L. Verdegay, and M.A Vila. Linguistic decision making models. *International Journal of Intelligent Systems*, 7:479–492, 1992.

[6] M. Delgado, J.L. Verdegay, and M.A Vila. On aggregation operations of linguistic labels. *International Journal of Intelligent Systems*, 8:351–370, 1993.

[7] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Kluwer Academic, New York, 1980.

[8] F. Herrera, E. Herrera-Viedma, and J.L. Verdegay. A sequential selection process in group decision making with linguistic assessment. *Information Sciences*, 85:223–239, 1995.

[9] F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.

[10] F. Herrera and L. Martínez. The 2-tuple linguistic computational model. Advantages of its linguistic description, accuracy and consistency. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(Suppl.):33–49, 2001.

[11] L. Martínez, M.J. Barranco, L.G. Perez, and et al. A knowledge based recommender system with multigranular linguistic information. In *Proceedings of the international conference on intelligent systems and knowledge engineering (ISKE 2007)*, pages 312–319, Chengdu, China, 2007.

[12] J.M. Merigo and A.A. Gil-Lafuente. The linguistic generalized owa operator and its application instrategic decision making. In *Proceedings of the tenth international conference on enterprise information systems (ICEIS 2008)*, pages 219–224, Barcelona, Spain, 2008.

[13] J. Wang and J. Hao. A new version of 2-tuple fuzzy linguistic representation model for computing with words. *IEEE transactions on fuzzy systems*, 14:435, 2006.

[14] Y.J. Xu and L. Huang. An approach to group decision making problems based on 2-tuple linguistic aggregation operators. In *International Colloquium on Computing, Communication, Control, and Management*, pages 73–77, Guangzhou, China, 2008.

[15] Z.S Xu. A method based on linguistic aggregation operators for group decision making with linguistic preference relations. *Information Science*, 166:19–30, 2004.

[16] R.R. Yager. *Fuzzy Logic: State of the Art*, chapter Fuzzy screening systems, pages 251–261. Kluwer Academic Publishers, 1993.

[17] R.R. Yager. An approach to ordinal decision making. *International Journal of Approximate Reasoning*, 12:237–261, 1995.

[18] L.A. Zadeh. The concept of a linguistic variable and its applications to approximate reasoning. *Information Sciences, Part I, II, III*, 8,8,9:199–249,301–357,43–80, 1975.

ISBN: 978-989-95079-6-8

Universal Integrals Based on Level Dependent Capacities

Erich Peter Klement[1] Anna Kolesárová[2] Radko Mesiar[3,4] Andrea Stupňanová[3]

1. Department of Knowledge-Based Mathematical Systems, Johannes Kepler University
Linz, Austria
2. Institute IAM, Faculty of Chemical and Food Technology, Slovak University of Technology
Bratislava, Slovakia
3. Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology
Bratislava, Slovakia
4. Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: ep.klement@jku.at, anna.kolesarova@stuba.sk, mesiar@math.sk, andy@math.sk

Abstract— *Three different types of universal integral based on level dependent capacities are introduced and discussed. Two extremal types are based on Caratheodory's idea of inner and outer measures, while the third type is introduced for copula-based universal integrals only.*

Keywords— Choquet integral, Copula, Fuzzy measure, Level dependent capacities, Sugeno integral, Universal integral.

1 Introduction

Capacities (also called fuzzy measures) express the weights of (measurable) subsets of a given universe X in a consistent way. If, for example, X represents a set of criteria then a capacity m on X assigns to each group of criteria $A \subseteq X$ a weight $m(A)$. Universal integrals aggregate the information contained in a capacity m and in a (measurable) function f to a single representative value. The range of f has no influence on the capacity we are exploiting in the aggregation process. In practical applications, however, there is often a need for a different treatment of functions with small values and functions with large values (or even in a more sophisticated way).

This idea of different weights for sets (groups of criteria) at different levels can be expressed by means of a system of capacities (level dependent capacities, see [1, 2, 3]). Note that, motivated by multi-criteria decision problems, one approach to a Choquet integral based on level dependent capacities was proposed and discussed in [1, 2]. On the other hand, one type of a Sugeno integral based on level dependent capacities was introduced in [3] as a solution of the comonotone maxitivite problem for aggregation functions.

The aim of this contribution is to introduce and to discuss universal integrals based on level dependent capacities, i.e., we are looking for extensions of the concept of universal integrals [4] which originally was defined for capacities only. Similar ideas generalizing classical measures are related to the notion of Markov kernels [5], see also [6].

2 Preliminaries

Recall that a *(binary) aggregation operator* or *aggregation function* [7, 8] is a function $A\colon [0,1]^2 \to [0,1]$ which is non-decreasing (in each component) and satisfies $A(0,0) = 0$ and $A(1,1) = 1$. If a binary aggregation function A has neutral element 1, i.e., satisfies $A(a,1) = A(1,a) = a$ for all $a \in [0,1]$, it is called a *semicopula* [9].

Universal integrals were introduced and studied recently in [10, 4]. If the set $\mathcal{H}$ is defined by

$$\mathcal{H} = \{h\colon [0,1] \to [0,1] \mid h \text{ is non-increasing with } h(0) = 1\}.$$

and if $\otimes\colon [0,1]^2 \to [0,1]$ is a semicopula then a non-decreasing functional $J\colon \mathcal{H} \to [0,1]$ is called $\otimes$-*consistent* if, for all $a, b \in [0,1]$, we have

$$J(b \cdot \mathbf{1}_{[0,a]} + (1-b) \cdot \mathbf{1}_{\{0\}}) = a \otimes b. \tag{1}$$

A *capacity space* $(X, \mathcal{A}, m)$ is a triplet consisting of a non-empty universe X, a σ-algebra $\mathcal{A}$ of subsets of X (the elements of $\mathcal{A}$ are called *measurable subsets of* X), and a capacity $m\colon \mathcal{A} \to [0,1]$, i.e., m is isotone with boundary conditions $m(\emptyset) = 0$ and $m(X) = 1$. Furthermore, denote by $\mathcal{F}$ the system of all $\mathcal{A}$-measurable functions $f\colon X \to [0,1]$. Note that, in our context, a function $f\colon X \to [0,1]$ is called $\mathcal{A}$-*measurable* if, for each Borel subset B of $[0,1]$, its preimage $f^{-1}(B) = \{x \in X \mid f(x) \in B\}$ is a measurable subset of X, i.e., belongs to the σ-algebra $\mathcal{A}$. For details about capacity spaces (also called *fuzzy measure spaces*) see [11, 12, 13, 14].

Definition 2.1 Let $J\colon \mathcal{H} \to [0,1]$ be a $\otimes$-consistent functional. A *J-universal integral* $\mathbf{I}_J$ is a mapping which can be defined for each capacity space $(X, \mathcal{A}, m)$ via $\mathbf{I}_{J,m}\colon \mathcal{F} \to [0,1]$ given by

$$\mathbf{I}_{J,m}(f) = J(h_{m,f}), \tag{2}$$

where the function $h_{m,f}\colon [0,1] \to [0,1]$ is defined by $h_{m,f}(t) = m(\{f \geq t\})$ (observe that, because of the monotonicity of m, we have $h_{m,f} \in \mathcal{H}$).

Note that two prototypical universal integrals are the *Choquet integral* [15, 16, 17], in which case we have $J = R$, the standard Riemann integral given by

$$R(h) = \int_0^1 h(t)\, dt \tag{3}$$

(here $\otimes = \Pi$, the standard product), and the *Sugeno integral* [18, 11, 12, 17] which is related to $J = S$ with

$$S(h) = \sup\{\min(t, h(t)) \mid t \in [0,1]\} \tag{4}$$

ISBN: 978-989-95079-6-8

(here $\otimes = \mathrm{Min}$, the greatest semicopula).

Next, we recall two important classes of universal integrals (for more details see [10, 4]):

Proposition 2.2 *Let $\otimes$ be a semicopula. The smallest universal integral $\mathbf{I}_\otimes$ based on $\otimes$ is given by $J_\otimes\colon \mathcal{H} \to [0,1]$,*

$$J_\otimes(h) = \sup\{t \otimes h(t) \mid t \in [0,1]\}, \tag{5}$$

i.e., $\mathbf{I}_\otimes = \mathbf{I}_{J_\otimes}$.

Evidently, because of (5), we have

$$\mathbf{I}_{\otimes,m}(f) = \sup\{t \otimes m(\{f \geq t\}) \mid t \in [0,1]\}. \tag{6}$$

Observe that for $\otimes = \mathrm{Min}$, $\mathbf{I}_{\mathrm{Min}} = \mathbf{I}_S$ is just the Sugeno integral, while $\mathbf{I}_\Pi$ is the *Shilkret integral* [19]. If $T\colon [0,1]^2 \to [0,1]$ is a strict t-norm [20], then $\mathbf{I}_T$ is the *Weber integral* [21].

Recall that a *(two-dimensional) copula* [22] is a binary aggregation function $C\colon [0,1]^2 \to [0,1]$ with annihilator 0 and neutral element 1, i.e., satisfies $C(a,0) = C(0,a) = 0$ and $C(a,1) = C(1,a) = a$ for all $a \in [0,1]$, which is also 2-*increasing*, i.e., for all $a_1, a_2, b_1, b_2 \in [0,1]$ with $a_1 \leq a_2$ and $b_1 \leq b_2$ we have

$$C(a_1,b_1) - C(a_1,b_2) + C(a_2,b_2) - C(a_2,b_1) \geq 0. \tag{7}$$

This means that each copula C induces a probability measure P_C on the Borel subsets of $[0,1]^2$ via

$$\begin{aligned} &P_C([a_1,a_2] \times [b_1,b_2]) \\ &\quad = C(a_1,b_1) - C(a_1,b_2) + C(a_2,b_2) - C(a_2,b_1). \end{aligned} \tag{8}$$

Equivalently, a copula C is a semicopula which is *supermodular*, i.e.,

$$\begin{aligned} &C(a_1,b_1) + C(a_2,b_2) \\ &\quad = C((a_1,b_1) \wedge (a_2,b_2)) + C((a_1,b_1) \vee (a_2,b_2)) \end{aligned} \tag{9}$$

for all $(a_1,b_1),(a_2,b_2) \in [0,1]^2$, where $\wedge$ and $\vee$ are the (pointwise) lattice operations on $[0,1]^2$, i.e, $\wedge = \min$ and $\vee = \max$.

Proposition 2.3 *If $C\colon [0,1]^2 \to [0,1]$ is a copula and P_C the probability measure on the Borel subsets of $[0,1]^2$ induced by C, then the functional $J_C\colon \mathcal{H} \to [0,1]$ given by*

$$J_C(h) = P_C(\{(x,y) \in [0,1]^2 \mid y \leq h(x)\}) \tag{10}$$

is C-consistent.

Given a copula $C\colon [0,1]^2 \to [0,1]$, we shall denote the universal integral $\mathbf{I}_{J_C}$ simply by $\mathbf{I}_{(C)}$. Since Π and Min are copulas, we see that $\mathbf{I}_{(\Pi)} = \mathbf{I}_R$ is the Choquet integral and $\mathbf{I}_{(\mathrm{Min})} = \mathbf{I}_S$ is the Sugeno integral. Therefore, we have two different ways to define the Sugeno integral.

3 Level dependent capacities

The notion of level dependent capacities was introduced in [1], see also [2].

Definition 3.1 Let $(X,\mathcal{A})$ be a measurable space. A *level dependent capacity* on $(X,\mathcal{A})$ is a system $M = (m_t)_{t\in[0,1]}$, where each $m_t\colon \mathcal{A} \to [0,1]$ is a capacity on $(X,\mathcal{A})$.

A special example of a level dependent capacity is a *Markov kernel* [5, 6], where each m_t is a probability measure on $(X,\mathcal{A})$ and, for each $A \in \mathcal{A}$, the function $M_A\colon [0,1] \to [0,1]$ given by $M_A(t) = m_t(A)$ is $\mathcal{A}$-measurable.

Given a level dependent capacity $M = (m_t)_{t\in[0,1]}$ on $(X,\mathcal{A})$, for each $\mathcal{A}$-measurable function $f\colon X \to [0,1]$ we define the function $h_{M,f}\colon [0,1] \to [0,1]$, which accumulates all the information contained in M and f, by

$$h_{M,f}(t) = m_t(\{f \geq t\}). \tag{11}$$

In general, the function $h_{M,f}$ is neither monotone nor even $\mathcal{A}$-measurable (compare with the function $h_{m,f} \in \mathcal{H}$ discussed in Section 2). Following the ideas of inner and outer measures in Caratheodory's approach [23], we introduce the two functions $(h_{M,f})^*, (h_{M,f})_*\colon [0,1] \to [0,1]$ by

$$(h_{M,f})^* = \inf\{h \in \mathcal{H} \mid h \geq h_{M,f}\}, \tag{12}$$

$$(h_{M,f})_* = \sup\{h \in \mathcal{H} \mid h \leq h_{M,f}\}. \tag{13}$$

Note that it is possible to show that for all $t \in [0,1]$

$$(h_{M,f})^*(t) = \sup\{h_{M,f}(u) \mid u \in [t,1]\}, \tag{14}$$

$$(h_{M,f})_*(t) = \inf\{h_{M,f}(u) \mid u \in [0,t]\}. \tag{15}$$

Evidently, both functions $(h_{M,f})^*$ and$(h_{M,f})_*$ are non-increasing and, therefore, belong to $\mathcal{H}$. If the level dependent capacity $M = (m_t)_{t\in[0,1]}$ is constant (i.e., $m_t = m$ for all $t \in [0,1]$) then the three functions considered in (11)–(13) coincide, i.e., we have

$$h_{m,f} = h_{M,f} = (h_{M,f})^* = (h_{M,f})_*. \tag{16}$$

The three functions $h_{M,f}$, $(h_{M,f})^*$ and $(h_{M,f})_*$ allow us to introduce three different extensions of universal integrals for level dependent capacities.

4 Extensions of universal integrals

If $\mathbf{I}_J$ is a J-universal integral then a mapping $\widetilde{\mathbf{I}}_J$ which can be defined on arbitrary measurable spaces $(X,\mathcal{A})$ and arbitrary level dependent capacities M on $(X,\mathcal{A})$ is called an *extension of* $\mathbf{I}_J$ if, whenever the level dependent capacity $M = (m_t)_{t\in[0,1]}$ is constant (i.e., $m_t = m$ for all $t \in [0,1]$), we have $\widetilde{\mathbf{I}}_{J,M} = \mathbf{I}_{J,m}$.

Using the functions in (12) and (13), two extremal extensions of universal integrals can be introduced.

Definition 4.1 Let $\mathbf{I}_J$ be a J-universal integral. The *upper extension* $(\mathbf{I}_J)^*$ (respectively the *lower extension* $(\mathbf{I}_J)_*$) of $\mathbf{I}_J$ for an arbitrary measurable space $(X,\mathcal{A})$, a level dependent capacity M on $(X,\mathcal{A})$, and an $\mathcal{A}$-measurable function $f \in \mathcal{F}$ are given by, respectively,

$$(\mathbf{I}_{J,M})^*(f) = J((h_{M,f})^*), \tag{17}$$

$$(\mathbf{I}_{J,M})_*(f) = J((h_{M,f})_*). \tag{18}$$

If the system $M = (m_t)_{t\in[0,1]}$ is constant (i.e., $m_t = m$ for all $t \in [0,1]$) then we clearly have

$$\mathbf{I}_{J,m} = (\mathbf{I}_{J,M})_* = (\mathbf{I}_{J,M})^*, \tag{19}$$

i.e., both $(\mathbf{I}_J)_*$ and $(\mathbf{I}_J)^*$ are indeed extensions of $\mathbf{I}_J$.

ISBN: 978-989-95079-6-8

For extensions $\widetilde{\mathbf{I}}_{J,M}$ of $\mathbf{I}_J$ which are non-decreasing in M we get

$$(\mathbf{I}_{J,M})_* \leq \widetilde{\mathbf{I}}_{J,M} \leq (\mathbf{I}_{J,M})^*. \tag{20}$$

The approach of extending a known integral to more general situations was applied, e.g., in the case of belief and plausibility measures. Indeed, in the case of a belief measure m [14, 16], the standard Lebesque integral (for probability measures) was applied to probability measures $P \geq m$, and the integral with respect to the capacity m (which, in general, is non-additive) was in this case defined as the infimum of all Lebesgue integrals (with respect to probability measures P with $P \geq m$). Similarly, for a plausibility measure m, all probability measures $P \leq m$ were taken into account, and then the supremum over all Lebesgue integrals (with respect to probability measures P with $P \leq m$) yields the integral for the (non-additive) plausibility measure m. Observe that in both cases the resulting integral is the Choquet integral with respect to m. In our case, having a universal integral $\mathbf{I}_J$ defined for any pair (m, f) of a capacity and a measurable function on the same space, we can compare such pairs based on the corresponding $h_{m,f}$: $(m_1, f_1) \leq (m_2, f_2)$ whenever $h_{m_1,f_1} \leq h_{m_2,f_2}$ (here (m_1, f_1) and (m_2, f_2) need not be defined on the same measurable space, in general). Then $(\mathbf{I}_{J,M})^*(f)$ is just the infimum of all values $(\mathbf{I}_{J,m})(g)$, the infimum being taken over all (m, g) with $h_{M,f} \leq h_{m,g}$. Similarly, $(\mathbf{I}_{J,M})_*(f)$ is just the supremum of $(\mathbf{I}_{J,m})(g)$, where the supremum is taken over all (m, g) with $h_{M,f} \geq h_{m,g}$.

For a copula-based universal integral $\mathbf{I}_C$ there is a third extension — however, it cannot necessarily be applied to any measurable function $f \in \mathcal{F}$.

Definition 4.2 Let C be a copula, $(X, \mathcal{A})$ be a measurable space and $M = (m_t)_{t\in[0,1]}$ be a level dependent capacity on $(X, \mathcal{A})$.

(i) A function $f \in \mathcal{F}$ is called *M-integrable* if $h_{M,f}$ is a measurable function.

(ii) For each M-integrable function $f \in \mathcal{F}$ the corresponding C-based universal integral $\widetilde{\mathbf{I}}_{(C)}$ is defined by $\widetilde{\mathbf{I}}_{(C),M}(f) = J_C(h_{M,f})$ (compare (10)), i.e.,

$$\widetilde{\mathbf{I}}_{(C),M}(f) = P_C(\{(u, v) \in [0, 1]^2 \mid v \leq h_{M,f}(u)\}). \tag{21}$$

Note that a similar extension is possible in the case of a universal integral $\mathbf{I}_\mu$ based on a capacity μ on the Borel subsets of $]0, 1[^2$ as proposed in [24], in which case

$$\mathbf{I}_{\mu,M}(f) = \mu(\{(u, v) \in]0, 1[^2 \mid v < h_{M,f}(u)\}). \tag{22}$$

Remark 4.3

(i) The generalization of the Choquet integral for level dependent capacities as proposed in [1] is closely related to the Riemann-integrability of the function $h_{M,f}$. For example, if X is a finite set (and $\mathcal{A} = 2^X$), and if the level dependent capacity $M = (m_t)_{t\in[0,1]}$ has the same measurability property as a Markov kernel, i.e., for each $A \subseteq X$ the function $M_A\colon [0, 1] \to [0, 1]$ given by $M_A(t) = m_t(A)$ is measurable, then also $h_{M,f}$ is measurable for each $f\colon X \to [0, 1]$. Since $h_{M,f}$ is not necessarily monotone, the Riemann integral in the original definition of the Choquet integral should be replaced by the Lebesgue integral (with respect to the standard Lebesgue measure λ on the Borel subsets of $[0, 1]$), i.e., then (21) turns into

$$\widetilde{\mathbf{I}}_{(\Pi),M}(f) = \int_{[0,1]} h_{M,f}\, d\lambda. \tag{23}$$

However, based on Definition 4.1, we have two other extensions of the Choquet integral given by

$$(\mathbf{I}_{(\Pi),M})^*(f) = \int_0^1 (h_{M,f})^*(x)\, dx, \tag{24}$$

$$(\mathbf{I}_{(\Pi),M})_*(f) = \int_0^1 (h_{M,f})_*(x)\, dx. \tag{25}$$

(ii) Similarly, we have three possible extensions of the Sugeno integral, namely,

$$\widetilde{\mathbf{I}}_{(\mathrm{Min}),M}(f) = \lambda(\{t \in [0, 1] \mid t \leq h_{M,f}(t)\}), \tag{26}$$

$$(\mathbf{I}_{(\mathrm{Min}),M})^*(f) = \sup_{t\in[0,1]} \min(t, h_{M,f}(t)), \tag{27}$$

$$(\mathbf{I}_{(\mathrm{Min}),M})_*(f) = \sup_{t\in[0,1]} \min\Big(t, \inf_{u\in[0,t]} h_{M,f}(u)\Big). \tag{28}$$

In [3] comonotone maxitivity of aggregation functions was investigated and, without any reference to integrals, $(\mathbf{I}_{(\mathrm{Min}),M})^*$ was found to be a solution, compare also [25].

Example 4.4 Let $X = [0, 1]$, $\mathcal{A}$ be the σ-algebra of Borel subsets of $[0, 1]$, and define $M = (m_t)_{t\in[0,1]}$ by

$$m_t = \begin{cases} m^* & \text{if } t \in \left[0, \frac{1}{4}\right], \\ \sqrt{\lambda} & \text{if } t \in \left[\frac{1}{2}, 1\right], \\ m_* & \text{otherwise,} \end{cases} \tag{29}$$

where m^* and m_* are the greatest and the smallest capacity on $(X, \mathcal{A})$, respectively, given by

$$m^*(A) = \begin{cases} 0 & \text{if } A = \emptyset, \\ 1 & \text{otherwise,} \end{cases} \tag{30}$$

$$m_*(A) = \begin{cases} 1 & \text{if } A = X, \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

If $f = \mathrm{id}_{[0,1]}$ then we get

$$(h_{M,f})^* = \mathbf{1}_{[0,\frac{1}{4}]} + \tfrac{1}{\sqrt{2}} \cdot \mathbf{1}_{]\frac{1}{4},\frac{1}{2}[} + \sqrt{1 - f} \cdot \mathbf{1}_{[\frac{1}{2},1]}, \tag{32}$$

$$h_{M,f} = \mathbf{1}_{[0,\frac{1}{4}]} + \sqrt{1 - f} \cdot \mathbf{1}_{[\frac{1}{2},1]}, \tag{33}$$

$$(h_{M,f})_* = \mathbf{1}_{[0,\frac{1}{4}]}. \tag{34}$$

Consequently, we have for the corresponding extensions of the Choquet integral

$$(\mathbf{I}_{(\Pi),M})^*(f) \approx 0.663, \tag{35}$$

$$\widetilde{\mathbf{I}}_{(\Pi),M}(f) \approx 0.486, \tag{36}$$

$$(\mathbf{I}_{(\Pi),M})_*(f) = 0.25. \tag{37}$$

ISBN: 978-989-95079-6-8

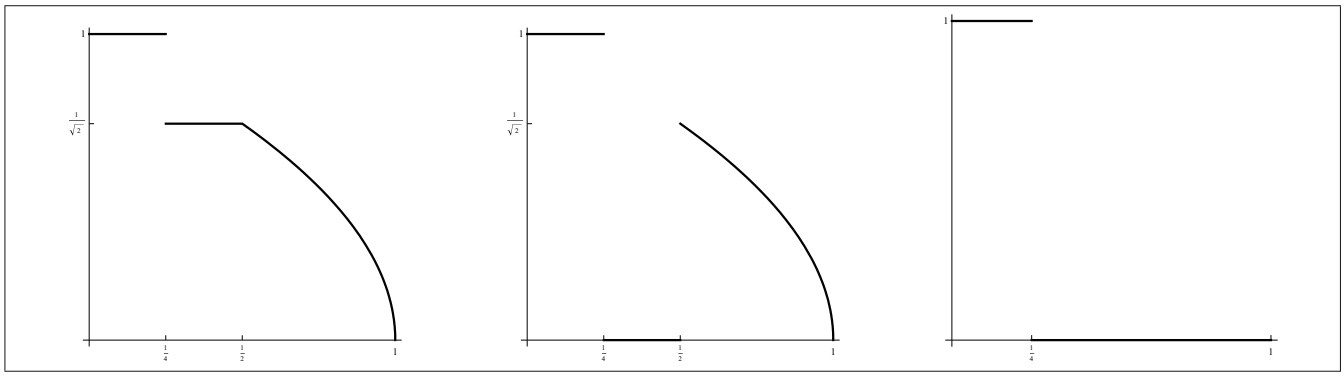


Figure 1: The functions $(h_{M,f})^*$ (left), $h_{M,f}$ (center), and $(h_{M,f})_*$ in Example 4.4

For the extensions of the Sugeno integral we get

$$(\mathbf{I}_{(\mathrm{Min}),M})^*(f) \approx 0.618, \tag{38}$$

$$\widetilde{\mathbf{I}}_{(\mathrm{Min}),M}(f) \approx 0.368, \tag{39}$$

$$(\mathbf{I}_{(\mathrm{Min}),M})_*(f) = 0.25. \tag{40}$$

The Shilkret integral [19] can be considered either as the smallest universal integral based on $\otimes = \Pi$ or as μ-based universal integral, where μ is the capacity on the Borel subsets of $]0,1[^2$ given by $\mu(E) = \sup\{x \cdot y \mid (x,y) \in E\}$. Then

$$(\mathbf{I}_{(\Pi),M})^*(f) \approx 0.385, \tag{41}$$

$$\widetilde{\mathbf{I}}_{\mu,M}(f) \approx 0.385, \tag{42}$$

$$(\mathbf{I}_{(\Pi),M})_*(f) = 0.25. \tag{43}$$

5 Conclusions

Universal integrals based on level dependent capacities can be seen as natural extensions of capacity-based universal integrals acting on different subdomains with (possibly) different capacities. Take, for example, $X = \{1,2\}$ and define the capacities $v_1, v_2 \colon 2^X \to [0,1]$ by $v_1(\{1\}) = \frac{1}{3}$, $v_1(\{2\}) = \frac{2}{3}$, and $v_2(\{1\}) = \frac{3}{4}$, $v_2(\{2\}) = \frac{1}{4}$. Both capacities are additive (i.e., discrete probability measures), and the corresponding Choquet integrals are just weighted arithmetic means, i.e., $W_1(x,y) = \frac{x+2y}{3}$ and $W_2(x,y) = \frac{3x+y}{4}$. Consider the level dependent capacity $M = (m_t)_{t\in[0,1]}$ given by

$$m_t = \begin{cases} v_1 & \text{if } t \le \frac{1}{2}, \\ v_2 & \text{otherwise.} \end{cases} \tag{44}$$

Then, for $(x,y) \in \left[0,\frac{1}{2}\right]^2$, we have

$$(\mathbf{I}_{(\Pi),M})^*(x,y) = \widetilde{\mathbf{I}}_{(\Pi),M}(x,y) = (\mathbf{I}_{(\Pi),M})_*(x,y) = W_1(x,y), \tag{45}$$

and for $(x,y) \in \left[\frac{1}{2},1\right]^2$

$$(\mathbf{I}_{(\Pi),M})^*(x,y) = \widetilde{\mathbf{I}}_{(\Pi),M}(x,y) = (\mathbf{I}_{(\Pi),M})_*(x,y) = W_2(x,y), \tag{46}$$

If $(x,y) \in \left[0,\frac{1}{2}\right] \times \left[\frac{1}{2},1\right]$ then

$$(\mathbf{I}_{(\Pi),M})^*(x,y) = \widetilde{\mathbf{I}}_{(\Pi),M}(x,y) = (\mathbf{I}_{(\Pi),M})_*(x,y) = \tfrac{8x+6y+5}{24}. \tag{47}$$

However, our three extensions of the Choquet integral lead to three different extensions of W_1 (restricted to $\left[0,\frac{1}{2}\right]^2$) and W_2 (restricted to $\left[\frac{1}{2},1\right]^2$): if $(x,y) \in \left]\frac{1}{2},1\right] \times \left[0,\frac{1}{2}\right[$ then

$$\widetilde{\mathbf{I}}_{(\Pi),M}(x,y) = \tfrac{18x+16y-5}{24}, \tag{48}$$

$$(\mathbf{I}_{(\Pi),M})^*(x,y) = W_2(x,y), \tag{49}$$

$$(\mathbf{I}_{(\Pi),M})_*(x,y) = W_1(x,y). \tag{50}$$

Note that $\widetilde{\mathbf{I}}_{(\Pi),M}$ is a continuous aggregation function while $(\mathbf{I}_{(\Pi),M})^*$ and $(\mathbf{I}_{(\Pi),M})_*$ are non-continuous. Finally, observe that $\widetilde{\mathbf{I}}_{(\Pi),M}$ is the ordinal sum extension of the aggregation functions W_1 and W_2 as proposed in [26].

We expect applications of the functionals introduced here in multi-criteria decision making, especially in situations when the weights of the criteria are related to the cardinal values of the score values.

Acknowledgment

The second, third and fourth author gratefully acknowledge the support of the grants VEGA 1/4209/07, VEGA 1/0198/09 and APVV-0012-07. The third author was also supported by the grant MSMVZ6198898701.

References

[1] S. Giove, S. Greco, and B. Matarazzo. Level dependent capacities and the generalized Choquet integral. In D. Dubois, E. P. Klement, and R. Mesiar, editors, *Proceedings 28th Linz Seminar on Fuzzy Set Theory: "Fuzzy Sets, Probability, and Statistics—Gaps and Bridges"*, pages 54–55, Linz (Austria), 2007. Johannes Kepler Universität.

[2] S. Greco, S. Giove, and B. Matarazzo. The Choquet integral with respect to a level dependent capacity. *Fuzzy Sets and Systems*. (in press).

[3] R. Mesiar, A. Mesiarová-Zemánková, and K. Ahmad. Level-dependent Sugeno integral. *IEEE Trans. Fuzzy Systems*. (in press).

[4] E. P. Klement, R. Mesiar, and E. Pap. A universal integral as common frame for Choquet and Sugeno integral. Submitted for publication.

[5] H. Bauer. *Probability Theory and Elements of Measure Theory*. Academic Press, London, 1981.

[6] E. P. Klement. Characterization of finite fuzzy measures using Markoff-kernels. *J. Math. Anal. Appl.*, 75:330–339, 1980.

ISBN: 978-989-95079-6-8

[7] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In Calvo et al. [8], pages 3–104.

[8] T. Calvo, G. Mayor, and R. Mesiar, editors. *Aggregation Operators. New Trends and Applications*, volume 97 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2002.

[9] F. Durante and C. Sempi. Semicopulæ. *Kybernetika (Prague)*, 41:315–328, 2005.

[10] E. P. Klement, R. Mesiar, and E. Pap. A universal integral. In *Proceedings EUSFLAT 2007, Ostrava*, volume I, pages 253–256. 2007.

[11] Z. Wang and G. J. Klir. *Fuzzy Measure Theory*. Plenum Press, New York, 1992.

[12] E. Pap. *Null-Additive Set Functions*. Kluwer Academic Publishers, Dordrecht, 1995.

[13] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy Measures and Integrals. Theory and Applications*, volume 40 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000.

[14] E. Pap, editor. *Handbook of Measure Theory*. Elsevier Science, Amsterdam, 2002.

[15] G. Choquet. Theory of capacities. *Ann. Inst. Fourier (Grenoble)*, 5:131–292, 1953–1954.

[16] D. Denneberg. *Non-additive Measure and Integral*. Kluwer Academic Publishers, Dordrecht, 1994.

[17] P. Benvenuti, R. Mesiar, and D. Vivona. Monotone set functions-based integrals. In Pap [14], chapter 33, pages 1329–1379.

[18] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[19] N. Shilkret. Maxitive measure and integration. *Indag. Math.*, 33:109–116, 1971.

[20] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, volume 8 of *Trends in Logic. Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 2000.

[21] S. Weber. Two integrals and some modified versions: critical remarks. *Fuzzy Sets and Systems*, 20:97–105, 1986.

[22] R. B. Nelsen. *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*. Springer, New York, second edition, 2006.

[23] P. R. Halmos. *Measure Theory*. Van Nostrand Reinhold, New York, 1950.

[24] E. P. Klement, R. Mesiar, and E. Pap. Measure-based aggregation operators. *Fuzzy Sets and Systems*, 142:3–14, 2004.

[25] M. Couceiro and J.-L. Marichal. Representations and characterizations of polynomial functions on chains. arXiv:0811.0309v, 2008.

[26] B. De Baets and R. Mesiar. Ordinal sums of aggregation operators. In B. Bouchon-Meunier, J. Gutiérrez-Ríos, L. Magdalena, and R. R. Yager, editors, *Technologies for Constructing Intelligent Systems*, volume 2: Tools, pages 137–148. Physica-Verlag, Heidelberg, 2002.

ISBN: 978-989-95079-6-8

A Conceptual Framework for Understanding a Fuzzy System

José M. Alonso, Luis Magdalena

European Centre for Soft Computing, Edificio Científico-Tecnológico,
Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain
Email: {jose.alonso,luis.magdalena}@softcomputing.es

Abstract— *The word Interpretability is becoming more and more frequent in the fuzzy literature. It is admitted as the main advantage of fuzzy systems and it should be given a main role in fuzzy modeling. However, although researchers talk a lot about Interpretability, it is even not clear what it really means. Understanding of fuzzy systems is a subjective task which strongly depends on the person (experience, preferences, knowledge, etc.) who makes the assessment. The general context and the specific problem under consideration have a huge influence too. This paper makes a review on works related to Interpretability and presents the proposal of a conceptual framework that can help to understand fuzzy systems. Moreover, it can be considered as a starting point in order to propose a fuzzy index, easily adaptable to the context of each problem as well as to the user quality criteria, for measuring Interpretability.*

Keywords— Fuzzy systems, Interpretability definition and measurement.

1 Introduction

The concept of interpretability appears in many fields (education, medicine, computer science, etc.) under several names like understandability, comprehensibility, intelligibility, transparency, readability, etc. All these terms are usually considered as synonymous what could yield some confusion. However, some authors [1] distinguish between the term "transparency" (readability) referred as an inherent systemic property (related to the view of the model structure as a white-box) and the term "interpretability" (comprehensibility) which has more cognitive aspects because it is always related to human beings, or more specifically to *humanistic systems* (defined by Zadeh as *those systems whose behavior is strongly influenced by human judgment, perception or emotions* [2]). Notice that readability is assumed as a prerequisite for comprehensibility.

Understanding is likely to be one of the most valuable human abilities. Of course, it is related to the human intelligence and the natural language processing capabilities, because human reasoning is mainly supported by language. The most usual way of explaining something to someone is through the use of words, sentences, linguistic expressions, etc. Of course, gestures and symbols are also used as additional communication tools but they only represent other kinds of languages. Unfortunately, knowledge about these kinds of cognitive tasks is still quite reduced. However, let us underline that this work belongs to the field of artificial intelligence and it will focus on analyzing the interpretability of knowledge-based systems, and more specifically of fuzzy rule-based systems (FRBSs). The main goal of this work is to study how comprehensible are such systems from a human point of view, opening a constructive discussion.

The use of linguistic variables [2] to overcome the ineffectiveness of computers in dealing with systems whose behavior is strongly influenced by human judgment, perceptions or emotions was pointed out by Zadeh long time ago: *In order to be able to make significant assertions (...) it may be necessary to abandon the high standards of rigor and precision that we have become conditioned to expect of our mathematical analyses (...) and become more tolerant of approaches which are approximate in nature* [2]. Following the Zadeh's advice if we really want to define a useful index for system modeling, it is necessary to change our mind. Numerical indices should be forgotten and in turn fuzzy indices should be defined, i.e., the focus must be shifted *from computing with numbers to computing with words, from manipulation of measurements to manipulation of perceptions* [3]. In consequence, the right approach to assess interpretability in an effective way consists in proposing a fuzzy index instead of a numerical one. A first attempt was presented in [4] where a hierarchical fuzzy system was used to get an interpretability measure. That proposal opened this way but a lot of work remains to do.

The expressivity of linguistic rules [5] is acknowledged to be quite close to natural language what favors the interpretability because human understanding is made in terms of natural language. That is why it is useful to take into account the experience gained by natural language processing researchers. For instance, the philosopher Paul Grice established the next four conversational maxims [6] which arise from the pragmatics of natural language and they are based on the common sense:

1. *Maxim of Quality*: Do not say what you believe to be false. Do not say anything without adequate evidence.
2. *Maxim of Quantity*: Make your contribution as informative as required for the current purposes of the exchange.
3. *Maxim of Relation*: Be relevant.
4. *Maxim of Manner*: Avoid obscurity of expression. Avoid ambiguity. Be brief. Be orderly.

Keeping the Grice's maxims in mind during the fuzzy modeling process can help to make easier the understanding of FRBSs. The rule base must be coherent avoiding the use of inconsistent rules (*Maxim of Quality*), redundant rules (*Maxim of Quantity*), and ambiguity rules (*Maxim of Manner*). In addition, selecting the most relevant rules (*Maxim of Relation*) will yield more compact and robust systems.

The rest of the paper is structured as follows. Section 2 has a look on definitions of interpretability found in the literature. In addition, it makes a global review on all the aspects that should be taken into account in the interpretability

ISBN: 978-989-95079-6-8

assessment, setting a conceptual framework for characterizing interpretability. Section 3 describes how to combine the main factors included in the proposed framework for measuring interpretability of FRBSs. Finally, section 4 offers some conclusions and points out future works.

2 Understanding a fuzzy system

Authors talk a lot about interpretability but it is not easy to find a formal definition in the literature. Thus, it is necessary to think on the following question: How can interpretability be defined? The first bid to set a formal definition was made by Tarski et al. [7] a very long time ago (in 1953). He formulated a mathematical definition in the context of classical logic, setting the basis for identifying interpretable theories. In short, *assuming T and S are formal theories, T is interpretable in S if and only if there is a way to pass from T to S, assuring that every theorem of T can be translated and proved into S*.

Regarding the fuzzy literature, a similar definition is included as part of the formal framework proposed in [8]. It distinguishes between a formal language *L* (fuzzy logic) used for describing the model under consideration, and a user-oriented language *L'* (usually the natural language) used for explaining the model to the user. If the system is interpretable, the translation from *L* to *L'* should be made by the user with a small effort. In an informal way, people say that a model is interpretable if they are able to describe it easily.

A more formal definition was given by Bodenhofer and Bauer [9]: *Interpretability means possibility to estimate the system's behavior by reading and understanding the rule base only*. Since the rule base understanding strongly depends on the readability of the involved linguistic expressions, the authors focused on analyzing the interpretability at the level of fuzzy partitioning (linguistic variables) from an intuitive and mathematically exact point of view: *The obvious orderings and inclusions of linguistic terms must not be violated by the corresponding fuzzy sets*. As a result, fuzzy partitioning readability was assumed to be a prerequisite to build interpretable FRBSs.

The comprehensibility of a FRBS depends on all its components, i.e., it depends on the knowledge base (KB) transparency but also on the inference mechanism understanding. Previous works [10, 11] have thoroughly analyzed the main factors that influence the KB readability. Also, a complete study on the interpretability constraints most frequently used in fuzzy modeling has been recently published [1].

Fig. 1 describes the main factors to be considered regarding interpretability of FRBSs. It is inspired on the taxonomy of interpretability of fuzzy systems introduced by [12], which is extended adding our own notation and concepts, and also including some of the most significant constraints extracted from [1]. There are two main points of view to be considered when assessing interpretability of FRBSs (*Global description* and *Local explanation*). The global view presents the system as a whole explaining its global behavior and trend. However, the local view focuses on each individual situation, explaining specific behaviors for specific events. For instance, if we had a fuzzy controller for driving a car, the global view would give an idea on the kind of operations it can do (go straight forward, turn on the right/left, speed up, brake, etc.) and even on the driver style (aggressive, sluggish, etc.). On the contrary, the local view would explain each specific manoeuvre.

Additional information is detailed in the following subsections. Pay attention to the fact that both viewpoints could lead to contradictory goals. The first one (*Global description*) prefers rules as compact as possible, while the second one (*Local explanation*) favors the use of complete rules (*The more general rules, the larger the number of rules that can be fired at the same time*).

2.1 Global description (system structure)

In order to assess the simplicity of a FRBS the following assumption is made: *The more compact the KB, the simpler its understanding, i.e., the higher the interpretability*. This reasoning follows the principle of incompatibility formulated by Zadeh [13] in 1973: *As the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance become almost mutually exclusive characteristics. The closer one looks at a real-world problem, the fuzzier becomes its solution*.

The global description of a linguistic FRBS can be analyzed looking at different abstraction levels as illustrated on left part of Fig. 1. First, the lowest level corresponds to the level of individual fuzzy sets. It includes those constraints demanded to build interpretable fuzzy sets, regarding mathematical properties of the membership functions. At the second level, there are several constraints with respect to the combination of several fuzzy sets to form a fuzzy partition. The use of linguistic variables favors the readability, but it is not enough to ensure interpretability. Hence, some linguistic constraints must be superimposed to the fuzzy partition definition to be interpretable. Fortunately, Ruspini defined (in 1969) a special kind of partition called Strong Fuzzy Partition (SFP) [14] that satisfies most demanded semantic constraints (distinguishability, coverage, normality, convexity, etc). In practice, satisfying all constraints is almost impossible and useless because they represent a very restrictive set of conditions that usually yield systems with very small accuracy. Notice that looking for a good accuracy-interpretability trade-off is the most complex task of fuzzy modeling. Especially relevant are some recent successful biomedical applications [15].

Once a set of linguistic terms with their associated semantics has been defined, they can be used to express linguistic propositions. Then, several propositions are combined to form fuzzy rules describing the system behavior. However, in addition to the analysis of each individual rule it is needed to study the combination of several rules, achieving the highest abstraction level. Notice that defining a global semantics previous to the rule definition makes easier the rule understanding. Only if all the rules use the same linguistic terms (defined by the same fuzzy sets) it will be possible to make a rule comparison at the linguistic level. In order to get fully meaningful partitions the right linguistic terms should be selected according to the problem context. Nevertheless, matching linguistic terms and fuzzy sets is not a straightforward task, for instance finding good linguistic terms for fuzzy partitions automatically generated from data is sometimes not feasible.

To sum up, the satisfaction of all constraints enumerated on the left part of Fig. 1 guarantees the interpretability of the FRBS from the structural point of view.

ISBN: 978-989-95079-6-8

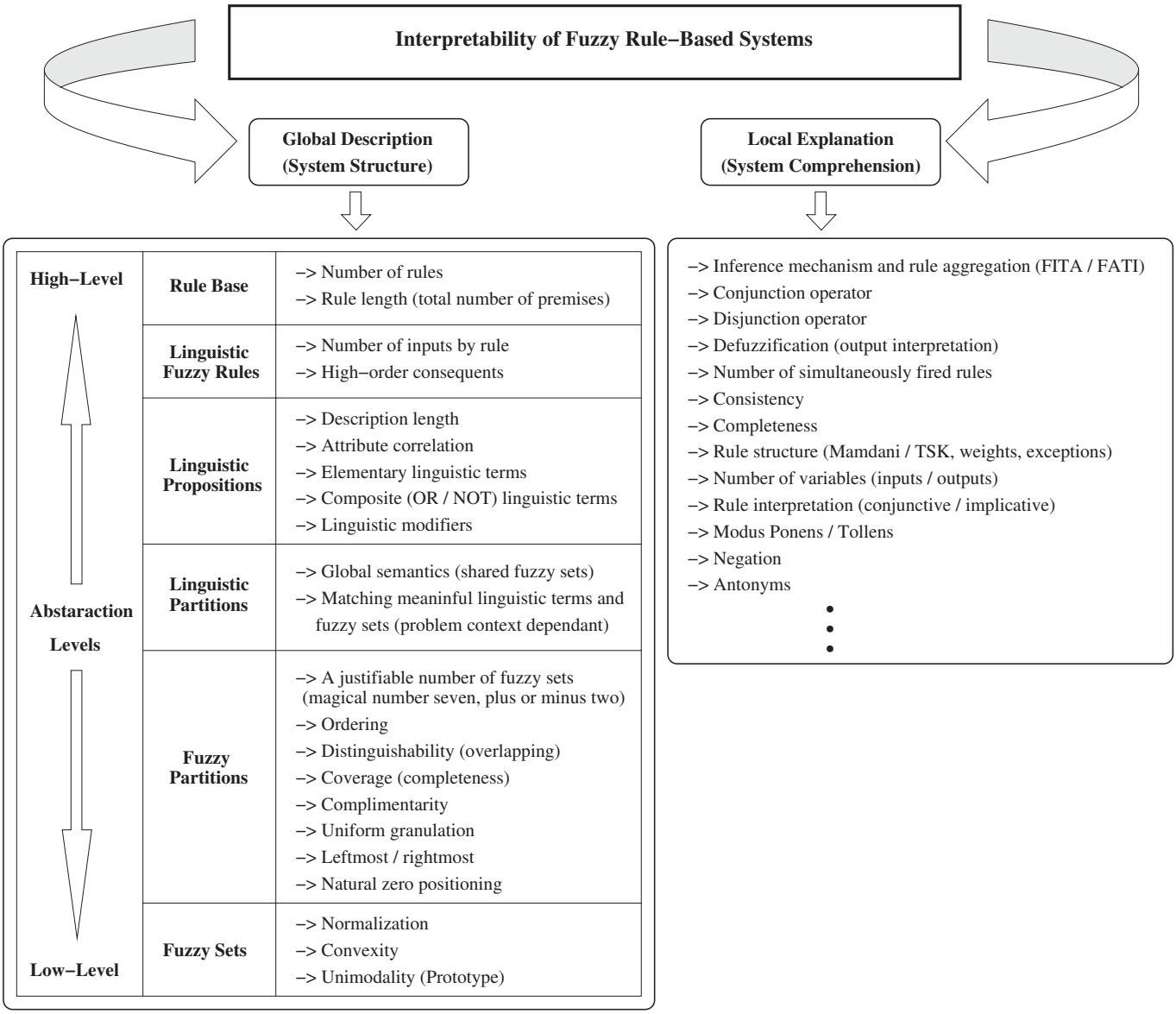


Figure 1: A conceptual framework for characterizing interpretability of FRBSs.

2.2 Local explanation (system comprehension)

Understanding the system behavior from its linguistic description is a very hard task that involves the inference level going beyond the former analysis of the system structure.

In addition, there is a need to give some comments about the inference mechanism implementation distinguishing between FITA (First Infer Then Aggregate) and FATI (First Aggregate Then Infer). It includes the fuzzy operator definitions for conjunction, disjunction, aggregation, and defuzzification. Furthermore, taking into account that as the result of a fuzzy inference several rules can be fired at the same time for a given input vector, the interpretability strongly depends on the number of rules that can be simultaneously fired. The smaller that value, the higher the interpretability. In fact, a model made up of thousand rules (where at maximum ten rules are fired together) may be seen as more interpretable than a model including only one hundred rules (where most of them are simultaneously fired). Notice that the whole rule base should be consistent (not including redundancies, contradictions, etc.) and it should cover most possible situations.

Although Mamdani rules are widely admitted as the more interpretable kind of rules, there are many other rule formats. The second most used rules are the well-known Takagi-Sugeno rules, but there are also rules with exceptions, rules with weights, and so on. The different rule formats can be compared and it is possible to discuss which one is better regarding interpretability from a structural point of view but it is a controversial issue. For instance, for many people the most interpretable rule format is the one they usually work with disregarding its complexity. This proves that many psychological aspects make influence when assessing interpretability. It is a clear example of the "Hammer principle" formulated by Zadeh [16]: *When the only tool you have is a hammer, everything begins to look like a nail*.

Finally, modus Ponens/Tollens must be carefully taken into account. Notice that the fact the all rules are fired at the same time make not easy to establish logical chains of reasoning. It is also necessary to remark that the use of negation and antonyms are quite usual in natural language but their representation using fuzzy logic is still a matter of research.

ISBN: 978-989-95079-6-8

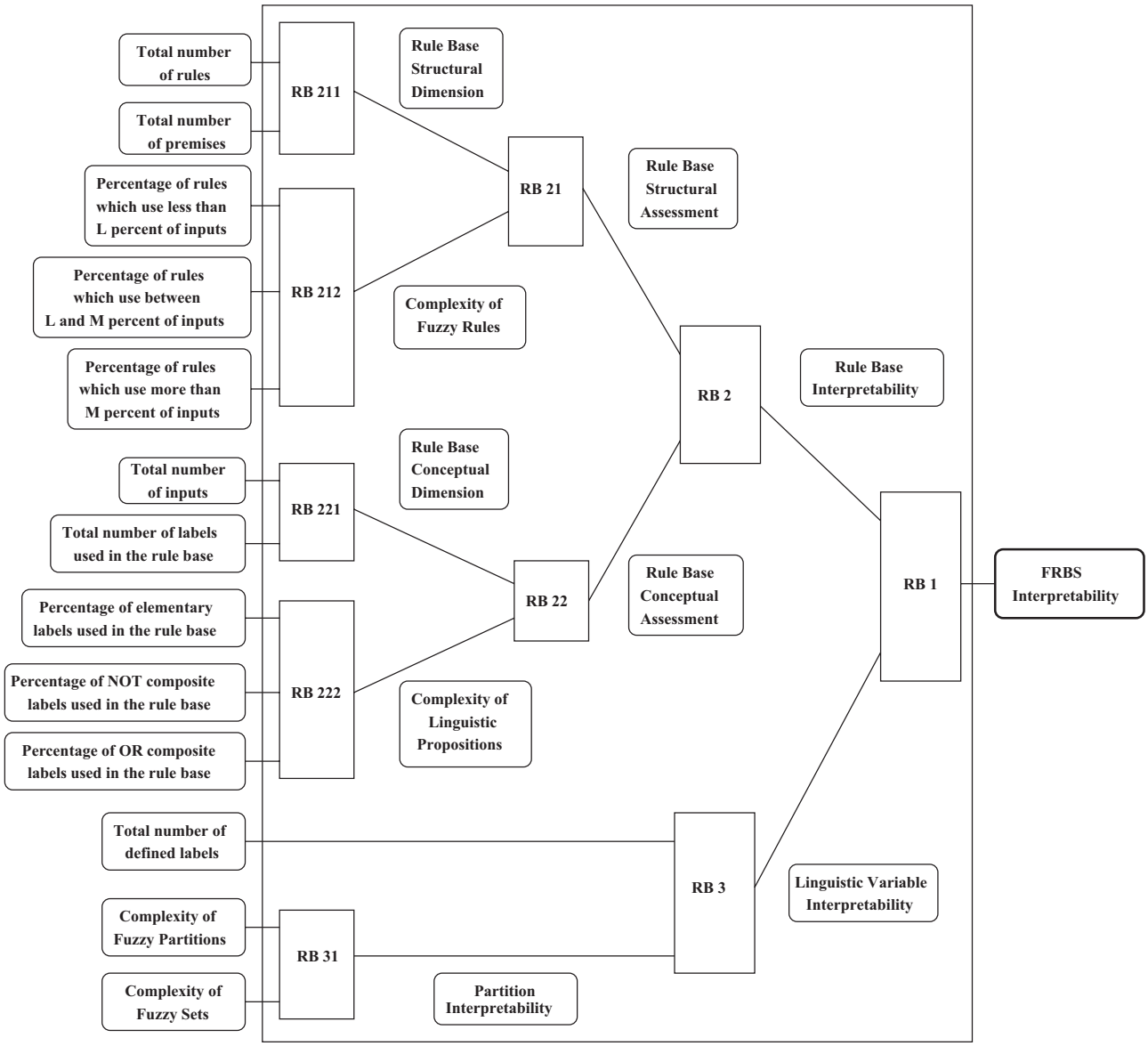


Figure 2: A conceptual framework for assessing interpretability (*Global description*) of FRBSs.

3 Measuring interpretability

Once identified all factors that should be kept in mind regarding interpretability, the definition of a universal interpretability index able to combine all of them becomes a great challenge.

The aim of this section is to introduce a conceptual framework for assessing interpretability. It is represented in the form of a hierarchical diagram in Fig. 2. Of course, this diagram only takes into account interpretability from the structural point of view (*Global description*). The local explanation will be addressed in future works.

To start with, the whole set of factors represented in left part of Fig. 1 has been summarized by a small subset that could be extended in the future. The global diagram can be seen as a flow chart with thirteen inputs (measurable factors to take into account regarding interpretability) and one output (Interpretability measure). The diagram keeps the abstraction levels shown in Fig. 1 (low-level at the bottom of the figure and high-level at the top) and selected inputs are grouped according to the information they convey.

Interpretability of a FRBS is estimated as a combination of two estimators at both low and high level. On the one hand, partition interpretability regards the complexity of each fuzzy set but also the complexity of the whole partition. An estimation of the low-level interpretability (regarding the description of all linguistic variables) is computed adding the number of labels (linguistic terms). Notice that the simple diagram depicted in Fig. 2 does not include anything with respect to the interpretability of linguistic partitions. It is assumed the use of SFPs and global semantics, not entering to the way how linguistic terms are named. On the other hand, high-level interpretability (called *rule base interpretability* in the figure above) involves analysis at the three highest sublevels (linguistic propositions, linguistic fuzzy rules, and rule base).

Drawing an analogy between a set of fuzzy rules and a set of sentences in natural language, the interpretability is assessed regarding both Syntax and Semantics. On the one hand, Syntax can be defined as the *arrangement of words in sentences, clauses, and phrases, and the study of the formation of sentences and the relationship of their component parts*.

ISBN: 978-989-95079-6-8

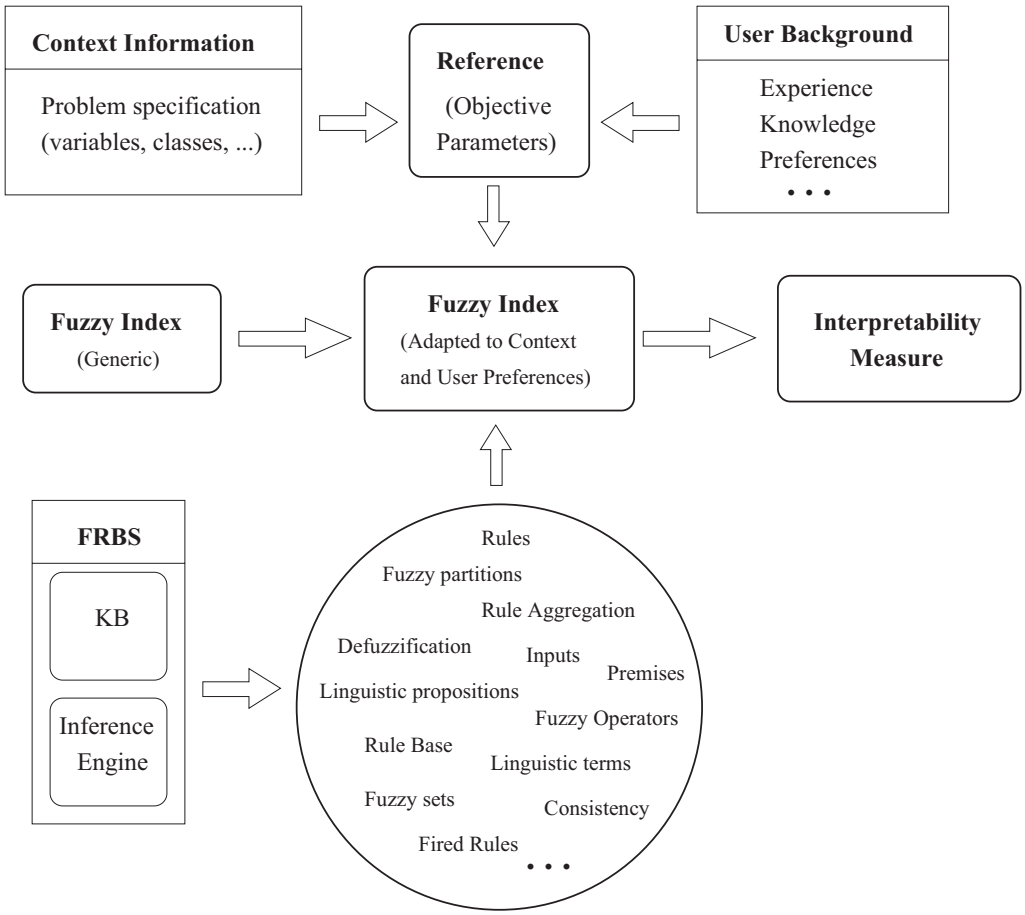


Figure 3: A fuzzy index (adaptable to context problem and user preferences) for measuring Interpretability.

On the other hand, Semantics makes reference to the *Study of meaning*[1]. In our context Syntax is related to dimension and complexity of the rule base (what we have named as *Rule Base Structural Assessment*), computed in a very simplistic way, only considering the total number of rules, premises and premise by rule. It covers both linguistic fuzzy rules and rule base abstraction levels. Semantics in turn takes into account the complexity at the level of linguistic propositions used in the rules (what we have named as *Rule Base Conceptual Assessment*).

The process of measuring something consists in comparing it with a reference (standard unit of measurement) such us a meter for measuring length. However, finding out the suitable reference is not always feasible and the task is especially difficult when measuring non-physical properties. It is widely admitted that interpretability assessment is clearly context dependant. There is not a universal reference; on the contrary the reference will change depending on the problem and depending on the person who makes the assessment. Therefore, the general proposed framework has to be adapted to the specific features of each problem under consideration as well as to the user's background and preferences. For instance, *Total number of rules* may be defined as a linguistic variable made up of five linguistic terms (*Very low*, *Low*, *Medium*, *High*, *Very high*). Nevertheless, the meaning of each linguistic term needs to be defined carefully. *What value should be taken as a prototype for small number of rules (three, ten, one hundred, etc)?* If we were analyzing a FRBS for classification among three kinds of wines, the minimum number of rules should be three, but we need to ask to the people who are going to interact with the system in order to really know how to characterize the linguistic variables. In fact, the perception of interpretability will change depending on the kind of user. The point of view of a system designer who is used to work with fuzzy systems is likely to be very different from the point of view of the domain expert who perfectly knows the problem and how should be the system behavior, but it will be even much more different from the final user who could have only a superficial knowledge of the problem, and who probably has not heard anything about fuzzy logic.

Fig. 3 describes how to build an easily adaptable fuzzy index for assessing interpretability. A generic fuzzy index like the one presented in Fig. 2 has to be tuned and adapted for each problem regarding both the problem definition and the user quality criteria. The system is flexible enough for making an easy adaptation. It consists in defining a reference with all collected information about the problem and the evaluator user (system designer, domain expert, and/or final user). Such reference yields the ranges (universes of discourse) along with the modal points of the fuzzy partitions used to define the input variables (*Total number of rules*, *Total number of premises*, *Percentage of rules which use less than L percent of inputs*, etc.) of the generic fuzzy index. The linked rule bases as well as the intermediate input-output variables could be tuned too.

[1]Both definitions were got from the Encyclopedia Britannica (http://www.britannica.com)

ISBN: 978-989-95079-6-8

4 Final remarks

Previous works has made a great effort to establish the basis for building interpretable fuzzy systems. There are many different works regarding interpretability on the fuzzy literature. Recently, some works have made a global review of the literature putting together contributions of different authors. Following that way, this work has formalized a conceptual framework for characterizing and assessing interpretability of fuzzy systems.

The use of multi-objective approaches is becoming a more and more important topic in fuzzy modeling [17] because of interpretability and accuracy are conflictive goals. In this specific field the interpretability of the model is usually only considered from the point of view of the fuzzy designer. First, it is necessary to make a qualitative and quantitative comparison of all obtained solutions. Then, the best solutions can be selected from a Pareto front regarding the accuracy-interpretability trade-off. It is possible to set a qualitative ranking of solutions based on a comparison per couples, without measuring the interpretability of each individual solution, setting some kind of pre-order is enough. Although there are several accuracy indices, interpretability is measured taking into account only basic parameters what is a strong limitation. Thus, the use of interpretability indices guiding the modeling process could help to achieve better solutions.

Setting qualitative rankings is quite common in the context of semantic web search where retrieved documents have to be ranked before presenting them as answer to a query. For instance, BUDI [18] is a meta-searcher based on fuzzy logic which uses a fuzzy similarity function for comparing documents. It regards the size of the documents, the number of series of words in the same position in both documents, but also the complexity and rarity of words and linguistic propositions. This approach could be extended to the interpretability assessment problem, considering that instead of documents what are going to be compared are the linguistic descriptions of FRBSs.

In the future, experimental analysis must be carried out in order to adapt the theoretical developments to the real worlds. In order to get a universal index adaptable to the user preferences, it is necessary to study how different kinds of user (fuzzy designer, domain expert, and final user) interact with fuzzy systems in a different way and they have different interpretability requirements.

Finally, it is necessary to advance on the paradigm of computing with words and perceptions (CWW/P) [3] which marks an evolution of fuzzy logic, an extension of current theories of fuzzy sets. It is strongly related with meaning, captured by the use of linguistic expressions, words and connectives. As a result, works regarding interpretability assessment can take profit from current research on CWW/P but also they will help to develop new ideas in relation with this new paradigm.

Acknowledgment

This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias) and Spanish government (CICYT) under grant: TIN2008-06890-C02-01.

References

[1] C. Mencar and A. M. Fanelli. Interpretability constraints for fuzzy information granulation. *Information Sciences*, 178(24):4585–4618, 2008.

[2] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Parts I, II, and III. Information Sciences*, 8, 8, 9:199–249, 301–357, 43–80, 1975.

[3] L. A. Zadeh. From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems - I: Fundamental theory and applications*, 45(1):105–119, 1999.

[4] J. M. Alonso, L. Magdalena, and S. Guillaume. HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems*, 23(7):761–794, 2008.

[5] E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Transactions on Computers*, 26(12):1182–1191, 1977.

[6] H. P. Grice. Logic and conversation. *Cole, P. and Morgan, J. (eds.) Syntax and semantics, New York: Academic Press*, 3:43–58, 1975.

[7] A. Tarski, A. Mostowski, and R. Robinson. *Undecidable Theories*. North-Holland, 1953.

[8] C. Mencar, G. Castellano, and A. M. Fanelli. Some fundamental interpretability issues in fuzzy modeling. In *Joint EUSFLAT-LFA*, pages 100–105, Barcelona, Spain, September, 7-9, 2005.

[9] U. Bodenhofer and P. Bauer. A formal model of interpretability of linguistic variables. *In [19]*, pages 524–545, 2003.

[10] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm. *IEEE Transactions on Fuzzy Systems*, 8(5):591–600, 2000.

[11] S. Guillaume and B. Charnomordic. Generating an interpretable family of fuzzy partitions from data. *IEEE Transactions on Fuzzy Systems*, 12 (3):324–335, 2004.

[12] S.-M. Zhou and J. Q. Gan. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. *Fuzzy Sets and Systems*, 159(23):3091–3131, 2008.

[13] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(1):28–44, 1973.

[14] E. H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

[15] P. Fazendeiro, J. V. de Oliveira, and W. Pedrycz. A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller. *IEEE Transactions on Biomedical Engineering*, 54(9):1667–1678, 2007.

[16] L. A. Zadeh. *Applied Soft Computing - Foreword*, 1(1):1–2, 2001.

[17] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4–31, 2007.

[18] J. Serrano-Guerrero et al. BUDI: Architecture for fuzzy search in documental repositories. *Mathware and Soft Computing*, 16(1):71–85, 2009.

[19] J. Casillas et al. *Interpretability issues in fuzzy modeling*, volume 128. Studies in Fuzziness and Soft Computing, Springer-Verlag, Heidelberg, 2003.

ISBN: 978-989-95079-6-8

An Experimental Study on the Interpretability of Fuzzy Systems

José M. Alonso, Luis Magdalena

European Centre for Soft Computing, Edificio Científico-Tecnológico,
Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain
Email: {jose.alonso,luis.magdalena}@softcomputing.es

Abstract— *Interpretability is one of the most significant properties of Fuzzy Systems which are widely acknowledged as gray boxes against other Soft Computing techniques such as Neural Networks usually regarded as black boxes. It is essential for applications with high human interaction (decision support systems in medicine, economics, etc). The use of accuracy indices to guide the fuzzy modeling process is broadly extended. In turn, although there have been a few attempts to define Interpretability indices, we are still far away from having a universal index. With the aim of evaluating the most used indices an experimental analysis (in the form of a web poll) was carried out yielding some useful clues to keep in mind regarding Interpretability assessment. Results extracted from the poll show the inherent subjectivity of the measure because we collected a huge diversity of answers. Nevertheless, comparing carefully all the answers, it was possible to find out some interesting user profiles.*

Keywords— Fuzzy modeling, Interpretability assessment.

1 Introduction

Fuzzy modeling (FM), i.e., system modeling with fuzzy rule-based systems (FRBSs), is an important and active research line inside the fuzzy community. Fuzzy Logic (FL) was introduced by Zadeh [1] (in 1965) and its semantic expressivity, using linguistic variables [2] and linguistic rules [3], favors the interpretability of the modeled system (at least from the structural transparency viewpoint) because it is quite close to expert natural language. From 1965 to 1990, fuzzy designers focus on modeling highly interpretable systems, mainly working with expert knowledge and a few simple linguistic rules. Then, researchers realized that to deal with complex systems expert knowledge was not enough. Thus, from 1990 to 2000, the main effort was made regarding the accuracy of the final model, building complicated fuzzy rules with high accuracy (applying machine learning techniques to extract knowledge from data) but disregarding the model interpretability because automatically generated rules are rarely as readable as desired. Nowadays, a new challenge lies in looking for compact and robust systems with a good accuracy-interpretability trade-off.

Regarding the interpretability assessment of FRBSs, their comprehensibility depends on all their components, i.e., it depends on the knowledge base (KB) transparency but also on the inference mechanism understanding. There are also some crucial psychological factors; for instance, for some people the most interpretable models are the ones they are used to work with, disregarding the model complexity. Anyhow, previous works [4, 5] have thoroughly analyzed the main factors (rule base and fuzzy partitioning) that influence the KB readability. Also, a complete study on the interpretability constraints most frequently used in literature has been recently published [6]. However, once identified all such elements, the current challenge lies in how to combine them in order to obtain a good index. Unfortunately, only a few works have dealt with this issue. As explained by [7] it is possible to distinguish two main interpretability levels: Low-level or fuzzy set level, and high-level or fuzzy rule level. There are some works regarding interpretability measurement at low-level [8, 9, 10, 11, 12] which consist of mathematical formulas to evaluate the main partition properties such as distinguishability, similarity, coverage, overlapping, etc. These indices are usually used to preserve the interpretability of fuzzy systems automatically generated from data. They are also used as part of tuning processes devoted to increase the accuracy of the final model while keeping good interpretability. Furthermore, there are some simple indices, mainly applied to multi-objective fuzzy genetics-based machine learning, regarding the rule base interpretability [13]: (1) *Number of rules*; (2) *Total rule length* (addition of the number of premises defined in all the rules); (3) *Average rule length* (total rule length divided by the number of rules).

However, only a few researchers have tackled with the challenge of defining an index covering both low and high levels. The first one was the *Nauck's index* [14], a numerical index designed (in 2003) to evaluate fuzzy rule-based classification systems, and computed as the product of three terms: $I_{Nauck} = Comp \times Part \times Cov$. *Comp* represents the complexity of a classifier measured as the number of classes divided by the total number of premises. *Part* stands for the average normalized partition index overall input variables. It is computed as the inverse of the number of labels minus one (two is the minimum number of linguistic terms in a partition) for each input variable. Finally, *Cov* is the average normalized coverage degree of the fuzzy partition. It is equal to one for strong fuzzy partitions (SFPs).

A second global index was defined in 2006 [15] and improved in 2008 [16]. It consists of a *fuzzy index* which was initially inspired on the Nauck's index. Six variables (total number of rules, total number of premises, number of rules which use one input, number of rules which use two inputs, number of rules which use three or more inputs, and total number of labels defined by input) are considered as inputs of a fuzzy system and they are grouped, according to the information they convey. In consequence, the interpretability index is computed as the result of inference of a hierarchical fuzzy system made up of four linked KBs. It is specially designed for the context of classification problems solved by means of a specific kind of FRBSs generated following the HILK (Highly Interpretable Linguistic Knowledge) [16] methodology, and assuming SFPs for all system variables.

The rest of the paper is structured as follows. Section 2

ISBN: 978-989-95079-6-8

Table 1: Comparison of interpretability indices (measures).

	Method	Number of rules	Total rule length	Average rule length	Nauck's index	Fuzzy index
KB1	CL-FDT-DS-FDT	20	43	2.15	0.0174	0.452
KB2	CL-FDT-DS-FDT-S	5	9	1.8	0.1667	0.92
KB3	CL-FDT-DS-WM	53	643	12.132	0.0011	0.144
KB4	CL-FDT-DS-WM-S	8	16	2	0.1484	0.839
KB5	CL-WM-DS-FDT	21	49	2.333	0.0153	0.444
KB6	CL-WM-DS-FDT-S	5	10	2	0.2625	0.919
KB7	CL-WM-DS-WM	46	545	11.848	0.0013	0.192
KB8	CL-WM-DS-WM-S	3	6	2	0.3056	0.924
KB9	FDT-S	8	19	2.375	0.0763	0.814
KB10	WM-S	6	18	3	0.0873	0.742
KB11	FDT-P	32	94	2.937	0.0079	0.392
KB12	FDT-P-S	6	15	2.5	0.1714	0.837

Table 2: Comparison of interpretability indices (ranking).

Index	+ Interpretability -
Number of rules	KB8, KB2/KB6, KB10/KB12, KB4/KB9, KB1, KB5, KB11, KB7, KB3
Total rule length	KB8, KB2, KB6, KB12, KB4, KB10, KB9, KB1, KB5, KB11, KB7, KB3
Average rule length	KB2, KB4/KB6/KB8, KB1, KB5, KB9, KB12, KB11, KB10, KB7, KB3
Nauck's index	KB8, KB6, KB12, KB2, KB4, KB10, KB9, KB1, KB5, KB11, KB7, KB3
Fuzzy index	KB8, KB2, KB6, KB4, KB12, KB9, KB10, KB1, KB5, KB11, KB7, KB3

makes a comparison of several interpretability indices in an experimental analysis. Results extracted from a web poll show clearly the intrinsic subjectivity of the measure. As it is explained in section 3, although we got a huge diversity of answers, at first glance completely different, after looking carefully it was possible to find out some interesting user profiles. Finally, section 4 offers some conclusions and points out future works.

2 Experimental analysis

With the aim of making a fair (qualitative and quantitative) comparison of the five indices remarked in the previous section (Number of rules, Total rule length, Average rule length, Nauck's index, and fuzzy index) this experimental study deals with the well known benchmark classification problem called WINE which data set is freely available at the UCI[1] (University of California, Irvine) machine-learning repository. It contains 178 instances coming from results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

Following the HILK modeling methodology [16] twelve KBs, of several sizes, have been generated for the WINE recognition problem. Looking for maximizing the interpretability of final KBs, a global semantics (based on the use of SFPs) is defined previously to rule definition. As a result, for each KB all the rules use the same linguistic terms (defined by the same fuzzy sets), and rule comparison can be done at the linguistic level. The original data set was randomly divided into two subsets taking 50% of data for training and the remaining part for test. HILK copes with different rule induction techniques in order to get enough diversity. Second column in Table 1 contains the abbreviations of the combined methods. CL means clustering previous to rule induction, WM represents the well-known Wang and Mendel's algorithm [17], FDT stands for the popular Fuzzy Decision Tree algorithm [18], DS is data selection in training set previous to rule induction, P means pruning of the tree, and S stands for simplification procedure. All selected algorithms are implemented in KBCT [19], a free software tools for designing fuzzy systems. WM and FDT implementations differ from the original ones in the fuzzy partition design step. Interpretable fuzzy partitions are defined previous to rule induction (in our study five labels per variable were initially defined). Details about the induction algorithms are explained in the cited literature.

The five interpretability indices have been compared from both quantitative (Table 1) and qualitative (Table 2) viewpoints. Table 1 includes results of computing the five selected interpretability indices for the twelve generated KBs. The comparison of the obtained values lets us set rankings from the interpretability point of view. Notice that KBs with equivalent interpretability are set at the same level separated by symbol "/" (see Table 2). From a quantitative point of view the ranges of values are completely different among all the indices. Each individual value only makes sense in comparison with the other values obtained by the same index. From a qualitative point of view we would rather choose those indices yielding a ranking without ambiguities (Total rule length, Nauck's index, and Fuzzy index), i.e., those indices able to produce a full order distinguishing among all pairs of KBs. As expected, we have achieved five different rankings because each interpretability index follows different criteria. Nevertheless, looking carefully it is easy to appreciate that all rankings are somehow similar, the same KBs usually appear at the beginning (KB2, KB4, KB6, KB8) or at the end (KB3, KB7).

However, a last question still remains to be answered: How to know which index is the best one? Since the measure of interpretability is clearly subjective the only way to answer this question is asking people. For that reason, a web poll was addressed to FL experts (50%) as well as people who are not familiar with FL (50%). The study is made regarding the

[1] http://www.ics.uci.edu/-mlearn/MLSummary.html

ISBN: 978-989-95079-6-8

twelve KBs generated for the WINE problem. The goal is to compare the most popular interpretability criteria, including people used (and not used) to work with fuzzy systems that can be (or not) fond of wines. Since interpretability extremely depends on the kind of user, let us add a short comment. In the context of FM, there are three kinds of users:

- The *final user* of the modeled system. In most cases, he/she will interact with the system providing data and/or receiving system suggestions and advices for making decisions. The user will only trust on the system if the system output is coherent according to his/her background. Notice that the use of a comprehensible model can help the final user to understand the system output.
- The *system designer* who has to be an expert on fuzzy logic in order to produce a good model useful for the final user of the application. A transparent (gray-box) model structure is really appreciated for the future model maintenance and update.
- The *domain expert* who will explain the system behavior to the system designer during the model design stage. In addition, he/she will be in charge of validating system running. Since domain experts usually do not know anything about fuzzy logic a clearly readable model description is required to make easier the validation stage.

In our study, FL experts are mainly considered as system designers but due to the nature of the problem they also can act as domain experts and even as final users. In turn, non-FL users are only viewed as domain experts or final users. Twenty six answers were collected. They show a huge diversity what clearly illustrates how different users have very different criteria to measure interpretability. Three main questions were asked as part of the poll:

1. *How much interpretable are the twelve KBs?* Each user was asked to give an interpretability measure for each KB. Such measure was represented by an interval (min-max), i.e., the range in which it should be included, between zero and one hundred. However, only a few users were willing to answer to this question with numerical values. In fact, we realize that people find much more natural to make approximate reasoning based on the use of linguistic terms like *Highly interpretable*, *Moderately interpretable*, etc. In addition collected values show a huge variance. In consequence, it does not make sense drawing statistical conclusions from stored data. According to these results it can be argued that people get into difficulties when they have to give numerical indices as computers usually do.
2. *What is the KB interpretability ranking?* Users were asked to rank the KBs according to their preferences from the interpretability point of view: One for the most interpretable KB, and twelve for the least interpretable one. Since all users were willing to answer this question, an interesting conclusion can be drawn: *People feel much more confident setting rankings than giving numerical values*. In order to set a ranking it is necessary to compare all the KBs (by couples) but it does not imply setting individual measures. First column of Table 3 includes the user identifier, setting in brackets if the user is used to work with fuzzy systems (F) or not (NF). Second column of the table shows rankings given by users. As it can be seen at first glance there is a huge variance. Only two couples of users (1-26 and 4-11) gave exactly the same order. Nevertheless, looking carefully answers are not so different. The global order is more or less the same for all users but when two KBs are quite close regarding interpretability the final ranking choice depends in many subtle details, and as a result, there is a clearly subjective choice at the end.

   The comparison between rankings provided by the users (Table 3) and rankings derived from the computed interpretability indices (Table 2) lets us evaluate the goodness of such indices. However, only user 3 and fuzzy index yield the same ranking. In order to make a deeper analysis, we have computed Euclidean distance from each of the five interpretability indices, x (first line of the table), to all the twenty six users, y, according to equation 1 where x_i means the ranking position of KB_i regarding index x and y_i is the ranking position of KB_i regarding user y.

$$d_{x,y} = \sqrt{\sum_{i=1}^{12} |x_i - y_i|} \tag{1}$$

   Computed distances give an idea on how different (comparing positions of each KB in selected rankings) indices and user's answers are. Most user's answers are closer to the rankings obtained using *Number of rules*. There are also many answers closer to *Total rule length* and *Fuzzy index*. In fact, it is possible to identify several groups of users (look at Table 4) whose answers fit better with some indices but there is a lot of overlapping among groups. Moreover, approximately the same number of F and NF users belongs to all the groups. This is due to the fact that in general none of the computed indices fit properly in user's answers.

3. *What are the most relevant aspects to consider when assessing interpretability?* Each user was asked to give short comments explaining what he/she considers good strategies and/or key criteria to measure interpretability. Some of the most useful comments are listed below:
   - *A common heuristic reasoning is the following. First look at the total number of rules. Second, if there is ambiguity between some of the knowledge bases it is needed to check the total number of premises. Then, if there is still ambiguity it is necessary to analyze the complexity of the linguistic terms*. This suggests making the ranking in different abstraction levels, adding new criteria only when they are needed.
   - *I prefer short rules considering at most 5 features than fewer rules with a long size*. This remarks that the number of inputs by rule is a main criterion. However, different people have different views about what must be considered as a small

ISBN: 978-989-95079-6-8

Table 3: Ranking of KBs extracted from the poll results.

User	+ Interpretability -
user1 (F)	KB8, KB2, KB6, KB12, KB10, KB4, KB9, KB1, KB5, KB11, KB7, KB3
user2 (F)	KB6, KB2, KB4, KB1, KB8, KB5/KB9, KB10, KB12, KB11, KB7, KB3
user3 (F)	KB8, KB2, KB6, KB4, KB12, KB9, KB10, KB1, KB5, KB11, KB7, KB3
user4 (NF)	KB2, KB6, KB8, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3
user5 (F)	KB2, KB6, KB8, KB9, KB10/KB12, KB4, KB1, KB5, KB11, KB7, KB3
user6 (F)	KB8/KB12, KB2/KB6, KB4/KB9/KB10, KB1/KB5, KB11, KB3, KB7
user7 (F)	KB8, KB6, KB2, KB12, KB10, KB9, KB4, KB1, KB5, KB3, KB11, KB7
user8 (F)	KB8, KB2, KB6, KB12, KB9, KB4, KB10, KB1, KB5, KB11, KB7, KB3
user9 (NF)	KB2, KB9, KB12, KB8, KB6, KB10, KB5, KB4, KB1, KB11, KB3, KB7
user10 (NF)	KB6, KB2, KB9, KB12, KB4, KB8, KB5, KB1, KB11, KB10, KB7, KB3
user11 (NF)	KB2, KB6, KB8, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3
user12 (NF)	KB8, KB12, KB2, KB6, KB9, KB4, KB10, KB5, KB1, KB11, KB7, KB3
user13 (F)	KB2, KB6, KB8, KB9, KB12, KB4/KB10, KB1/KB5/KB11, KB3/KB7
user14 (NF)	KB8, KB2/KB6, KB12, KB9, KB4, KB10, KB1, KB5, KB11, KB7, KB3
user15 (NF)	KB8, KB2, KB12, KB6, KB10, KB1, KB5, KB9, KB4, KB11, KB7, KB3
user16 (NF)	KB8, KB6, KB2, KB12, KB10, KB4, KB9, KB5, KB1, KB11, KB7, KB3
user17 (NF)	KB8, KB6, KB2, KB12, KB4, KB9, KB5, KB1, KB11, KB10, KB3, KB7
user18 (NF)	KB2/KB12, KB4/KB6/KB8/KB10, KB5/KB9, KB11, KB1, KB7, KB3
user19 (NF)	KB2, KB4, KB6, KB11, KB5, KB1, KB9, KB12, KB10, KB8, KB7, KB3
user20 (F)	KB2, KB8, KB9, KB12, KB4, KB6, KB10, KB11, KB1, KB5, KB7, KB3
user21 (NF)	KB2, KB6, KB8, KB12, KB9, KB4, KB10, KB11, KB5, KB1, KB7, KB3
user22 (F)	KB8, KB6, KB2, KB9, KB12, KB10, KB4, KB5, KB11, KB1, KB3, KB7
user23 (NF)	KB8, KB2, KB6, KB12, KB10, KB9, KB4, KB1, KB5, KB11, KB7, KB3
user24 (F)	KB2, KB8, KB6, KB12, KB4, KB9, KB1, KB10, KB5, KB11, KB7, KB3
user25 (F)	KB8, KB2/KB4/KB6/KB9/KB10/KB12, KB1/KB5/KB11, KB3/KB7
user26 (F)	KB8, KB2, KB6, KB12, KB10, KB4, KB9, KB1, KB5, KB11, KB7, KB3

Table 4: Groups of users regarding computed interpretability indices.

Index	Users	F	NF
Number of rules	4, 5, 7, 9, 11, 13, 15, 16, 22, 23, 25	5	6
Total rule length	1, 8, 12, 14, 18, 21, 25, 26	4	4
Average rule length	2, 19	1	1
Nauck's index	6, 12, 25	2	1
Fuzzy index	3, 8, 10, 14, 17, 20, 21, 24, 25	5	4

number of inputs by rule. This problem arises from the intrinsic ambiguity of natural language: What means small? The same word has different meanings in different contexts, but even in the same context it has different meanings for different people. The use of fuzzy logic formalizes a precise meaning for each word coping with this kind of ambiguity.

- *With respect to words (linguistic variables and terms), the better choice of words within the context of the problem, the more accurate interpretation*. Understanding strongly depends on the context of the problem. For instance, it is easy to see how different the meaning of *High* is when talking about people, buildings, or mountains.

3 User profiles

With the aim of finding out some user profiles from Table 3 we have applied a hierarchical clustering analysis [20]. Two dendrograms were built using Ward's method [21] and squared Euclidean distance (see Fig. 1): The first one (on the top part of the figure) only regarding fuzzy expert users where two groups (SF1 and SF2) are clearly identified. The second one (on the bottom part of the figure) including only non-fuzzy users where only one group (SNF1) can be defined. The rest of users are progressively added by the clustering algorithm but they stay at long distance.

It is possible to extract a prototype user profile from each group. According to our experience designing and assessing interpretable fuzzy systems, and keeping in mind the conclusions derived from the web poll, ten variables were selected as tentative interpretability indicators: (1) Number of rules; (2) Total rule length; (3) Percentage of rules which use less than ten percent of inputs; (4) Percentage of rules which use between ten and thirty percent of inputs; (5) Percentage of rules which use more than thirty percent of inputs; (6) Number of inputs; (7) Number of labels used in the rule base; (8) Percentage of elementary labels used in the rule base; (9) Percentage of OR composite labels used in the rule base; (10) Percentage of NOT composite labels used in the rule base.

The task consists of discovering those indicators that can be considered as key to distinguish among groups. From the rankings provided by users it can be induced the order relation between KBs. Assuming that each ranking is based on a comparison per couples of all KBs, Table 3 is translated into a data set with the following format. Each column give the difference between the ten selected indicators (listed above) for each couple of KBs (A and B). The last column includes one

ISBN: 978-989-95079-6-8

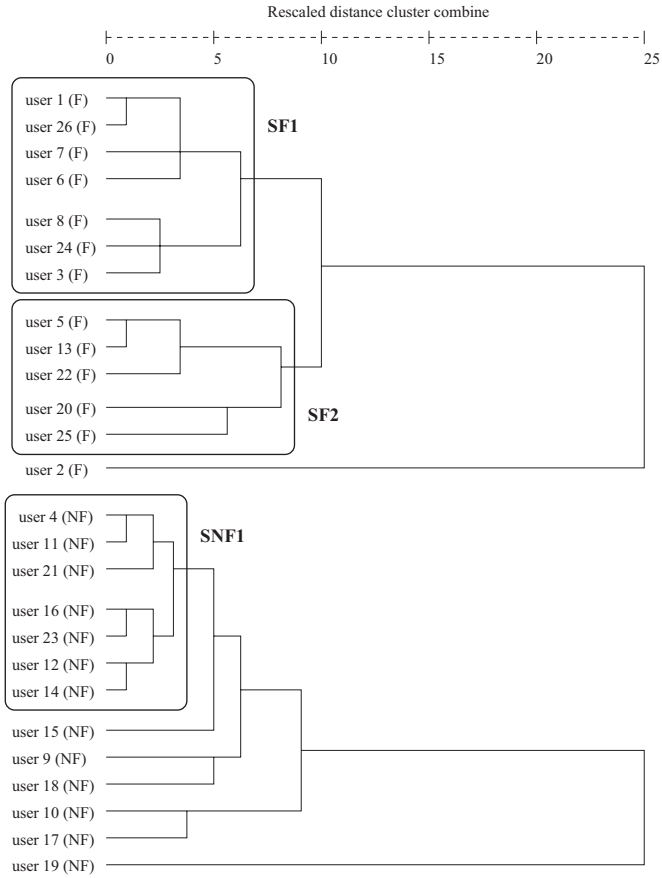


Figure 1: Groups of users by hierarchical clustering.

of three labels (1:*No*, 2:*Yes*, 3:*I don't know*) answering to the question: is A more interpretable than B? The whole data set was divided into three subsets including only the comparisons related to the users belonging to each group.

Using HILK methodology [16] (FDT-S method) and the previously generated data subsets we have built three fuzzy classifiers, one for each group of users. For each couple of KBs (A and B) the fuzzy classifiers give as output (1) *A is more interpretable than B*, (2) *B is more interpretable than A*, or (3) *I don't know*, what means that A and B are similar from an interpretability point of view. The comparison among all KBs yields a ranking for each group. Figure 2 presents obtained rankings as well as the interpretability indicators (inputs of the fuzzy classifier) that define the six prototype user profiles we were looking for. Three indicators are not selected by any of the prototypes: (3) Percentage of rules which use less than ten percent of inputs; (5) Percentage of rules which use more than thirty percent of inputs; and (9) Percentage of OR composite labels used in the rule base. It seems that all users agree that a small percentage of inputs by rule is good for interpretability and a large percentage is bad, while composite propositions including OR are seen as easily readable. Notice that each prototype is defined using at least three indicators, i.e., the use of basic indices is not enough.

In addition, for each group we have computed the distance (mean and variance) between the prototype and all users included in the group. SF1 represents a quite compact group where the prototype is defined using only three indicators, while SF2 and SNF1 are less compact and because of that they need more indicators. Although clustering of F and NF users should intuitively yield more homogeneous groups, in practice there is still a lot of diversity inside each group. In the case of F users, prototypes achieve medium mean distance with small variance. SF2 takes into account six indicators because it is made up of only five users quite heterogeneous. It results specially interesting the fact that *Total rule length* which seemed to be a very important basic indicator according to Table 4 only is taken into account by group SF2. It is not nearly the most relevant indicator in comparison with the subsets of indicators emerging from the clustering analysis. Finally, it could be argued that fuzzy users (SF1 and SF2) give more homogeneous answers but regarding more complex criteria than non-fuzzy users (SNF1). Finally, as a result of the heterogeneity of NF users only seven of them give more or less similar rankings. In consequence, four indicators were relevant for this group but yielding large mean and variance.

4 Conclusions

Assessing interpretability is a very challenging and complex task due to the inherent subjectivity of the measure. In order to evaluate existing indices we have set up a first experimental study, for simplicity limited to twelve rule bases assuring most interpretability constraints described as essential in the literature. As a result, assuming knowledge bases under study are interpretable the study focus on quantifying interpretability and comparing obtained results with assessment provided by people in a web poll. None of the evaluated indices gave good results in comparison with rankings provided by human beings. A lot of work remains still to be done so that finding a universal index. However, results derived from our experimental study offer some interesting clues.

First, it is necessary to define a new index flexible enough to be easily adaptable to the problem context and user preferences. Such index must take into account many subtle details combined in the context of computing with words and perceptions based on fuzzy logic technology. In addition, obtaining a numerical value is not needed in most applications where the important thing is to set an appropriate ranking.

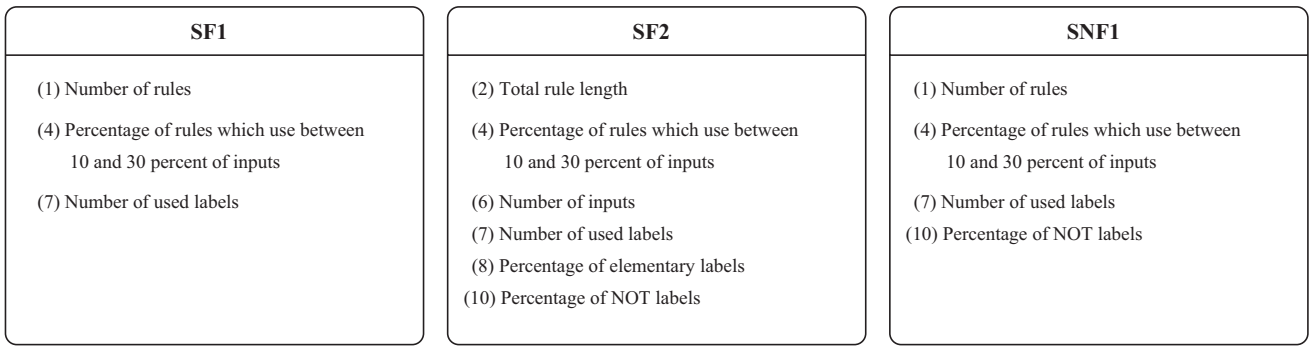
Second, a hierarchical fuzzy framework has been proved as a powerful tool to imitate the usual way of people reasoning. It mainly consists of taking a few interpretability indicators as guide to discriminate between two knowledge bases, adding more criteria only when it is necessary because the compared knowledge bases are not distinguishable at first glance.

Finally, we have focused on interpretability from a structural point of view but there are many cognitive aspects that should be analyzed. Therefore more experimental studies are needed. Obviously, as a first step our study has been limited to a very specific kind of fuzzy rule-base systems. Of course, in the future it would be interesting to make a comparison of different rule base structures.

Acknowledgment

Authors would like to acknowledge people who have filled up the web poll. Their suggestions were very interesting and they have contributed so much to this work.

This work has been partially funded by the Foundation for the Advancement of Soft Computing (Asturias) and Spanish government (CICYT) under grant: TIN2008-06890-C02-01.



	Ranking	Distance	
		Mean	Variance
SF1	KB8, KB2, KB6, KB12, KB4, KB9/KB10, KB1, KB5, KB11, KB7, KB3	2.616	0.597
SF2	KB2/KB8, KB6, KB9/KB12, KB4, KB10, KB1, KB5, KB11, KB3, KB7	3.842	0.551
SNF1	KB8, KB2, KB6, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3	2.789	3.086

Figure 2: Prototype user profiles.

References

[1] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Parts I, II, and III. Information Sciences*, 8, 8, 9:199–249, 301–357, 43–80, 1975.

[3] E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Transactions on Computers*, 26(12):1182–1191, 1977.

[4] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm. *IEEE Transactions on Fuzzy Systems*, 8(5):591–600, 2000.

[5] S. Guillaume and B. Charnomordic. Generating an interpretable family of fuzzy partitions. *IEEE Transactions on Fuzzy Systems*, 12(3):324–335, 2004.

[6] C. Mencar and A. M. Fanelli. Interpretability constraints for fuzzy information granulation. *Information Sciences*, 178(24):4585–4618, 2008.

[7] S.-M. Zhou and J. Q. Gan. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. *Fuzzy Sets and Systems*, 159(23):3091–3131, 2008.

[8] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. *Soft Computing*, 13(5):437–449, 2009.

[9] P. Fazendeiro and J. V. de Oliveira. A working hypothesis on the semantics/accuracy synergy. In *Joint EUSFLAT-LFA*, pages 266–271, Barcelona, Spain, September, 7-9, 2005.

[10] Y. Jin, W. von Seelen, and B. Sendhoff. On generating FC^3 fuzzy rule systems from data using evolution strategies. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(6):829–845, 1999.

[11] C. Mencar, G. Castellano, and A. M. Fanelli. Distinguishability quantification of fuzzy sets. *Information Sciences*, 177(1):130–149, 2007.

[12] M. Setnes, R. Babuska, U. Kaymak, and H. R. van Nauta Lemke. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man, and Cybernetics. Part B*, 28(3):376–386, 1998.

[13] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4–31, 2007.

[14] D. D. Nauck. Measuring interpretability in rule-based classification systems. In *FUZZ-IEEE 2003*, volume 1, pages 196–201, St. Louis, Missouri, USA, May, 25-28, 2003.

[15] J. M. Alonso, S. Guillaume, and L. Magdalena. A hierarchical fuzzy system for assessing interpretability of linguistic knowledge bases in classification problems. In *IPMU06. Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 348–355, Paris, France, July, 2-7, 2006.

[16] J. M. Alonso, L. Magdalena, and S. Guillaume. HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems*, 23(7):761–794, 2008.

[17] L.-X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, 22(6):1414–1427, 1992.

[18] H. Ichihashi, T. Shirai, K. Nagasaka, and T. Miyoshi. Neuro-fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. *Fuzzy Sets and Systems*, 81(1):157–167, 1996.

[19] J. M. Alonso, S. Guillaume, and L. Magdalena. KBCT: A knowledge management tool for fuzzy inference systems. *Free software under GPL license, available in http://www.mat.upm.es/projects/advocate/kbct.htm*, 2003.

[20] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[21] J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

ISBN: 978-989-95079-6-8

Games and capacities on partitions

Michel Grabisch
Centre d'Economie de la Sorbonne, Université Paris I 106-112, Bd de l'Hôpital, 75013 Paris, France
Email: michel.grabisch@univ-paris1.fr

Abstract— We consider capacities and games whose value $v(S)$ depend on the organization of $N \setminus S$, this organization being represented by a partition. Hence we deal with capacities and games depending on two arguments: a subset (coalition) S and a partition π containing S. We call embedded coalition any such pair (S, π). We first provide a suitable structure for the set of embedded coalitions, study its properties, and find the Möbius function for this structure. Then, we define games and capacities on this structure and study their properties. We propose also a concept similar to the Shapley value, and provide an axiomatization of it. Lastly, we give some insights on the Choquet integral.

Keywords— capacity, fuzzy measure, partition, Möbius transform, Shapley value

1 Introduction

Capacities [1] or fuzzy measures [2] are usually defined on the set of subsets of the universe, assumed to be finite in this paper, or on a subcollection of it. Specifically, denoting by N the universe, let us take a capacity v defined on 2^N, and consider some subset $S \in 2^N$ together with the quantity $v(S)$. The interpretation of $v(S)$ may differ according to the context, but usually falls into two categories: either $v(S)$ represents a degree of certainty, plausibility, etc. that the true (but unknown) state of the world is contained in S, or $v(S)$ represents some power, importance, strength, monetary value of the group S of entities (agents, players, voters, criteria, etc.). Using a common word, we may say that $v(S)$ is the *worth* of S.

Let us consider more carefully the second interpretation, which is related to cooperative game theory. If N represents a society of individuals, speaking of $v(S)$ implicitly means that the society is formed of the group (or coalition) S and the coalition $N \setminus S$. Under this organization, the worth of S is $v(S)$. We may however consider less simple situations, where the set of remaining agents $N \setminus S$ is also organized in some way, say $N \setminus S = S_2 \cup S_3 \cup \cdots \cup S_k$, where $\{S_2, S_3, \ldots, S_k\}$ is a partition of $N \setminus S$. Then it is natural to think that the worth of S may depend on the organization of the remaining agents. Therefore, the expression $v(S)$ is no more enough precise, and we need to consider $v(S, \{S, S_2, S_3, \ldots, S_k\})$, for all possible organizations of the society containing S as a block of the society. The usual term for the argument of v, namely (S, π) with π the partition $\{S, S_2, S_3, \ldots, S_k\}$, is *embedded coalition*, and v is called a *game in partition function form*, or *PFF-game* for short [3]. We will call them in this paper games (or capacities) on partitions.

Although games on partitions have been studied in the game theoretic literature (essntially the Shapley value, the core), no explicit study has been done on the mathematical object (S, π) and its structure, nor on the properties of v. The aim of this paper is precisely to fill this gap. We will provide a structure for embedded coalitions, study its properties, and find the Möbius function on this structure. Then we will study properties of games, and define a Shapley value for v, and finally give some remarks on the possibility to define a Choquet-like integral. A part of this paper is based on the working paper [4].

2 The structure of embedded coalitions

Let $N := \{1, 2, \ldots, n\}$ be the universal set (set of agents, players, etc.). We denote by $S, T, \ldots$ the subsets of N (coalitions), and their cardinality by $s, t, \ldots$. We consider the set $\Pi(N)$ (denoted also by $\Pi(n)$) of all possible partitions of N (coverings of N by disjoint subsets). For a partition $\pi := \{S_1, \ldots, S_k\}$, subsets $S_1, \ldots, S_k$ are called *blocks* of π. A partition in k blocks is called a *k-partition*. A natural ordering of partitions is given by the notion of "coarsening" or "refinement". Taking π, π' partitions in $\Pi(N)$, we say that π is a *refinement* of π' (or π' is a *coarsening* of π), denoted by $\pi \leq \pi'$, if any block of π is contained in a block of π' (or every block of π' fully decomposes into blocks of π). Then $(\Pi(N), \leq)$ is a lattice, called the *partition lattice*. With this ordering, the bottom element of the lattice is the finest partition $\pi^\perp := \{\{1\}, \ldots, \{n\}\}$, while the top element is the coarsest partition $\pi^\top := \{N\}$.

An *embedded coalition* is a pair (S, π), where $S \in 2^N \setminus \{\emptyset\}$, and $\pi \in \Pi(N)$ is such that $S \in \pi$. We denote by $\mathfrak{C}(N)$ (or by $\mathfrak{C}(n)$) the set of embedded coalitions on N. For the sake of concision, we often denote by $S\pi$ the embedded coalition (S, π), and omit braces and commas for subsets (example with $n = 3$: $12\{12, 3\}$ instead of $(\{1, 2\}, \{\{1, 2\}, \{3\}\})$). Remark that $\mathfrak{C}(N)$ is a proper subset of $2^N \times \Pi(N)$.

As mentionned in the introduction, works on games on partitions do not explicitly define a structure (that is, some order) on embedded coalitions. A natural choice is to take the product order on $2^N \times \Pi(N)$:

$$(S, \pi) \sqsubseteq (S', \pi') \Leftrightarrow S \subseteq S' \text{ and } \pi \leq \pi'.$$

Evidently, the top element of this ordered set is $(N, \pi^\top)$ (denoted more simply by $N\{N\}$ according to our conventions). However, due to the fact that the empty set is not allowed in (S, π), there is no bottom element in the ordered structure $(\mathfrak{C}(N), \sqsubseteq)$. Indeed, all elements of the form $(\{i\}, \pi^\perp)$ are minimal elements. For mathematical convenience, we introduce an artificial bottom element $\perp$ to $\mathfrak{C}(N)$ (it could be considered as $(\emptyset, \pi^\perp)$), and denote $\mathfrak{C}(N)_\perp := \mathfrak{C}(N) \cup \{\perp\}$. We give as illustration the partially ordered set $(\mathfrak{C}(N)_\perp, \sqsubseteq)$ with $n = 3$ (Fig. 1).

ISBN: 978-989-95079-6-8

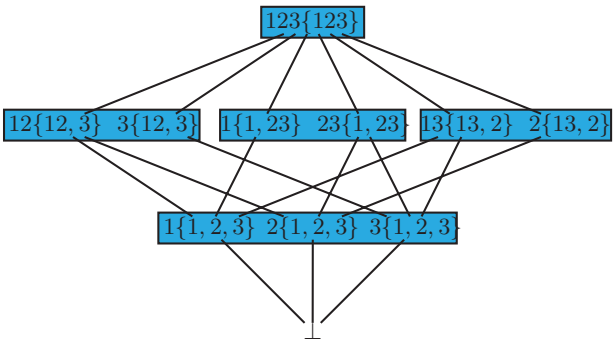


Figure 1: Hasse diagram of $(\mathfrak{C}(N)_\perp, \sqsubseteq)$ with $n = 3$. Elements with the same partition are framed in grey.

Another possibility to define an order on embedded coalitions is to basically take the Boolean lattice 2^N, and for any $S \in 2^N$, duplicate it into $S\pi_1, S\pi_2, \ldots$ for all partitions π containing S. Then $(S, \pi) \leq (S', \pi')$ if and only if $S \subseteq S'$. This seems to be the underlying structure in the work of Macho-Stadler et al. [5]. We illustrate this structure for $n = 3$ in Figure 2.

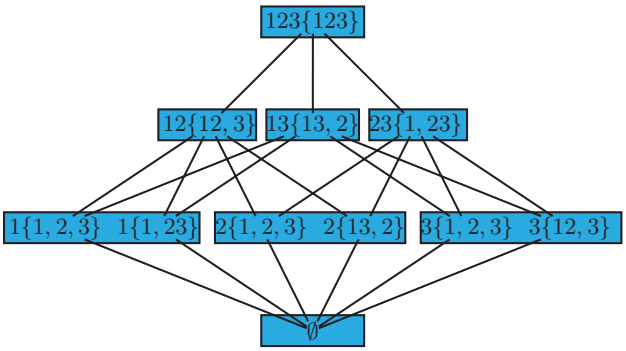


Figure 2: Hasse diagram of the structure of Macho-Stadler et al. for $n = 3$. Elements with same coalition (duplicates) are framed in grey.

In the sequel, we focus on the first structure, since it has better properties. The second one is not a lattice, since for example, $12\{12, 3\}$ and $13\{13, 2\}$ are two minimal upper bounds of $1\{1, 2, 3\}$ and $1\{1, 23\}$.

We recall that in a partially ordered set $(P, \leq)$ with a bottom element $\perp$ and a top element $\top$, a *chain from* $\perp$ *to* $\top$ is a totally ordered sequence of elements of P including $\perp, \top$. The chain is *maximal* if no other chain can contain it (equivalently, if between two consecutive elements x_i, x_{i+1} of the sequence, there is no element $x \in P$ such that $x_i < x < x_{i+1}$). If no ambiguity occurs, we say *maximal chain* instead of *maximal chain from* $\perp$ *to* $\top$. The set of maximal chains in P is denoted by $\mathcal{C}(P)$. The *length* of a maximal chain is the number of elements of the sequence minus 1.

The following is shown in [4].

Proposition 1. The following holds.

(i) $(\mathfrak{C}(N)_\perp, \sqsubseteq)$ is an upper semimodular lattice, whose top and bottom elements are $(N, \pi^\top)$ and $\perp$. All elements are complemented: for a given $S\pi$, any embedded coalition of the form $\overline{S}\pi_{\overline{S}}$ with $\overline{S}$ the complement of S and $\pi_{\overline{S}}$ any partition containing $\overline{S}$.

(ii) Every maximal chain from $\perp$ to an element (S, π) has the same length, which is $n - k + 1$, if π is a partition in k blocks. Hence, the height of the lattice is n.

(iii) The total number of elements of $\mathfrak{C}(N)_\perp$ is $\sum_{k=1}^{n} kS_{n,k} + 1$, where $S_{n,k}$ is the Stirling number of second kind.

n	1	2	3	4	5	6	7	8
$\lvert\mathfrak{C}(n)_\perp\rvert$	2	4	11	38	152	675	3264	17008

(iv) The number of maximal chains from $\perp$ to $N\{N\}$ is $\frac{(n!)^2}{2^{n-1}}$, which is also the number of maximal longest chains in $\mathfrak{C}(N)$.

n	1	2	3	4	5	6	7	8
$\lvert\mathcal{C}(\mathfrak{C}(n)_\perp)\rvert$	1	2	9	72	900	16 200	396 900	12 700 800

(v) Let (S, π) be an embedded coalition, with $\pi := \{S, S_2 \ldots, S_k\}$, and $|S| = s$. The number of maximal chains from $\perp$ to (S, π) is

$$|\mathcal{C}([\perp, (S, \pi)])| = \frac{s(n-k)!}{2^{n-k}} s!s_2! \cdots s_k!.$$

For example, there are 9 chains from $\perp$ to $123\{123, 4\}$.

(vi) Let (S, π) be an embedded coalition, with $\pi := \{S, S_2 \ldots, S_k\}$. The number of maximal chains from (S, π) to $N\{N\}$ is

$$|\mathcal{C}([(S, \pi), N\{N\}])| = \frac{1}{k}|\mathcal{C}(\mathfrak{C}(k)_\perp)| = \frac{k!(k-1)!}{2^{k-1}}.$$

The above results focus on maximal chains, since they are a fundamental concept for the sequel, especially for the Shapley value. It is important to note that this lattice is neither distributive, modular nor atomic (and hence not geometric). Atoms are all elements of the form $\{i\}\pi^\perp$, and elements representable by a join of atoms are only those of the form $S\pi^\perp_S$, where $\pi^\perp_S = \{S, i_1, \ldots, i_{n-s}\}$, with $N \setminus S = \{i_1, \ldots, i_{n-s}\}$.

3 Games and capacities on partitions

Definition 1. A *game on partitions* on N is a mapping $v : \mathfrak{C}(N)_\perp \to \mathbb{R}$, such that $v(\perp) = 0$. The set of all games on partitions on N is denoted by $\mathcal{PG}(N)$.

Although any such function could be called a game on partitions, to be meaningful, we need to assume that forming the grand coalition generates the largest total surplus, i.e., $v(N\{N\}) \geq \sum_{S \in \pi} v(S, \pi)$, for all $\pi \in \Pi(N)$.

The structure of lattice permits to define the usual notions on games and capacities.

ISBN: 978-989-95079-6-8

Definition 2. Let $v \in \mathcal{PG}(N)$.

(i) v is *monotone* if $S\pi \sqsubseteq S'\pi'$ implies $v(S\pi) \leq v(S'\pi')$. A monotone game on partitions is called a *capacity on partitions*. A capacity on partitions v is normalized if $v(N\{N\}) = 1$.

(ii) v is *supermodular* if for every $S\pi, S'\pi'$ we have

$$v(S\pi \vee S'\pi') + v(S\pi \wedge S'\pi') \geq v(S\pi) + v(S'\pi').$$

It is *submodular* if the reverse inequality holds.

(iii) A game is *additive* if it is both supermodular and submodular.

(iv) More generally, for a given $k \geq 2$, a game is *k-monotone* if for all families of k elements $S_1\pi_1, \ldots, S_k\pi_k$, we have

$$v(\bigvee_{i\in K} S_i\pi_i) \geq \sum_{J\subseteq K, J\neq\emptyset} (-1)^{|J+1|} v(\bigwedge_{i\in J} S_i\pi_i)$$

putting $K := \{1, \ldots, k\}$. A game is ∞-monotone if it is k-monotone for every $k \geq 2$.

(v) A game is a belief function if it is a normalized ∞-monotone capacity.

Above, $\vee, \wedge$ are the least upper bounds and greatest lower bounds. The greatest lower bound is simply given by:

$$(S, \pi) \wedge (S', \pi') = (S \cap S', \pi \wedge \pi') \text{ if } S \cap S' \neq \emptyset,$$

and $\perp$ otherwise. The least upper bound is more tricky to define:

$$(S, \pi) \vee (S', \pi') = (T \cup T', \rho)$$

where T, T' are blocks of $\pi \vee \pi'$ containing respectively S and S', and ρ is the partition obtained by merging T and T' in $\pi \vee \pi'$.

The following result is due to Barthélemy [6].

Proposition 2. Let L be a lattice. Then f is monotone and ∞-monotone on L if and only if it is monotone and $(|L| - 2)$-monotone.

In lattice theory, a *valuation* (or *2-valuation*) on a lattice L is a real-valued function on L being both super- and submodular (additive). More generally, for a given $k \geq 2$, a *k-valuation* satisfies

$$v(\bigvee_{i\in K} x_i) = \sum_{J\subseteq K, J\neq\emptyset} (-1)^{|J+1|} v(\bigwedge_{i\in J} x_i)$$

for every family of k elements. An ∞-valuation is a function f which is a k-valuation for every $k \geq 2$. The following results are useful (see [7, Ch. X], and also [6]).

Proposition 3. Let L be a lattice.

(i) L is modular if and only if it admits a strictly monotone valuation v (i.e., $x < y$ implies $v(x) < v(y)$).

(ii) L is distributive if and only if it admits a strictly monotone 3-valuation.

(iii) L is distributive if and only if it is modular and every strictly monotone valuation is a k-valuation for any $k \geq 2$.

(iv) Any lattice admits an ∞-valuation.

The consequence of (i) is that no strictly monotone additive game exists since $\mathfrak{C}(n)_\perp$ is not modular when $n > 2$. The question is: does it exist an additive game? The following proposition answers this question.

Proposition 4. $\mathfrak{C}(2)_\perp$ admits a strictly monotone 2-valuation (hence by Prop. 2, a strictly monotone ∞-monotone valuation). If $n > 2$, the only possible valuations on $\mathfrak{C}(n)_\perp$ are constant valuations.

Since any game v satisfies $v(\perp) = 0$, we have:

Corollary 1. The only additive game is the constant game $v = 0$.

Definition 3. Let $T\sigma \in \mathfrak{C}(N)$.

(i) The *unanimity game centered on $T\sigma$* is the game defined by

$$u_{T\sigma}(S\pi) = \begin{cases} 1, & \text{if } S\pi \sqsupseteq T\sigma \\ 0, & \text{otherwise.} \end{cases}$$

(ii) The *identity game centered on $T\sigma$* is the game defined by

$$\delta_{T\sigma}(S\pi) = \begin{cases} 1, & \text{if } S\pi = T\sigma \\ 0, & \text{otherwise.} \end{cases}$$

Unanimity games are normalized capacities while the latter ones are not. Identity games form a basis of the vector space $\mathcal{PG}(N)$ since for any $v \in \mathcal{PG}(N)$,

$$v = \sum_{T\sigma\in\mathfrak{C}(N)} v(T\sigma)\delta_{T\sigma}.$$

From Proposition 1 (iii), the dimension of the basis is $\sum_{k=1}^{n} kS_{n,k}$. A less trivial basis is given by unanimity games, through the Möbius transform, described in the next paragraph.

In capacity theory and partially ordered sets, a key point is to determine the Möbius function, enabling to compute the Möbius transform of games and capacities. We recall that for a partially ordered set $(P, \leq)$ which is locally finite, the Möbius function $\mu(x, y) : P^2 \to \mathbb{R}$ permits to solve the equation:

$$g(x) = \sum_{y\leq x} f(y)$$

where $f, g : P \to \mathbb{R}$, as follows:

$$f(x) = \sum_{y\leq x} g(y)\mu(y, x).$$

(see Rota [8], Aigner [9]). Based on properties of $\mathfrak{C}(N)_\perp$, the following fundamental result:

ISBN: 978-989-95079-6-8

Proposition 5. If P is a lattice with bottom element 0 and set of atoms $\mathcal{A}$, for every $x \in P$, the Möbius function reads

$$\mu(0,x) = \sum_{S \subseteq \mathcal{A} | \bigvee S = x} (-1)^{|S|},$$

and the fact that for the partition lattice we know that $\mu_{\Pi(n)}(\pi^{\perp}, \{N\}) = (-1)^{n-1}(n-1)!$ (see Aigner [9, p. 154]), the following can be shown.

Proposition 6. The Möbius function on $\mathfrak{C}(n)_{\perp}$ is given by:

$$\mu(\perp, S\pi) = \begin{cases} (-1)^{|S|}, & \text{if } \pi = \pi_S^{\perp} \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

$$\mu(S'\pi', S\pi) = (-1)^{k'-k}(l_1 - 1)! \cdots (l_k - 1)!, \quad (2)$$

for every $S'\pi' \leq S\pi$, where $\pi_S^{\perp}$ is the partition consisting in S and all singletons in $N \setminus S$, $S\pi = S\{S, S_2, \ldots, S_k\}$, $S'\pi' = S'\{S', S_{12}, \ldots, S_{1l_1}, S_{21}, \ldots, S_{2l_2}, \ldots, S_{k1}, \ldots, S_{kl_k}\}$, $S_i = S_{i1} \cup \cdots \cup S_{il_i}$ for $i = 2, \ldots, k$, and $k' := \sum_{i=1}^{k} l_i$.

Using this, the Möbius transform of any PFF-game v is given by

$$m(S\pi) = \sum_{S'\pi' \sqsubseteq S\pi} \mu(S'\pi', S\pi) v(S'\pi').$$

Now, $m(S\pi)$ gives the coordinates of v in the basis of unanimity games.

Example 1. Application for $n = 3$. Let us compute the Möbius transform of a game v. We have, for all distinct $i, j, k \in \{1, 2, 3\}$:

$$\begin{aligned} m(\perp) &= 0 \\ m(i\pi^{\perp}) &= v(i\pi^{\perp}) \\ m(ij\{ij,k\}) &= v(ij\{ij,k\}) - v(i\pi^{\perp}) - v(j\pi^{\perp}) \\ m(i\{i,jk\}) &= v(i\{i,jk\}) - v(i\pi^{\perp}) \\ m(123\{123\}) &= v(123\{123\}) - \sum_{i,j} v(ij\{ij,k\}) \\ &\quad - \sum_{i} v(i\{i,jk\}) + 2\sum_{i} v(i\pi^{\perp}). \end{aligned}$$

Barthélemy [6] proved the following.

Proposition 7. Let L be any lattice, and $f : L \to \mathbb{R}$. If the Möbius transform of f, denoted by m, satisfies $m(\perp) = 0$, $m(x) \geq 0$ for all $x \in L$, and $\sum_{x \in L} m(x) = 1$ (normalization), then f is a belief function.

The converse of this proposition does not hold in general (it holds for the Boolean lattice 2^N). A belief function is *invertible* if its Möbius transform is nonnegative, normalized and vanishes at $\perp$. The following counterexample shows that for $\mathfrak{C}(n)_{\perp}$, there exist belief functions which are not invertible.

Example 2. Let us take $\mathfrak{C}(3)_{\perp}$, and consider a function f whose values are given on the figure below. Monotonicity implies that $1 \geq \beta \geq \alpha \geq 0$. In order to check ∞-monotonicity, from Proposition 2 we know that it suffices to check till 7-monotonicity. We write below the most constraining inequalities only, keeping in mind that $1 \geq \beta \geq \alpha \geq 0$.

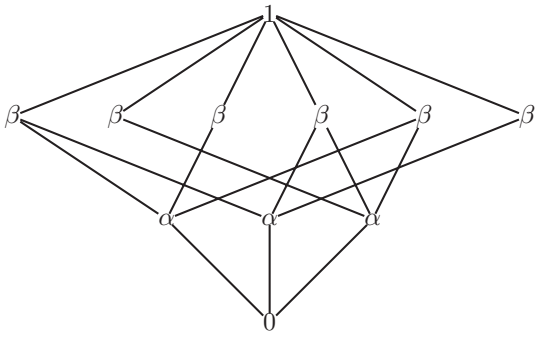


Figure 3: Hasse diagram of $(\mathfrak{C}(3)_{\perp}, \sqsubseteq)$

2-monotonicity is equivalent to $\beta \geq 2\alpha$ and $1 \geq 2\beta - \alpha$. 3-monotonicity is equivalent to $1 \geq 3\beta - \alpha$. 4-monotonicity is equivalent to $1 \geq 4\beta - 3\alpha$. 5-monotonicity is equivalent to $1 \geq 5\beta - 4\alpha$, while 6-monotonicity is equivalent to $1 \geq 6\beta - 9\alpha$. 7-monotonicity does not add further constraints.

From Example 1, nonnegativity of the Möbius transform implies $\alpha \geq 0$ (atoms), $\beta \geq 2\alpha$ (2nd level), and $6\alpha - 6\beta + 1 \geq 0$ for the top element. Then taking $\alpha = 0.1$, $\beta = 0.28$ make that f is a belief function, but $m(\top) = -0.08$.

The last point concerns necessity functions. A *necessity function* on a lattice L is a real-valued function f on L such that

$$f(x \wedge y) = \min(f(x), f(y)).$$

The following proposition by Barthélemy shows that necessity functions always exist on $\mathfrak{C}(n)_{\perp}$.

Proposition 8. Let L be a lattice and $f : L \to \mathbb{R}$ such that $f(\perp) = 0$, $f(\top) = 1$. Then f is a necessity function if and only if it is an invertible belief function whose Möbius transform is nonzero on a chain of $\mathfrak{C}(n)_{\perp}$.

In other words, taking any chain C in $\mathfrak{C}(n)_{\perp}$ and assigning nonnegative numbers (at least one of them being positive) on elements of C such that their sum is 1 generates (by Proposition 7) a belief function which is a necessity function.

4 The Shapley value

A key notion in game theory and also in applications of capacity theory is the Shapley value [10]. For classical games, it defines a rational way to share the quantity $v(N)$ among the individuals in N. Our aim is to define a similar notion for games on partitions. We give a simplified exposition of [4]. See also Macho-Stadler et al. [5] for another proposition based on their structure.

Basically, our Shapley value is defined through maximal chains in $\mathfrak{C}(N)_{\perp}$, which are called *scenarios*, because they depict a particular story of coalition formation, starting from the society of individuals and arriving at the grand coalition. The set of scenarios is denoted by $\mathfrak{S}$.

ISBN: 978-989-95079-6-8

In a scenario $\mathcal{S}$, some elements play a special role. We consider those elements $S\pi$ such that in the sequence of elements of $\mathcal{S}$ from bottom to top, $S\pi$ is the last element with S. They are called *terminal elements*. The set of terminal elements in $\mathcal{S}$ is denoted by $\mathcal{F}(\mathcal{S})$. For example, in the following scenario with $N = \{1, 2, 3, 4\}$, terminal elements are in bold:

$$\mathcal{S}_1 = \mathbf{1}\{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}\} \to 13\{13, 2, 4\} \to \mathbf{13}\{\mathbf{13}, \mathbf{24}\} \to \mathbf{N}\{\mathbf{N}\}.$$

Basically, a value represents the average contribution of players in the game. The contribution of player i is defined as the sum of his contribution in each scenario. However, this contribution is more complex to define than in classical games. Taking as example the above scenario $\mathcal{S}_1$, we see that three situations can arise:

(i) In one step, a single player enters: this happens for steps 1 and 2. This is exactly like classical games, and the contribution goes for the entering player.

(ii) In one step, several players enter together: this happens in the last step where players 2 and 4 enter together. In this case, the principle of insufficient reason tells us to divide the contribution equally among the entering players.

(iii) In one step, no new player enters: this happens in step 3 of $\mathcal{S}_1$. In this case, we wait till a new change occurs (i.e., a new player enters). Hence these steps are not taken into account.

Note that the third case implies that only steps which correspond to terminal elements are taken into account. Applying this methodology to the above scenario $\mathcal{S}_1$, we obtain:

(i) contribution of player 1: $v(1\{1, 2, 3, 4\}) - 0$

(ii) contribution of player 2: $\frac{1}{2}(v(1234\{1234\}) - v(13\{13, 24\}))$

(iii) contribution of player 3: $v(13\{13, 24\}) - v(1\{1, 2, 3, 4\})$

(iv) contribution of player 4: $\frac{1}{2}(v(1234\{1234\}) - v(13\{13, 24\}))$

Based on the above considerations, we can define the contribution of a player in a scenario.

Definition 4. The *contribution of player i in a given scenario* $\mathcal{S}$ is given by

$$\Delta_i^{\mathcal{S}}(v) := \frac{1}{|S' \setminus S|}(v(S'\pi') - v(S\pi)),$$

where $S\pi$ is the last element where i is not present (note that $S\pi \in \mathcal{F}(\mathcal{S})$), and $S'\pi'$ is the next terminal element.

Based on the idea of scenario, we define a scenario-value, from which the value will be constructed.

Definition 5. (i) A *scenario-value* is a mapping $\phi : \mathcal{PG} \to \mathbb{R}^{n\times|\mathfrak{S}|}$. Components of $\phi(v)$ are denoted by $\phi_i^{\mathcal{S}}(v)$ for scenario $\mathcal{S}$ and player i.

The *Shapley scenario-value* is defined by

$$\phi_i^{\mathcal{S}}(v) := \Delta_i^{\mathcal{S}}(v), \quad i \in N, \mathcal{S} \in \mathfrak{S}.$$

(ii) A *value* is a mapping $\phi : \mathcal{PG} \to \mathbb{R}^n$. Components of $\phi(v)$ are denoted by $\phi_i(v)$ for player i. Any scenario-value ϕ induces a value by:

$$\phi_i(v) := \frac{1}{c}\sum_{\mathcal{S}\in\mathfrak{S}} \phi_i^{\mathcal{S}}(v),$$

where c is the number of maximal chains in $\mathfrak{C}(N)_\perp$. The *Shapley value of v* is induced by the Shapley scenario-value.

It is possible to give a direct expression of the Shapley value without using the scenario-value, by expressing the game in the basis of identity games and using linearity of the Shapley value. Using the generic notation $T\sigma := T\{T, T_2, \ldots, T_k\}$, we have

$$\begin{aligned}\phi_i(v) = \frac{1}{n}v(N\{N\}) + \\ \sum_{T\sigma, T\ni i, T\neq N} \frac{2(n-k)!}{n!n!}(k-1)(k-1)!(k-2)!\times \\ t!t_2!\cdots t_k!v(T\sigma) \\ - \sum_{T\sigma, T\not\ni i} \frac{2t(n-k)!}{n!n!}(k-1)!(k-2)!\times \\ t!(t_2-1)!\cdots t_k!v(T\sigma),\end{aligned}$$

where it is assumed in the third term that $i \in T_2$, the second block of σ. An equivalent expression, although less computationally efficient but closer to the classical Shapley value, is:

$$\begin{aligned}\phi_i(v) = \frac{1}{n}v(N\{N\}) \\ + \sum_{T\sigma, T\not\ni i, T_2\supset\{i\}} \frac{2t(n-k)!}{n!n!}(k-1)!(k-2)!\times \\ t!(t_2-1)!\cdots t_k!\Big[\frac{t+1}{t}v(T\cup i\sigma_{T\cup i}) - v(T\sigma)\Big] \\ - \sum_{T\sigma, T\not\ni i, T_2=\{i\}} \frac{2t(n-k)!}{n!n!}(k-1)!(k-2)!\times \\ t!t_3!\cdots t_k!v(T\sigma),\end{aligned}$$

with $\sigma_{T\cup i}$ the partition obtained from σ by moving $i \in T_2$ to T. With $n = 3$, we obtain:

$$\begin{aligned}\phi_1(v) = \frac{1}{3}v(N\{N\}) + \frac{1}{9}v(12\{12, 3\}) + \frac{1}{9}v(13\{13, 2\}) \\ - \frac{2}{9}v(23\{23, 1\}) + \frac{1}{9}v(1\{1, 23\}) - \frac{1}{18}v(2\{2, 13\}) \\ - \frac{1}{18}v(3\{3, 12\}) + \frac{2}{9}v(1\{1, 2, 3\}) - \frac{1}{9}v(2\{1, 2, 3\}) \\ - \frac{1}{9}v(3\{1, 2, 3\}).\end{aligned}$$

We introduce axioms for the characterization of the Shapley scenario-value.

Definition 6. (i) A value is *efficient (E)* if $\sum_{i\in N}\phi_i(v) = v(N\{N\})$.

ISBN: 978-989-95079-6-8

(ii) A scenario-value is *scenario-efficient (SE)* if $\sum_{i\in N}\phi_i^{\mathcal{S}}(v) = v(N\{N\})$ for all $\mathcal{S}\in\mathfrak{S}$.

Evidently, scenario-efficiency implies efficieny of the induced value.

Proposition 9. The Shapley scenario-value is scenario-efficient. Hence, the Shapley value is efficient as well.

A scenario-value satisfies *linearity (L)* if it is linear on $\mathcal{PG}(N)$.

Definition 7. Let us consider $i\in N$, a scenario $\mathcal{S}$, and denote by $S\pi$ the last element in $\mathcal{S}$ not containing i, and $S'\pi'$ its successor in $\mathcal{F}(\mathcal{S})$. Player i is *null in scenario* $\mathcal{S}$ *for* v if $v(S\pi) = v(S'\pi')$.

> **Null axiom (N):** If i is null in scenario $\mathcal{S}$ for v, then $\phi_i^{\mathcal{S}}(v) = 0$.

The *symmetry axiom for the scenario-value (SS)* reads: For any $i\in N$, any $\mathcal{S}\in\mathfrak{S}$, and any permutation σ on N, it holds

$$\phi_i^{\mathcal{S}}(v) = \phi_{\sigma(i)}^{\sigma(\mathcal{S})}(v\circ\sigma^{-1})$$

with $\sigma(\mathcal{S}), \sigma(S,\pi)$ defined naturally as follows: $\sigma(S,\pi) = (\sigma(S),\sigma(\pi))$, where $\sigma(S) = \{\sigma(i)\mid i\in S\}$, $\sigma(\pi) = \{\sigma(S)\mid S\in\pi\}$, and $\sigma(\mathcal{S}) = \{\sigma(S,\pi)\mid (S,\pi)\in\mathcal{S}\}$.

We given the following simple characterization.

Proposition 10. The Shapley scenario-value is the unique scenario-value satisfying (L), (N), (SS), and (SE).

Another axiomatization avoiding the strong (SE) axiom is given in [4].

5 The Choquet integral

We end the paper by giving some considerations on the Choquet integral defined for capacities on partitions. The (classical) Choquet integral in its discrete version is also based on maximal chains. Specifically, taking a function $f: N\to\mathbb{R}_+$, the maximal chain induced by f is given by

$$\{\sigma(n)\}, \{\sigma(n),\sigma(n-1)\},\ldots,\{\sigma(n),\ldots,\sigma(1)\}$$

where σ is any permutation such that

$$f_{\sigma(1)}\le f_{\sigma(2)}\le\cdots\le f_{\sigma(n)}$$

where $f(i)$ is denoted by f_i for simplicity. Then the Choquet integral of f is defined by:

$$\int f\,dv := \sum_{i=1}^{n}(f_{\sigma(i)} - f_{\sigma(i-1)})v(\{\sigma(n),\ldots,\sigma(i)\}),$$

with the convention $f_{\sigma(0)} = 0$. Let us keep this idea for capacities on partitions, and consider for simplicity a function f such that $f_1 < f_2 < \cdots < f_n$. Then there is only one maximal chain in $\mathfrak{C}(N)_\perp$ induced by f, containing the subsets $\{n\}, \{n,n-1\},\ldots,N$, which is

$$\{n\}\pi^\perp, \{n,n-1\}\pi^\perp_{\{n,n-1\}},\ldots,N\{N\}.$$

Hence the Choquet integral should be defined by:

$$\int f\,dv := \sum_{i=1}^{n}(f_i - f_{i-1})v(\{n,\ldots,i\}\{\{n,\ldots,i\},1,2,\ldots,i-1\}).$$

Consider now that $f_1 = f_2 < f_3\cdots < f_n$. Since two permutations can order this function, and the result should be the same whatever the permutation used, it is not difficult to check that this can be achieved only if the above formula is used, thus excluding the use of embedded coalitions which are not of the form $S\pi^\perp_S$. But then it suffices to put $\tilde{v}(S) := v(S\pi^\perp_S)$, and we are back to the classical Choquet integral.

As a conclusion, it does not seem to make sense to consider capacities on partitions for the Choquet integral.

References

[1] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[2] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[3] R. M. Thrall and W. F. Lucas. n-person games in partition function form. *Naval Research Logistic Quarterly*, 10:281–298, 1963.

[4] M. Grabisch and Y. Funaki. A coalition formation value for games in partition function form. submitted.

[5] I. Macho-Stadler, D. Pérez-Castillo, and D. Wettstein. Sharing the surplus: An extension of the Shapley value for environments with externalities. *J. of Economic Theory*, 135:339–356, 2007.

[6] J.-P. Barthélemy. Monotone functions on finite lattices: an ordinal approach to capacities, belief and necessity functions. In J. Fodor, B. De Baets, and P. Perny, editors, *Preferences and Decisions under Incomplete Knowledge*, pages 195–208. Physica Verlag, 2000.

[7] G. Birkhoff. *Lattice Theory*. American Mathematical Society, 3d edition, 1967.

[8] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:340–368, 1964.

[9] M. Aigner. *Combinatorial Theory*. Springer Verlag, 1979.

[10] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.

ISBN: 978-989-95079-6-8

The Minimization of the Risk of Falling in Portfolios under Uncertainty

Yuji Yoshida
Faculty of Economics and Business Administration, University of Kitakyushu
4-2-1 Kitagata, Kokuraminami, Kitakyushu 802-8577, Japan
Email: yoshida@kitakyu-u.ac.jp

Abstract— *A portfolio model to minimize the risk of falling under uncertainty is discussed. The risk of falling is represented by the value-at-risk of rate of return. Introducing the perception-based extension of the value-at-risk, this paper formulates a portfolio problem to minimize the risk of falling with fuzzy random variables. In the proposed model, randomness and fuzziness are evaluated respectively by the probabilistic expectation and the mean with evaluation weights and λ-mean functions. The analytical solutions of the portfolio problem regarding the risk of falling are derived. This paper gives formulae to show the explicit relations among the following important parameters in portfolio: The expected rate of return, the risk probability of falling and bankruptcy, and the rate of falling regarding the asset prices. A numerical example is given to explain how to obtain the optimal portfolio and these parameters from the asset prices in the stock market. Several figures are shown to observe the relation among these parameters at the optimal portfolios.*

Keywords— Value-at-risk (VaR), risk-sensitive portfolio, fuzzy random variable, perception-based extension, the risk probability, the rate of falling.

1 Introduction

In financial market, the portfolio is one of the most useful risk allocation technique for stable asset management. The minimization of the financial risk as well as the maximization of the return are important themes in the asset management. In a classical portfolio theory, *Markowitz's mean-variance model* is studied by many researchers and fruitful results have been achieved, and then the variance is investigated as the risk for portfolios ([8, 9, 11]). In this paper we focus on the drastic decline of asset prices. Recently, *value-at-risk (VaR)* is used widely to estimate the risk that asset prices decline based on worst scenarios. VaR is a risk-sensitive criterion based on percentiles, and it is one of the standard criteria in asset management ([17]). VaR is a kind of risk-level values of the asset at a specified probability of decline and it is used to select portfolios after due consideration of worst scenarios in investment. Many researchers and financial traders usually use VaR by mathematical programming since it is not easy to analyze the VaR portfolio model mathematically. Because Markowitz's mean-variance criterion and the variance-minimizing criterion are represented by quadratic programming, but VaR criterion in portfolio is neither linear nor quadratic ([17]). In this paper, by use of VaR regarding the rates of return, we discuss a *portfolio selection problem not only to minimize the rates of falling and but also to maximize the expected rates of return*. In the proposed portfolio model, we can maximize the expected rate of return after due consideration of the worst scenarios. This paper derives analytical solutions for the portfolio problem to minimize the rate of falling. The main purpose in this paper is to derive the explicit relations among the important parameters γ, p and δ regarding the financial risk from the obtained analytical results, where γ is the expected rate of return, p is the risk probability of falling and bankruptcy, and δ is the rate of falling regarding the asset at the optimal portfolios.

Soft computing like fuzzy logic works well for financial models in uncertain environment. Estimation of uncertain quantities is important in decision making. To represent uncertainty in this portfolio model, we use *fuzzy random variables* which have two kinds of uncertainties, i.e. randomness and fuzziness. In this paper, randomness is used to represent the uncertainty regarding the belief degree of frequency, and fuzziness is applied to linguistic imprecision of data because of a lack of knowledge regarding the current stock market. At the financial crisis in October 2008, we have observed the serious distrust of the market that the risky information regarding banks and security companies, for example the amounts of trouble loans, risky accounts, debts and so on, may not disclose to the investors and the public, and it is surely a kind of risks occurring from the imprecision of information. The fuzziness comes from the imprecision of data because of a lack of knowledge, and such serious distrust in the stock market will be represented by the fuzziness of information in finance models. We extend the VaR for real random variables to one regarding fuzzy random variables from the viewpoint of perception-based approach in Yoshida [14]. We formulate the a portfolio problem with fuzzy random variables, and we discuss the fundamental properties of the extended VaR. Recently, Yoshida [13] introduced the mean, the variance and the measurement of fuzziness of fuzzy random variables, using *evaluation weights* and λ-mean functions. This paper estimates fuzzy numbers and fuzzy random variables by the probabilistic expectation and these criteria, which are characterized by *possibility and necessity criteria* for subjective estimation and a *pessimistic-optimistic index* for subjective decision. These parameters are decided by the investor and are based on the degree of his certainty regarding the current information in the market. In this portfolio model, we use triangle-type fuzzy numbers and fuzzy random variables for computation in actual models. Regarding falling and bankruptcy, we observe the direct relations among the rate of falling, the expected rate of return and the risk probability at the optimal portfolios by figures derived from the obtained analytical results.

2 A portfolio model and the rate of falling

In this section, we explain a portfolio model with n stocks, where n is a positive integer. Let $\mathbb{T} := \{0, 1, 2, \cdots, T\}$ be the time space with an expiration date T, and $\mathbb{R}$ denotes the set of all real numbers. Let (Ω, P) be a probability space, where P is a non-atomic probability on a sample space Ω. For an asset $i = 1, 2, \cdots, n$, a *stock price process* $\{S_t^i\}_{t=0}^T$ is given

ISBN: 978-989-95079-6-8

by *rates of return* R_t^i as follows. Let

$$S_t^i := S_{t-1}^i(1 + R_t^i) \tag{1}$$

for $t = 1, 2, \cdots, T$, where $\{R_t^i\}_{t=1}^T$ is assumed to be an integrable sequence of independent real random variables. Hence $w_t = (w_t^1, w_t^2, \cdots, w_t^n)$ is called a *portfolio weight vector* if it satisfies $w_t^1 + w_t^2 + \cdots + w_t^n = 1$, and further a portfolio $(w_t^1, w_t^2, \cdots, w_t^n)$ is said to *allow for short selling* if $w_t^i \geq 0$ for all $i = 1, 2, \cdots, n$. Then the rate of return with a portfolio $(w_t^1, w_t^2, \cdots, w_t^n)$ is given by

$$R_t := w_t^1 R_t^1 + w_t^2 R_t^2 + \cdots + w_t^n R_t^n. \tag{2}$$

Therefore, the reward at time $t = 1, 2, \cdots, T$ follows

$$S_t := S_{t-1} \sum_{i=1}^{n} w_t^i (1 + R_t^i) = S_{t-1}(1 + R_t). \tag{3}$$

In this paper, we present a portfolio model where stock price processes S_t^i take fuzzy values using fuzzy random variables. The falling of asset prices is one of the most important risks in stock markets. In this section, we discuss a portfolio model where the risk is estimated by the rate of falling. Regarding the asset (3) with the portfolio w_t, the theoretical *bankruptcy* at time t occurs on scenarios ω satisfying $S_t(\omega) \leq 0$, i.e. it follows $1 + R_t(\omega) \leq 0$ from (3). Similarly, for a constant δ satisfying $0 \leq \delta \leq 1$, a set of sample paths

$$\{\omega \in \Omega | 1 + R_t(\omega) \leq 1 - \delta\} = \{\omega \in \Omega | R_t(\omega) \leq -\delta\} \tag{4}$$

is the event of scenarios where the asset price S_t will fall from the current price S_{t-1} to a lower level than $100(1 - \delta)\,\%$ of the current price S_{t-1}, i.e. the rate of falling is $100\,\delta\,\%$. Then we say shortly that the asset is $100\,\delta\,\%$*-falling* at time t and the parameter δ is called *the rate of falling*. The probability of $100\,\delta\,\%$-falling is also given by

$$p_\delta := P(R_t \leq -\delta). \tag{5}$$

For example, p_δ denotes the probability of the falling below par value if '$\delta = 0$' and it indicates the probability of the bankruptcy if '$\delta = 1$'. In this paper, we discuss portfolios to minimize the rate of falling δ.

For a positive probability p, a value-at-risk (VaR) regarding the rate of return R_t at the probability p is given by a real number v satisfying

$$P(R_t \leq v) = p \tag{6}$$

since P is non-atomic. The value-at-risk v is the upper bound of the rate of return R_t at the worst scenarios under a given risk probability p, and then the value-at-risk v in (6) is denoted by $\mathrm{VaR}_p(R_t)$. From (5) and (6), for a risk probability $p = p_\delta$, the rate of falling is

$$\delta = -\mathrm{VaR}_p(R_t). \tag{7}$$

To minimize the rate of falling (7) under a fuzzy and random environment, in next section we discuss the fundamental properties of value-at-risks.

3 A portfolio model with fuzzy random variables

In this section, we introduce fuzzy numbers and fuzzy random variables and we give a portfolio model under uncertainty. A fuzzy number is denoted by its membership function $\tilde{a} : \mathbb{R} \mapsto [0, 1]$ which is normal, upper-semicontinuous and quasi-concave and has a compact support ([18]). $\mathcal{R}$ denotes the set of all fuzzy numbers. In this paper, we identify fuzzy numbers with their corresponding membership functions. The α-cut of a fuzzy number $\tilde{a}(\in \mathcal{R})$ is given by $\tilde{a}_\alpha := \{x \in \mathbb{R} \mid \tilde{a}(x) \geq \alpha\}$ $(\alpha \in (0, 1])$ and $\tilde{a}_0 := \mathrm{cl}\{x \in \mathbb{R} \mid \tilde{a}(x) > 0\}$, where cl denotes the closure of an interval. We write the closed intervals as $\tilde{a}_\alpha := [\tilde{a}_\alpha^-, \tilde{a}_\alpha^+]$ for $\alpha \in [0, 1]$. Hence we also introduce a partial order $\succeq$, so called the *fuzzy max order*, on fuzzy numbers $\mathcal{R}$: Let $\tilde{a}, \tilde{b} \in \mathcal{R}$ be fuzzy numbers. Then, $\tilde{a} \succeq \tilde{b}$ means that $\tilde{a}_\alpha^- \geq \tilde{b}_\alpha^-$ and $\tilde{a}_\alpha^+ \geq \tilde{b}_\alpha^+$ for all $\alpha \in [0, 1]$. An addition, a subtraction and a scalar multiplication for fuzzy numbers are defined by Zadeh's extension principle as follows: For $\tilde{a}, \tilde{b} \in \mathcal{R}$ and $\xi \in \mathbb{R}$, the addition and subtraction $\tilde{a} \pm \tilde{b}$ of $\tilde{a}$ and $\tilde{b}$ and the scalar multiplication $\xi\tilde{a}$ of ξ and $\tilde{a}$ are fuzzy numbers given by their α-cuts $(\tilde{a}+\tilde{b})_\alpha := [\tilde{a}_\alpha^- + \tilde{b}_\alpha^-, \tilde{a}_\alpha^+ + \tilde{b}_\alpha^+]$, $(\tilde{a}-\tilde{b})_\alpha := [\tilde{a}_\alpha^- - \tilde{b}_\alpha^+, \tilde{a}_\alpha^+ - \tilde{b}_\alpha^-]$ and $(\xi\tilde{a})_\alpha := [\xi\tilde{a}_\alpha^-, \xi\tilde{a}_\alpha^+]$ if $\xi \geq 0$.

A fuzzy-number-valued map $\tilde{X} : \Omega \mapsto \mathcal{R}$ is called a *fuzzy random variable* if the maps $\omega \mapsto \tilde{X}_\alpha^-(\omega)$ and $\omega \mapsto \tilde{X}_\alpha^+(\omega)$ are measurable for all $\alpha \in (0, 1]$, where $\tilde{X}_\alpha(\omega) = [\tilde{X}_\alpha^-(\omega), \tilde{X}_\alpha^+(\omega)] = \{x \in \mathbb{R} \mid \tilde{X}(\omega)(x) \geq \alpha\}$ ([7, 10]). We need to introduce expectations of fuzzy random variables in order to describe a portfolio model. A fuzzy random variable $\tilde{X}$ is said to be integrably bounded if both $\omega \mapsto \tilde{X}_\alpha^-(\omega)$ and $\omega \mapsto \tilde{X}_\alpha^+(\omega)$ are integrable for all $\alpha \in (0, 1]$. Let $\tilde{X}$ be an integrably bounded fuzzy random variable. The expectation $E(\tilde{X})$ of the fuzzy random variable $\tilde{X}$ is defined by a fuzzy number $E(\tilde{X})(x) := \sup_{\alpha \in [0,1]} \min\{\alpha, 1_{E(\tilde{X})_\alpha}(x)\}$ for $x \in \mathbb{R}$, where $E(\tilde{X})_\alpha := [\int_\Omega \tilde{X}_\alpha^-(\omega)\,\mathrm{d}P(\omega), \int_\Omega \tilde{X}_\alpha^+(\omega)\,\mathrm{d}P(\omega)]$ for $\alpha \in (0, 1]$ ([6, 10]).

Now we deal with a case where the rate of return $\{R_t^i\}_{t=1}^T$ has some imprecision. In this paper, we use triangle-type fuzzy random variables for computation, however we can apply the similar approach to general fuzzy random variables. We define a *rate of return process with imprecision* $\{\tilde{R}_t^i\}_{t=0}^T$ by a sequence of triangle-type fuzzy random variables

$$\tilde{R}_t^i(\cdot)(x) = \begin{cases} 0 & \text{if } x < R_t^i - c_t^i \\ \frac{x - R_t^i + c_t^i}{c_t^i} & \text{if } R_t^i - c_t^i \leq x < R_t^i \\ \frac{x - R_t^i - c_t^i}{-c_t^i} & \text{if } R_t^i \leq x < R_t^i + c_t^i \\ 0 & \text{if } x \geq R_t^i + c_t^i, \end{cases} \tag{8}$$

where c_t^i is a positive number. We call c_t^i a *fuzzy factor* for asset i at time t. Hence we can represent $\tilde{R}_t^i$ by the sum of the real random variable R_t^i and a fuzzy number $\tilde{a}_t^i$:

$$\tilde{R}_t^i(\omega)(\cdot) := 1_{\{R_t^i(\omega)\}}(\cdot) + \tilde{a}_t^i(\cdot) \tag{9}$$

for $\omega \in \Omega$, where $1_{\{\cdot\}}$ denotes the characteristic function of a singleton and $\tilde{a}_t^i$ is a triangle-type fuzzy number defined by

$$\tilde{a}_t^i(x) = \begin{cases} 0 & \text{if } x < -c_t^i \\ \frac{x + c_t^i}{c_t^i} & \text{if } -c_t^i \leq x < 0 \\ \frac{x - c_t^i}{-c_t^i} & \text{if } 0 \leq x < c_t^i \\ 0 & \text{if } x \geq c_t^i. \end{cases} \tag{10}$$

ISBN: 978-989-95079-6-8

For assets $i = 1, 2, \cdots, n$, we define *stock price processes* $\{\tilde{S}_t^i\}_{t=0}^T$ by the *rates of return with imprecision* $\tilde{R}_t^i$ as follows: $S_0^i := S_0^i$ is a positive number and

$$\tilde{S}_t^i = \tilde{S}_0^i \prod_{s=1}^{t} (1 + \tilde{R}_s^i) \tag{11}$$

for $t = 1, 2, \cdots, T$. For a portfolio $w = (w^1, w^2, \cdots, w^n)$, the rate of return with imprecision is given by a linear combination of fuzzy random variables

$$\tilde{R}_t := w^1 \tilde{R}_t^1 + w^2 \tilde{R}_t^2 + \cdots + w^n \tilde{R}_t^n. \tag{12}$$

In Section 4 we investigate the value-at-risk to apply the fuzzy random variable (12), and in Section 5 we discuss the portfolio problem to minimize the rate of falling regarding (12).

4 A perception-based extension of VaR

First we introduce mathematical notations of the value-at-risk for real random variables to apply it to the rates of return (12). Let $\mathcal{X}$ be the set of all integrable real random variables X on Ω with a continuous distribution function $x \mapsto F_X(x) := P(X < x)$ for which there exists a non-empty open interval I such that $F_X(\cdot) : I \mapsto (0, 1)$ is a strictly increasing and onto. Then there exists a strictly increasing and continuous inverse function $F_X^{-1} : (0, 1) \mapsto I$. We note that $F_X(\cdot) : I \mapsto (0, 1)$ and $F_X^{-1} : (0, 1) \mapsto I$ are one-to-one and onto, and we put $F_X(\inf I) := \lim_{x \downarrow \inf I} F_X(x) = 0$ and $F_X(\sup I) := \lim_{x \uparrow \sup I} F_X(x) = 1$. Then, the *value-at-risk*, shortly for *VaR*, at a risk probability p is given by the $100\,p$-percentile of the distribution function F_X: For a probability p $(0 < p < 1)$,

$$\mathrm{VaR}_p(X) := \sup\{x \in I \mid F_X(x) \le p\}, \tag{13}$$

and then we have $F_X(\mathrm{VaR}_p(X)) = p$ and $\mathrm{VaR}_p(X) = F_X^{-1}(p)$ for $0 < p < 1$. In this paper, we assume that VaR v in (6) has the following representation (14).

$$\begin{aligned}(\text{VaR } v) \quad &= (\text{the mean}) - (\text{a positive constant } \kappa) \\ &\quad \times (\text{the standard deviation}),\end{aligned} \tag{14}$$

where the positive constant κ is given corresponding to the probability p. The details are as follows: For any probability p satisfying $0 < p < 1$, there exists a positive constant κ such that a real number $v := \mu_t - \kappa\,\sigma_t$ satisfies (14) for all portfolios, where μ_t and σ_t are the expectation and the standard deviation of R_t respectively. One of the most popular sufficient condition for (14) is what the distribution of the rate of return R_t is normal ([2, 17]). In this paper, we obtain an exact estimation of κ in Theorem 1.

Let $\tilde{\mathcal{X}}$ be the set of all fuzzy random variables $\tilde{X}$ on Ω such that their α-cuts $\tilde{X}_\alpha^\pm$ are integrable and $\lambda\tilde{X}_\alpha^- + (1-\lambda)\tilde{X}_\alpha^+ \in \mathcal{X}$ for all $\lambda \in [0, 1]$ and $\alpha \in [0, 1]$. Hence, from (8) we introduce a value-at-risk for a fuzzy random variable $\tilde{X}(\in \tilde{\mathcal{X}})$ at a positive risk probability p as follows.

$$\mathrm{VaR}_p(\tilde{X})(x) := \sup_{X \in \mathcal{X} \,:\, \mathrm{VaR}_p(X) = x} \inf_{\omega \in \Omega} \tilde{X}(\omega)(X(\omega)) \tag{15}$$

for $x \in \mathbb{R}$. Yoshida [14] has studied *perception-based estimations* extending the concept of the expectations in Kruce and Meyer [6]. This definition (15) is an extension from the value-at-risk on real random variables to one on fuzzy random variables based on the perception. The value-at-risk (15) on fuzzy random variables is characterized by the following representation, which is from the continuity and the monotonicity of $\mathrm{VaR}_p(\cdot)$.

Lemma 1 ([14]). *Let $\tilde{X} \in \tilde{\mathcal{X}}$ be a fuzzy random variable and let p be a positive probability. Then the value-at-risk $\mathrm{VaR}_p(\tilde{X})$ defined by (15) is a fuzzy number whose α-cuts are*

$$\mathrm{VaR}_p(\tilde{X})_\alpha = [\mathrm{VaR}_p(\tilde{X}_\alpha^-), \mathrm{VaR}_p(\tilde{X}_\alpha^+)] \tag{16}$$

for $\alpha \in (0, 1]$.

The value-at-risk (15) on fuzzy random variables has the following properties.

Lemma 2 ([14]). *Let $\tilde{X}, \tilde{Y} \in \tilde{\mathcal{X}}$ be fuzzy random variables and let p be a positive probability. Then the value-at-risk VaR_p defined by (15) has the following properties:*

(i) *If $\tilde{X} \preceq \tilde{Y}$, then $\mathrm{VaR}_p(\tilde{X}) \preceq \mathrm{VaR}_p(\tilde{Y})$.*

(ii) *$\mathrm{VaR}_p(\zeta\tilde{X}) = \zeta\,\mathrm{VaR}_p(\tilde{X})$ for $\zeta > 0$.*

(iii) *$\mathrm{VaR}_p(\tilde{X} + \tilde{a}) = \mathrm{VaR}_p(\tilde{X}) + \tilde{a}$ for a fuzzy number $\tilde{a} \in \mathcal{R}$.*

Next since the value-at-risk $\mathrm{VaR}_p(\tilde{R}_t)$ for the rate of return (12) with a portfolio is a fuzzy number, we need to evaluate the fuzziness of fuzzy numbers and fuzzy random variables. There are many studies regarding the evaluation of fuzzy numbers. Two major approaches of them are as follows. One is to use weighting functions([1, 3, 13]) and the other is to use possibility and necessity criteria([4]). Here we adopt the former evaluation method of fuzzy numbers and fuzzy random variables. In the rest of this section we introduce the definitions from [13, 15], and in the next section we estimate the VaR regarding the rate of return (12) by the evaluation method. Yoshida [13] has studied an evaluation of fuzzy numbers by *evaluation weights* which are induced from fuzzy measures to evaluate a confidence degree that a fuzzy number takes values in an interval. With respect to fuzzy random variables, the randomness is evaluated by the probabilistic expectation and the fuzziness is estimated by the evaluation weights and the following function. Let $g^\lambda : \mathcal{I} \mapsto \mathbb{R}$ be a map such that

$$g^\lambda([x, y]) := \lambda x + (1 - \lambda)y, \quad [x, y] \in \mathcal{I}, \tag{17}$$

where λ is a constant satisfying $0 \le \lambda \le 1$ and $\mathcal{I}$ denotes the set of all bounded closed intervals. This scalarization is used for the estimation of fuzzy numbers to give a mean value of the interval $[x, y]$ with a weight λ, and g^λ is called a *λ-mean function* and λ is called a *pessimistic-optimistic index* which indicates the pessimistic degree of attitude in decision making ([3]). Let a fuzzy number $\tilde{a} \in \mathcal{R}$. A mean value of the fuzzy number $\tilde{a}$ with respect to λ-mean functions g^λ and an evaluation weight $w(\alpha)$, which depends only on $\tilde{a}$ and α, is given as follows ([15]):

$$\tilde{E}(\tilde{a}) := \frac{\int_0^1 g^\lambda(\tilde{a}_\alpha)\, w(\alpha)\, \mathrm{d}\alpha}{\int_0^1 w(\alpha)\, \mathrm{d}\alpha}, \tag{18}$$

where $\tilde{a}_\alpha = [\tilde{a}_\alpha^-, \tilde{a}_\alpha^+]$ is the α-cut of the fuzzy number $\tilde{a}$. In (18), $w(\alpha)$ indicates a *confidence degree that the fuzzy*

ISBN: 978-989-95079-6-8

number $\tilde{a}$ takes values in the interval $\tilde{a}_\alpha$ at each level α. Hence, an evaluation weight $w(\alpha)$ is called the *possibility evaluation weight* $w^P(\alpha)$ if $w^P(\alpha) := 1$ for $\alpha \in [0,1]$, and $w(\alpha)$ is called the *necessity evaluation weight* $w^N(\alpha)$ if $w^N(\alpha) := 1-\alpha$ for $\alpha \in [0,1]$. Especially, for a fuzzy number $\tilde{a} \in \mathcal{R}$, the means in the possibility and necessity cases are represented respectively by $\tilde{E}^P(\tilde{a})$ and $\tilde{E}^N(\tilde{a})$, and we consider their combination $\nu\tilde{E}^P(\tilde{a}) + (1-\nu)\tilde{E}^P(\tilde{a})$ with a parameter $\nu \in [0,1]$ ([13, 14, 16]). The mean $\tilde{E}$ has the following natural properties of the linearity and the monotonicity regarding the fuzzy max order.

Lemma 3 ([13, 15]). *Let $\lambda \in [0,1]$. For fuzzy numbers $\tilde{a}, \tilde{b} \in \mathcal{R}$ and real numbers θ, ζ, the following (i) – (iv) hold.*

(i) $\tilde{E}(\tilde{a} + 1_{\{\theta\}}) = \tilde{E}(\tilde{a}) + \theta$.

(ii) $\tilde{E}(\zeta\tilde{a}) = \zeta\tilde{E}(\tilde{a})$ *if* $\zeta \geq 0$.

(iii) $\tilde{E}(\tilde{a} + \tilde{b}) = \tilde{E}(\tilde{a}) + \tilde{E}(\tilde{b})$.

(iv) *If* $\tilde{a} \succeq \tilde{b}$, *then* $\tilde{E}(\tilde{a}) \geq \tilde{E}(\tilde{b})$ *holds.*

For a fuzzy random variable $\tilde{X}$, *the mean of the expectation* $E(\tilde{X})$ is a real number

$$E(\tilde{E}(\tilde{X})) = E\left(\frac{\int_0^1 g^\lambda(\tilde{X}_\alpha)\, w(\alpha)\, \mathrm{d}\alpha}{\int_0^1 w(\alpha)\, \mathrm{d}\alpha}\right). \tag{19}$$

Then, from Lemma 3, we obtain the following results regarding fuzzy random variables.

Lemma 4 ([13, 15]). *Let $\lambda \in [0,1]$. For a fuzzy number $\tilde{a} \in \mathcal{R}$, integrable fuzzy random variables $\tilde{X}, \tilde{Y}$, an integrable real random variable Z and a nonnegative real number ζ, the following (i) – (v) hold.*

(i) $E(\tilde{E}(\tilde{X})) = \tilde{E}(E(\tilde{X}))$.

(ii) $E(\tilde{E}(\tilde{a})) = \tilde{E}(\tilde{a})$ *and* $E(\tilde{E}(Z)) = E(Z)$.

(iii) $E(\tilde{E}(\zeta\tilde{X})) = \zeta E(\tilde{E}(\tilde{X}))$.

(iv) $E(\tilde{E}(\tilde{X} + \tilde{Y})) = E(\tilde{E}(\tilde{X})) + E(\tilde{E}(\tilde{Y}))$.

(v) *If* $\tilde{X} \succeq \tilde{Y}$, *then* $E(\tilde{E}(\tilde{X})) \geq E(\tilde{E}(\tilde{Y}))$ *holds.*

5 The Minimization of the Risk of Falling

In this section, we discuss portfolio problems under uncertainty. First we estimate the rate of return with imprecision for a portfolio. Let the mean, the variance and the covariance of the rate of return R_t^i, which are the real random variables in (2.1), by $\mu_t^i := E(R_t^i)$, $(\sigma_t^i)^2 := E((R_t^i - \mu_t^i)^2)$, and $\sigma_t^{ij} := E((R_t^i - \mu_t^i)(R_t^j - \mu_t^j))$ for $i, j = 1, 2, \cdots, n$. Hence we assume that the determinant of the variance-covariance matrix $\Sigma := [\sigma_t^{ij}]$ is not zero and there exists its inverse matrix Σ^{-1}. This assumption is natural and it can be realized easily by taking care of the combinations of assets. For a portfolio $w = (w^1, w^2, \cdots, w^n)$ satisfying $w^1 + w^2 + \cdots + w^n = 1$ and $w^i \geq 0$ $(i = 1, 2, \cdots, n)$, we calculate the expectation and the variance regarding $\tilde{R}_t = w^1\tilde{R}_t^1 + w^2\tilde{R}_t^2 + \cdots + w^n\tilde{R}_t^n$. From Lemma 4, the expectation $\tilde{\mu}_t := E(\tilde{E}(\tilde{R}_t))$ follows

$$\tilde{\mu}_t = E(\tilde{E}(\tilde{R}_t)) = \sum_{i=1}^n w^i E(\tilde{E}(\tilde{R}_t^i)) = \sum_{i=1}^n w^i \tilde{\mu}_t^i, \tag{20}$$

where $\tilde{\mu}_t^i := E(\tilde{E}(\tilde{R}_t^i))$ for $i = 1, 2, \cdots, n$. On the other hand, regarding this model, from [13] we can find that the variance $(\tilde{\sigma}_t)^2 := E((\tilde{E}(\tilde{R}_t) - \tilde{\mu}_t)^2)$ of $\tilde{R}_t$ equals to the variance $(\sigma_t)^2 := E((R_t - \mu_t)^2)$ of R_t:

$$(\tilde{\sigma}_t)^2 = (\sigma_t)^2 = \sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij}. \tag{21}$$

By (14), (20) and (21), the mean of $\mathrm{VaR}_p(\tilde{R}_t)$ is

$$\tilde{E}(\mathrm{VaR}_p(\tilde{R}_t)) = \sum_{i=1}^n w^i \tilde{\mu}_t^i - \kappa\sqrt{\sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij}} \tag{22}$$

with a positive constant κ. Now step by step we discuss a portfolio problem to minimize the rate of falling $\delta = -\tilde{E}(\mathrm{VaR}_p(\tilde{R}_t))$. First, we deal with a variance-minimizing model. For a given constant γ, which is the minimum expected rate of return to be guaranteed for the portfolio, we consider the following quadratic programming with respect to portfolios with allowance for short selling.

Variance-minimizing problem (P1): Minimize the variance

$$\sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij} \tag{23}$$

with respect to portfolios $w = (w^1, w^2, \cdots, w^n)$ satisfying $w^1 + w^2 + \cdots + w^n = 1$ under the following condition regarding the expected rate of return: $\sum_{i=1}^n w^i \tilde{\mu}_t^i = \gamma$.

We start from the following classical results regarding the variance-minimizing problem (P1), and we analyze the portfolio problem to minimize the rate of falling.

Lemma 5 ([15, Theorem 1]). *The solution of the variance-minimizing problem (P1) is given by*

$$w = \xi\Sigma^{-1}\mathbf{1} + \eta\Sigma^{-1}\tilde{\mu} \tag{24}$$

and then the corresponding variance is

$$\tilde{\rho} := \frac{A\gamma^2 - 2B\gamma + C}{\Delta}, \tag{25}$$

where $\tilde{\mu}^i := \mu_t^i + \tilde{E}^\lambda(\tilde{a}_t^i)$, $\tilde{\sigma}^{ij} := \sigma_t^{ij}$ $(i, j = 1, 2, \cdots, n)$, $\tilde{\Sigma} := [\tilde{\sigma}^{ij}]$, $\tilde{\mu} := [\tilde{\mu}^1\ \tilde{\mu}^2\ \cdots\ \tilde{\mu}^n]^{\mathrm{T}}$, $\mathbf{1} := [1\ 1\ \cdots\ 1]^{\mathrm{T}}$, $\xi := \frac{C - B\gamma}{\Delta}$, $\eta := \frac{A\gamma - B}{\Delta}$, $A := \mathbf{1}^{\mathrm{T}}\tilde{\Sigma}^{-1}\mathbf{1}$, $B := \mathbf{1}^{\mathrm{T}}\tilde{\Sigma}^{-1}\tilde{\mu}$, $C := \tilde{\mu}^{\mathrm{T}}\tilde{\Sigma}^{-1}\tilde{\mu}$, $\Delta := AC - B^2$ and $^{\mathrm{T}}$ denotes the transpose of a vector.

Hence, we consider a risk-sensitive model, which is not of mean-variance types but of mean-standard deviation types, in order to deal with a portfolio problem to minimize the rate of falling in the third step. For a constant γ and a positive constant κ, we discuss the following risk-sensitive portfolio problem with allowance for short selling.

Risk-sensitive problem (P2): Maximize a risk-sensitive expected rate of return

$$v(\gamma) := \sum_{i=1}^n w^i \tilde{\mu}_t^i - \kappa\sqrt{\sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij}} \tag{26}$$

ISBN: 978-989-95079-6-8

with respect to portfolios $w = (w^1, w^2, \cdots, w^n)$ $(w^1 + w^2 + \cdots + w^n = 1)$ under the condition $\sum_{i=1}^n w^i \tilde{\mu}_t^i = \gamma$.

Now we discuss the following VaR portfolio problem without allowance for short selling. The following form (27) comes from the value-at-risk $\tilde{E}(\mathrm{VaR}_p(\tilde{R}_t))$ given in (22).

Portfolio problem to minimize the rate of falling (P3):
Minimize the risk of falling

$$\delta = -\sum_{i=1}^n w^i \tilde{\mu}_t^i + \kappa \sqrt{\sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij}} \qquad (27)$$

with respect to portfolios $w = (w^1, w^2, \cdots, w^n)$ satisfying $w^1 + w^2 + \cdots + w^n = 1$ and $w^i \geq 0$ for $i = 1, 2, \cdots, n$.

Since we have

$$\inf_w (27) = \inf_\gamma \left(\inf_{w:\sum_{i=1}^n w^i \tilde{\mu}_t^i = \gamma} (27) \right) = -\sup_\gamma (26),$$

in the same way as [16] we arrive at the following analytical solutions of the portfolio problem to minimize the rate of falling (P3).

Lemma 5. *Let A and Δ be positive. Let κ satisfy $\kappa^2 > C$. The solution of the portfolio problem to minimize the rate of falling (P3) is given by $w^* = \xi\Sigma^{-1}\mathbf{1} + \eta\Sigma^{-1}\tilde{\mu}$, and then the corresponding rate of falling is $\delta(\gamma^*) = -\frac{B-\sqrt{A\kappa^2-\Delta}}{A}$ at the expected rate of return $\gamma^* := \frac{B}{A} + \frac{\Delta}{A\sqrt{A\kappa^2-\Delta}}$, where $\xi := \frac{C-B\gamma^*}{\Delta}$ and $\eta := \frac{A\gamma^*-B}{\Delta}$. Further, if $\Sigma^{-1}\mathbf{1} \geq \mathbf{0}$ and $\Sigma^{-1}\tilde{\mu} \geq \mathbf{0}$, then the portfolio (5.18) satisfies $w^* \geq \mathbf{0}$, i.e. the portfolio w^* is a portfolio without allowance for short selling. Here, $\mathbf{0}$ denotes the zero vector.*

In Lemma 5, we note that the optimal portfolio w^* not only to minimize the rate of falling $\delta(\gamma^*)$ but also to maximize the expected rate of return γ^*.

Theorem 1. *Let A and Δ be positive. Let δ satisfy $\delta > -2B/A$. Then, the assumptions in Lemma 5 are satisfied, and the following (i) and (ii) hold for the optimal portfolio in Lemma 5.*

(i) *For a rate of falling δ, the corresponding constant κ_δ and the expected rate of return γ_δ are given by*

$$\kappa_\delta := \sqrt{A\delta^2 + 2B\delta + C} \quad \text{and} \quad \gamma_\delta := \frac{B\delta + C}{A\delta + B}. \qquad (28)$$

Then the risk probability is $p_\delta = P(\tilde{E}(\tilde{R}_t) \leq -\delta)$.

(ii) *If R_t^i $(i = 1, 2, \cdots, n)$ have normal distributions, the risk probability p_δ in (i) is given by*

$$p_\delta := \Phi(-\kappa_\delta), \qquad (29)$$

where κ_δ is defined by (28) and Φ is the cumulative normal distribution function $\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}}\, dt$ for $z \in \mathbb{R}$.

Remark 1 From Theorem 1, for the optimal portfolio of (P3), the relations among the expected rate of return $\gamma = \gamma_\delta$, the rate of falling δ and the risk probability $p = p_\delta$ are illustrated as follows.

$$\left.\begin{array}{lcl} \gamma & : & \text{the expected rate of return} \\ \Updownarrow & \gamma = & \dfrac{B\delta + C}{A\delta + B} \\ \delta & : & \text{the rate of falling}(= -\mathrm{VaR}) \\ \Updownarrow & p = & \Phi(-\sqrt{A\delta^2 + 2B\delta + C}) \\ p & : & \text{the risk probability of falling} \end{array}\right\} \qquad (30)$$

These results hold for the crisp case ($c_t^i = 0$) and they will be useful not only for analytic study but also for actual management in finance. In the next section, by figures with a numerical example we show the relations (30) among the expected rate of return γ, the rate of falling δ and the risk probability p.

6 A numerical example

In this session, we give a simple example to illustrate our idea. For the numerical computation, we need to evaluate fuzzy numbers representing the rates of return (8). From [13, 14], we have evaluations of the fuzzy numbers $\tilde{a}_t^i$ by a combination $\nu\tilde{E}^P(\tilde{a}) + (1-\nu)\tilde{E}^P(\tilde{a})$ with a parameter $\nu \in [0, 1]$ with a parameter $\nu(\in [0, 1])$, which is called a *possibility-necessity weight* ([13]). Then from (20) we obtain the expected rate of return

$$\tilde{\mu}_t = \sum_{i=1}^n w^i \left(\mu_t^i + \frac{(1-2\lambda)(4-\nu)}{6} c_t^i \right) \qquad (31)$$

for the possibility-necessity weight $\nu(\in [0, 1])$ and the pessimistic-optimistic index $\lambda(\in [0, 1])$. In (31), the decision maker may choose the parameters $\lambda(\in [0, 1])$ and $\nu(\in [0, 1])$. The pessimistic-optimistic index is taken as $\lambda = 1$ if he has pessimistic personal forecast in the market and he takes careful decision, and $\lambda = 0$ if he has optimistic personal forecast and he is not nervous. The possibility-necessity weight is taken as $\nu = 1$ when he has enough confidential information about the market, and $\nu = 0$ when he does not have confidential information. In this model, $\nu = 0$ is reasonable since our objective function is VaR, which is a kind of risk, and we need to take into account of the fuzziness of information in the market. While λ depends on the decision maker's attitude in his investment. In this example, we compute the pessimistic case $\lambda = 1$.

Let $n = 4$ be the number of assets. Take the expected rate of return, a variance-covariance matrix and fuzzy factors as Table 1. We assume that the rate of return R_t^i has the normal distributions. We discuss a risk probability 1% in the normal distribution, and then the corresponding constant is $\kappa = 2.326$, which is given in (14) for VaR. Then, the conditions in Theorem 1 are satisfied and by formulae of Lemma 5 we easily obtain the optimal portfolio $w^* = (w^1, w^2, w^3, w^4) = (0.229604, 0.215551, 0.252000, 0.302845)$ for the portfolio problem to minimize the rate of falling (P3), and then the corresponding rate of falling is $\delta(\gamma^*) = 0.557042$ and the expected rate of return is $\gamma^* = 0.0498026$.

Owing to the equations (28) and (29), we can easily observe the explicit relations among the expected rate of return γ, the rate of falling δ and the risk probability p (Remark 1). Regarding this example, Fig.1 shows the relation among the rate of

ISBN: 978-989-95079-6-8

falling δ, the expected rate of return γ and the risk probability p. Figs.1 and 2 are drawn based on Remark 1. Especially Fig.2 shows the relation among the risk probability p such that the falling in the asset prices will be lower than $100(1-\delta)\,\%$, the rate of falling δ and the expected rate of return γ. For example, for a rate of falling $\delta = 0.2$, which implies that the asset price will fall lower than $80\,\%$, its risk probability is $p = 0.168993$, however we have the expected rate of return $\gamma = 0.0504023$ with $\kappa = 0.958154$, which are calculated easily from Remark 1 based on Theorem 1. We can also compute the cases of falling below par value and the complete bankruptcy, taking $\delta = 0$ and $\delta = 1$ respectively. The probability of falling below par value is $p_0 = 0.4211443$ and then the expected rate of return $\gamma_0 = 0.0545276$, and the probability of complete bankruptcy is $p_1 = 0.0000285992$ and then the expected rate of return $\gamma_1 = 0.0496258$.

Table 1: Expected rates of return, a variance-covariance matrix and fuzzy factors.

Asset	μ_t^i
w_1	0.05
w_2	0.07
w_3	0.06
w_4	0.04

Asset	c_t^i
w_1	0.006
w_2	0.008
w_3	0.007
w_4	0.005

σ_t^{ij}	w_1	w_2	w_3	w_4
w_1	0.35	0.03	0.02	-0.08
w_2	0.03	0.25	-0.06	0.08
w_3	0.02	-0.06	0.33	-0.02
w_4	-0.08	0.08	-0.02	0.24

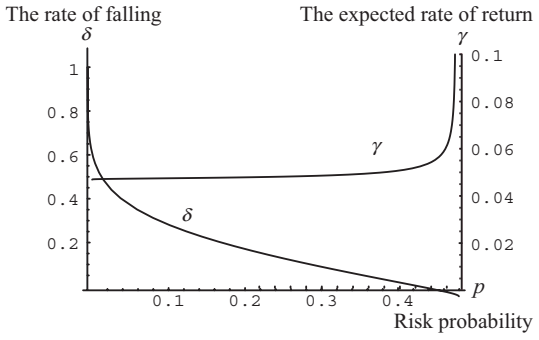


Figure 1: The rate of falling δ and the expected rate of return γ at risk probabilities p.

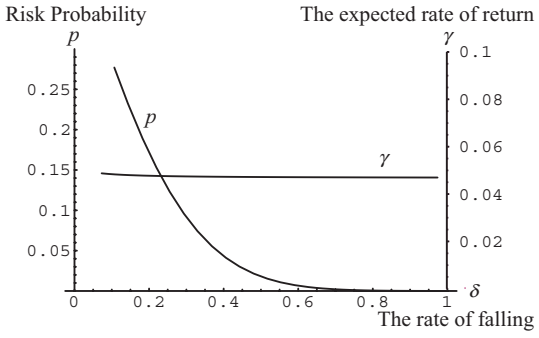


Figure 2: Risk probability p and the expected rate of return γ at the rates of falling δ.

7 Conclusion

We have derived the relation (30) among the expected rate of return γ, the rate of falling δ and the risk probability p. The relation holds for the crisp case and it will be useful not only for analytic study but also for actual management in finance.

References

[1] L.Campos and A.Munoz, A subjective approach for ranking fuzzy numbers, *Fuzzy Sets and Systems*, 29:145-153, 1989.

[2] L.El Chaoui, M.Oks and F.Oustry, Worst-case value at risk and robust portfolio optimization: A conic programming approach, *Operations Research*, 51:543-556, 2003.

[3] P. Fortemps and M. Roubens, Ranking and defuzzification methods based on area compensation, *Fuzzy Sets and Systems*, 82:319-330, 1996.

[4] M.Inuiguchi and T.Tanino, Portfolio selection under independent possibilistic information, *Fuzzy Sets and Systems*, 115:83-92, 2000.

[5] P.Jorion, *Value at Risk: Tne New Benchmark for Managing Financial Risk*, Third Edition McGraw-Hill, New York, 2007.

[6] R.Kruse and K.D.Meyer, *Statistics with Vague Data*, Riedel Publ. Co., Dortrecht, 1987.

[7] H.Kwakernaak, Fuzzy random variables – I. Definitions and theorem, *Inform. Sci.*, 15:1-29, 1978.

[8] H.Markowitz, *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Oxford, 1990.

[9] S.R.Pliska, *Introduction to Mathematical Finance: Discrete-Time Models*, Blackwell Publ., New York, 1997.

[10] M.L.Puri and D.A.Ralescu, Fuzzy random variables, *J. Math. Anal. Appl.*, 114:409-422, 1986.

[11] M.C.Steinbach, Markowitz revisited: Mean-variance model in financial portfolio analysis, *SIAM Review*, 43:31-85, 2001.

[12] H.Tanaka, P.Guo and L.B.Türksen, Portfolio selection based on fuzzy probabilities and possibilistic distribution, *Fuzzy Sets and Systems*, 111:397-397, 2000.

[13] Y.Yoshida, Mean values, measurement of fuzziness and variance of fuzzy random variables for fuzzy optimization, *Proceedings of SCIS & ISIS 2006*, 2277-2282, Sept. 2006.

[14] Y.Yoshida, A risk-minimizing model under uncertainty in portfolio, in: V.Torra et al. eds., *MDAI 2007*, LNAI 4529, 295-306, Springer, August, 2007.

[15] Y.Yoshida, Fuzzy extension of estimations with randomness: The perception-based approach, in: P.Melin et al. eds., *IFSA2007*, LNAI 4617, 381-391, Springer, Sept., 2007.

[16] Y.Yoshida, An estimation model of value-at-risk portfolio under uncertainty, to appear in *Fuzzy Sets and Systems*, doi:10.1016/j.fss2009.02.007.

[17] Z.Zmeškal, Value at risk methodology of international index portfolio under soft conditions (fuzzy-stochastic approach), *International Review of Financial Analysis*, 14:263-275, 2005.

[18] L.A.Zadeh, Fuzzy sets, *Inform. and Control*, 8:338-353, 1965.

ISBN: 978-989-95079-6-8

Approximations by interval, triangular and trapezoidal fuzzy numbers

Chi-Tsuen Yeh

Department of Mathematics Education, National University of Tainan
33, Sec. 2, Shu-Lin St., Tainan city 70005, Taiwan
Email: ctyeh@mail.nutn.edu.tw

Abstract— *Recently, many scholars investigated interval, triangular, and trapezoidal approximations of fuzzy numbers. These researches can be grouped into two classes: the Euclidean distance class and the non-Euclidean distance class. Most approximations in the Euclidean distance class can be calculated by formulas, but calculating approximations in the other class is more complicated. In this paper, we study interval, triangular, and trapezoidal approximations under a weighted Euclidean distance which generalize all approximations in the Euclidean distance class. First, we embed fuzzy numbers into a Hilbert space, and then introduce these weighted approximations by means of best approximations from closed convex subsets of the Hilbert space. Finally, we apply the reduction principle to simplify calculations of these approximations.*

Keywords— weighted trapezoidal approximation, triangular fuzzy number, Hilbert space

1 Introduction

In practice, fuzzy intervals are often used to represent uncertain or incomplete information. An interesting problem is to approximate general fuzzy intervals by interval, triangular, and trapezoidal fuzzy numbers, so as to simplify calculations. Recently, many scholars investigated these approximations of fuzzy numbers. According to the different aspects of distance, these researches can be grouped into two classes: the Euclidean distance class and the non-Euclidean distance class. The Euclidean distance class includes the interval approximation (proposed by Grzegorzewski in 2002 [11]), symmetric triangular approximation (proposed by Ma et al. in 2000 [18]), trapezoidal approximation (proposed by Abbasbandy and Asady in 2004 [1]), and weighted triangular approximation (proposed by Zeng and Li in 2007 [23]). The non-Euclidean distance class includes the rectangle approximation under the Hamming distance (proposed by Chanas in 2001 [6]), symmetric and non-symmetrical trapezoidal approximations under the Euclidean distance between the respective 1/2-levels (proposed by Delgado et al. in 1998 [7]), and trapezoidal approximation under the source distance (proposed by Abbasbandy and Amirfakhrian in 2006 [2]). Some other approximations are also investigated, such as the nearest parametric approximation (proposed by Nasibova and Peker in 2008 [19]), trapezoidal approximation preserving the expected interval (proposed by Grzegorzewski and Mrówka [12, 13, 14], and improved by Ban [5] and Yeh [22] in 2008, independently), approximation by π functions (proposed by Hassine et al. in 2006 [15]), and polynomial approximation (proposed by Abbasbandy and Amirfakhrian in 2006 [3]). Most approximations in the Euclidean distance class can be calculated by formulas, but calculating the approximations in the other class is more complicated. In this paper, we study interval, triangular, and trapezoidal approximations under a weighted Euclidean distance which generalize all approximations in the Euclidean distance class. In Section 2, we define a weighted L^2-distance on space of fuzzy numbers, and then embed the space into the Hilbert space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$ by applying the weighted L^2-distance. In Section 3, we introduce weighted approximations of fuzzy numbers by means of best approximations from closed convex subsets of the Hilbert space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$. Some preliminaries are presented. In Section 4-6, by applying the reduction principle [8, p.80] we compute straightforwardly these approximations of fuzzy numbers, and then propose several important theorems.

2 Embedding fuzzy numbers into the Hilbert space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$

By an *inner product space* we mean that it is a (real) vector space V equipped with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ obeying the following axioms:

1. $\langle u, u \rangle \geq 0$ for all $u \in V$, and $\langle u, u \rangle = 0$ iff (if and only if) $u = 0$,
2. $\langle u, v \rangle = \langle v, u \rangle$, for all $u, v \in V$,
3. $\langle au + bv, w \rangle = a\langle u, w \rangle + b\langle v, w \rangle$, for all $u, v, w \in V$ and all $a, b \in \mathbb{R}$.

An inner product is a metric space if the distance is defined by

$$\mathrm{d}(u, v) := \langle u - v, u - v \rangle^{\frac{1}{2}}.$$

A completely inner product space is often called a *Hilbert space*. It is well-known that the set of all L^2-integrable functions is a Hilbert space, denoted by $L^2_\lambda[0,1]$, on which the inner product is defined as

$$\langle f, g \rangle_\lambda := \int_0^1 f(t)g(t)\lambda(t)dt,$$

where $\lambda = \lambda(t)$ is a nonnegative function on $[0,1]$ with $\int_0^1 \lambda(t)dt > 0$.

Another important Hilbert space is the product space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$, which will be discussed in this paper. Its inner product is defined by

$$\langle (f_1, f_2), (g_1, g_2) \rangle_\lambda := \int_0^1 \left[f_1(t)g_1(t) + f_2(t)g_2(t) \right] \lambda(t)dt$$

ISBN: 978-989-95079-6-8

for all (f_1, f_2) and $(g_1, g_2) \in L^2_\lambda[0,1] \times L^2_\lambda[0,1]$. We hence obtain

$$\begin{aligned}&\mathrm{d}^2_\lambda((f_1,f_2),(g_1,g_2))\\ =&\langle (f_1-g_1, f_2-g_2), (f_1-g_1, f_2-g_2)\rangle_\lambda\\ =&\int_0^1 \left(|f_1(t)-g_1(t)|^2 + |f_2(t)-g_2(t)|^2\right)\lambda(t)dt.\end{aligned}$$

Recall that, a fuzzy number $\tilde{A}$ can be represented by an ordered pair of left continuous functions $[A_L(\alpha), A_U(\alpha)]$ (called the α-cuts of $\tilde{A}$), $0 \le \alpha \le 1$, which satisfy the following conditions: (1) A_L is increasing on [0,1], (2) A_U is decreasing on [0,1], (3) $A_L(1) \le A_U(1)$. Let $\tilde{\mathbb{F}}$ denote the set of all fuzzy numbers. The weighted L^2*-distance* (*Euclidean distance*) on $\tilde{\mathbb{F}}$ is defined as

$$\begin{aligned}\mathrm{d}_\lambda(\tilde{A},\tilde{B}) :=\big[&\int_0^1 |A_L(\alpha)-B_L(\alpha)|^2\lambda(\alpha)d\alpha\\ &+\int_0^1 |A_U(\alpha)-B_U(\alpha)|^2\lambda(\alpha)d\alpha\big]^{\frac{1}{2}}.\end{aligned} \tag{1}$$

For more generality, we refer to [10] in which Grzegorzewski proposed two families of general distances on $\tilde{\mathbb{F}}$. Let $\tilde{A}$ and $\tilde{B}$ be two fuzzy numbers. The fuzzy addition and fuzzy subtraction operations on $\tilde{\mathbb{F}}$ are defined as follows:

$$\begin{aligned}\tilde{A}+\tilde{B} &:= [A_L(\alpha)+B_L(\alpha), A_U(\alpha)+B_U(\alpha)],\\ \tilde{A}-\tilde{B} &:= [A_L(\alpha)-B_U(\alpha), A_U(\alpha)-B_L(\alpha)].\end{aligned}$$

The above conditions (1)-(3) (the definition of fuzzy numbers) imply that A_L and $A_U \in L^2_\lambda[0,1]$, hence we define

$$i : \tilde{A} \mapsto (A_L, A_U) \in L^2_\lambda[0,1] \times L^2_\lambda[0,1].$$

In the following, we always use the interval notation $[A_L, A_U]$ instead of (A_L, A_U), although it may make little sense. Notice that, the fuzzy addition operation coincides with the vector addition on $L^2_\lambda[0,1]\times L^2_\lambda[0,1]$ and its inverse operation (vector subtraction) is not the fuzzy subtraction "-". Let the symbol "$\ominus$" denote the inverse operation, that is

$$\tilde{A}\ominus\tilde{B} := [A_L(\alpha)-B_L(\alpha), A_U(\alpha)-B_U(\alpha)],$$

which is often called the Hukuhara difference, see [17]. In fact, $\tilde{A}\ominus\tilde{B}$ may be not in $\tilde{\mathbb{F}}$. From Eq.(1), we find that

$$\mathrm{d}^2_\lambda(\tilde{A},\tilde{B}) = \langle \tilde{A}\ominus\tilde{B}, \tilde{A}\ominus\tilde{B}\rangle_\lambda.$$

This shows that we can embed space of fuzzy numbers into the Hilbert space $L^2_\lambda[0,1]\times L^2_\lambda[0,1]$. We hence define an inner product on $\tilde{\mathbb{F}}$ inheriting from $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$, that is

$$\langle \tilde{A},\tilde{B}\rangle_\lambda := \int_0^1 \left[A_L(\alpha)B_L(\alpha)+A_U(\alpha)B_U(\alpha)\right]\lambda(\alpha)d\alpha. \tag{2}$$

3 Approximations of fuzzy numbers

Let Ω be a subset of a Hilbert space V, then we call that:

1. Ω is a *subspace* iff $u+v \in \Omega$ and $ru \in \Omega$ for all $u, v \in \Omega$ and all $r \in \mathbb{R}$,
2. Ω is *convex* iff $ru+(1-r)v \in \Omega$ for all $u, v \in \Omega$ and all $r \in [0,1]$,
3. Ω is *chebyshev* iff for each $u \in V$ there exists a unique element $P_\Omega(u) \in \Omega$ such that

$$\mathrm{d}(u, P_\Omega(u)) \le \mathrm{d}(u,x), \quad \forall x \in \Omega,$$

and then $P_\Omega(u)$ is called the *best approximation* of u from Ω.

It is well-known that every closed convex subset (closed subspace, finite dimensional subspace) is chebyshev, see [8, p.23-24]. For any closed convex subset Ω, there is a sufficient and necessary condition of the best approximation $P_\Omega(u)$, as follows

$$\langle u - P_\Omega(u), x - P_\Omega(u)\rangle \le 0, \quad \forall x \in \Omega.$$

Furthermore, we also have

$$\mathrm{d}(P_\Omega(u), P_\Omega(v)) \le \mathrm{d}(u,v), \quad \forall u, v \in V, \tag{3}$$

refer to [21, Appendix C]. This implies P_Ω is continuous. While Ω is a closed subspace, then the above condition becomes

$$\langle u - P_\Omega(u), x\rangle = 0, \quad \forall x \in \Omega,$$

or equivalently

$$\langle u, x\rangle = \langle P_\Omega(u), x\rangle, \quad \forall x \in \Omega. \tag{4}$$

In this paper, all elements in $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$ of the form

$$[r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_3)\alpha]$$

are called *extended trapezoidal* fuzzy numbers. Let $\mathbb{T}$ denote the subset of all extended trapezoidal fuzzy numbers. It is easy to see that, an element $\tilde{A} = [r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_3)\alpha]$ is *trapezoidal* iff $\tilde{A} \in \tilde{\mathbb{F}} \cap \mathbb{T}$, that is

$$r_1 \le r_2 \le r_3 \le r_4. \tag{5}$$

Also, a trapezoidal fuzzy number $\tilde{A}$ is *triangular* (resp. *symmetric trapezoidal*, *symmetric triangular*, *interval*) iff $r_2 = r_3$ (resp. $r_2 - r_1 = r_4 - r_3$, $r_2 = r_3$ and $r_2 - r_1 = r_4 - r_3$, $r_1 = r_2$ and $r_3 = r_4$). Let $\tilde{\mathbb{T}}$, $\tilde{\mathbb{T}}_s$, $\tilde{\Delta}$, $\tilde{\Delta}_s$, and $\tilde{\mathbb{I}}$ denote the sets of all trapezoidal, symmetric trapezoidal, triangular, symmetric triangular, and interval fuzzy numbers, respectively. It is easy to verify that

1. $\mathbb{T}$ is a closed subspace of $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$, and
2. $\tilde{\mathbb{F}}$, $\tilde{\mathbb{T}}$, $\tilde{\mathbb{T}}_s$, $\tilde{\Delta}$, $\tilde{\Delta}_s$, and $\tilde{\mathbb{I}}$ are all closed convex subsets.

Hence, all of them are chebyshev. The best approximations of u from $\mathbb{T}$, $\tilde{\mathbb{T}}$, $\tilde{\mathbb{T}}_s$, $\tilde{\Delta}$, $\tilde{\Delta}_s$, and $\tilde{\mathbb{I}}$ are called the *extended trapezoidal*, *trapezoidal*, *symmetric trapezoidal*, *triangular*, *symmetric triangular*, and *interval approximations* of u, respectively. Eq.(3) implies that these approximations are continuous.

Theorem 1 (The reduction principle [8, p.80])**.** *Let K be a closed convex subset of an inner product space V and M be any chebyshev subspace of V that contains K. Then, we have that*

$$P_K(u) = P_K(P_M(u)) \quad \textit{and}$$
$$\mathrm{d}(u, P_K(u))^2 = \mathrm{d}(u, P_M(u))^2 + \mathrm{d}(P_M(u), P_K(u))^2.$$

ISBN: 978-989-95079-6-8

Now, let's define four extended trapezoidal fuzzy numbers:

$$\tilde{E}_1 := [1-\alpha, 0], \qquad \tilde{E}_2 := [\alpha, 0],$$
$$\tilde{E}_3 := [0, \alpha], \qquad \tilde{E}_4 := [0, 1-\alpha].$$

Then, each element in $\mathbb{T}$ is a linear combination of $\tilde{E}_i$, $1 \leq i \leq 4$, for instance

$$[r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_3)\alpha] = \sum_{i=1}^{4} r_i \tilde{E}_i.$$

This implies $\mathbb{T} = \text{Span}\,\{\tilde{E}_1, \tilde{E}_2, \tilde{E}_3, \tilde{E}_4\}$. We also define two other subspaces of $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$ as follows

$$\Delta := \text{Span}\,\{\tilde{E}_1, \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4\},$$
$$\mathbb{I} := \text{Span}\,\{\tilde{E}_1 + \tilde{E}_2, \tilde{E}_3 + \tilde{E}_4\} = \text{Span}\,\{[1,0], [0,1]\}.$$

It is easy to see that $\tilde{\mathbb{I}} \subset \mathbb{I}$, $\tilde{\Delta}_s \subset \tilde{\Delta} \subset \Delta$, and $\tilde{\mathbb{T}}_s \subset \tilde{\mathbb{T}} \subset \mathbb{T}$. By applying the reduction principle, we obtain that

$$P_{\tilde{\mathbb{I}}}(u) = P_{\tilde{\mathbb{I}}}(P_{\mathbb{I}}(u)), \tag{6}$$
$$P_{\tilde{\Delta}}(u) = P_{\tilde{\Delta}}(P_{\Delta}(u)), \quad P_{\tilde{\Delta}_s}(u) = P_{\tilde{\Delta}_s}(P_{\Delta}(u)), \tag{7}$$
$$P_{\tilde{\mathbb{T}}}(u) = P_{\tilde{\mathbb{T}}}(P_{\mathbb{T}}(u)), \quad P_{\tilde{\mathbb{T}}_s}(u) = P_{\tilde{\mathbb{T}}_s}(P_{\mathbb{T}}(u)). \tag{8}$$

4 The interval approximations

In 2002, Grzegorzewski first proposed interval approximations of fuzzy numbers [11]. Let's extend his results to the case of weighted L^2-distance. We now start with computing the best approximation $P_{\mathbb{I}}(\tilde{A})$ of any fuzzy number $\tilde{A} = [A_L(\alpha), A_U(\alpha)]$ from the subspace $\mathbb{I}$. Unless otherwise stated, we fix the following real numbers:

$$\lambda_0 := \int_0^1 \lambda(\alpha) d\alpha > 0$$

and

$$L_0 := \int_0^1 A_L(\alpha)\lambda(\alpha)d\alpha, \quad U_0 := \int_0^1 A_U(\alpha)\lambda(\alpha)d\alpha.$$

From Eq.(2), we find that $\langle [1,0], [0,1] \rangle_\lambda = 0$ and

$$\langle [1,0], [1,0] \rangle_\lambda = \langle [0,1], [0,1] \rangle_\lambda = \lambda_0.$$

Since $\mathbb{I} = \text{Span}\,\{[1,0], [0,1]\}$, we may assume that

$$P_{\mathbb{I}}(\tilde{A}) = r[1,0] + s[0,1].$$

By applying Eq.(4), we can solve

$$\begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} \langle [1,0],[1,0] \rangle_\lambda & \langle [0,1],[1,0] \rangle_\lambda \\ \langle [1,0],[0,1] \rangle_\lambda & \langle [0,1],[0,1] \rangle_\lambda \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1,0] \rangle_\lambda \\ \langle \tilde{A}, [0,1] \rangle_\lambda \end{pmatrix}$$
$$= \lambda_0^{-1} \begin{pmatrix} L_0 \\ U_0 \end{pmatrix}.$$

Hence, we obtain

$$P_{\mathbb{I}}(\tilde{A}) = [\lambda_0^{-1} L_0, \lambda_0^{-1} U_0].$$

The fact $A_L(\alpha) \leq A_U(\alpha)$ implies $L_0 \leq U_0$, hence $P_{\mathbb{I}}(\tilde{A}) \in \tilde{\mathbb{I}}$. By applying Eq.(6), we obtain the following theorem.

Theorem 2. *Let $\tilde{A}$ be a fuzzy number. Then, its interval approximation is $P_{\tilde{\mathbb{I}}}(\tilde{A}) = [\lambda_0^{-1} L_0, \lambda_0^{-1} U_0]$.*

While $\lambda(\alpha) = 1$, we get that $\lambda_0 = 1$. Then, the above equation coincides with the Grzegorzewski's formula [11, Equations (15) and (16)]. Also, the interval $[L_0, U_0]$ is called the *expected interval* of $\tilde{A}$, which is introduced by Dubois and Prade [9] and Heilpern [16], independently.

5 The triangular approximations

Let $\lambda = \lambda(\alpha)$ be a nonnegative function on [0,1] with $\int_0^1 \lambda(\alpha)\, d\alpha > 0$. In what follows, we fix

$$a := \int_0^1 (1-\alpha)^2 \lambda(\alpha) d\alpha > 0,$$
$$b := \int_0^1 \alpha(1-\alpha)\lambda(\alpha) d\alpha > 0,$$
$$c := \int_0^1 \alpha^2 \lambda(\alpha) d\alpha > 0,$$

and

$$L_1 := \int_0^1 A_L(\alpha)\alpha\lambda(\alpha)d\alpha, \quad U_1 := \int_0^1 A_U(\alpha)\alpha\lambda(\alpha)d\alpha.$$

By applying Schwarz inequality, we get

$$ac - b^2 > 0.$$

In a similar manner, we start with computing $P_\Delta(\tilde{A})$. Recall that $\Delta := \text{Span}\,\{\tilde{E}_1, \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4\}$, so we compute

$$\begin{pmatrix} \langle \tilde{E}_1, \tilde{E}_1 \rangle_\lambda & \langle \tilde{E}_2 + \tilde{E}_3, \tilde{E}_1 \rangle_\lambda & \langle \tilde{E}_4, \tilde{E}_1 \rangle_\lambda \\ \langle \tilde{E}_1, \tilde{E}_2 + \tilde{E}_3 \rangle_\lambda & \langle \tilde{E}_2 + \tilde{E}_3, \tilde{E}_2 + \tilde{E}_3 \rangle_\lambda & \langle \tilde{E}_4, \tilde{E}_2 + \tilde{E}_3 \rangle_\lambda \\ \langle \tilde{E}_1, \tilde{E}_4 \rangle_\lambda & \langle \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4 \rangle_\lambda & \langle \tilde{E}_4, \tilde{E}_4 \rangle_\lambda \end{pmatrix}$$
$$= \begin{pmatrix} a & b & 0 \\ b & 2c & b \\ 0 & b & a \end{pmatrix}.$$

Let $P_\Delta(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4$. By applying Eq.(4), we can solve

$$\begin{pmatrix} r_1 \\ r_2 \\ r_4 \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ b & 2c & b \\ 0 & b & a \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, \tilde{E}_1 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_2 + \tilde{E}_3 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_4 \rangle_\lambda \end{pmatrix}$$
$$= \frac{1}{\delta} \begin{pmatrix} 2ac - b^2 & -ab & b^2 \\ -ab & a^2 & -ab \\ b^2 & -ab & 2ac - b^2 \end{pmatrix} \begin{pmatrix} L_0 - L_1 \\ L_1 + U_1 \\ U_0 - U_1 \end{pmatrix}. \tag{9}$$

where $\delta = 2a(ac - b^2) > 0$. If we assume $\int_0^1 \lambda(\alpha)d\alpha = \frac{1}{2}$ (that is $a + 2b + c = \frac{1}{2}$) additionally, then the above $P_\Delta(\tilde{A})$ is equal to Zeng and Li's weighted triangular approximation [23]. Notice that $P_\Delta(\tilde{A})$ may be not in $\tilde{\mathbb{F}}$, refer to [21]. That shows that $P_{\tilde{\Delta}}(\tilde{A}) \neq P_\Delta(\tilde{A})$.

Lemma 3. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_\Delta(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4,$$

where r_1, r_2, and r_4 are computed by Eq.(9). Then,

1. $r_1 \leq r_4$,
2. *if $r_2 \leq r_1$, then*

$$-(a+b)L_0 + (a+3b+2c)U_0 - 2(a+2b+c)U_1 \geq 0,$$

3. *if $r_2 \geq r_4$, then*

$$-(a+3b+2c)L_0 + 2(a+2b+c)L_1 + (a+b)U_0 \geq 0.$$

ISBN: 978-989-95079-6-8

Proof. Omitted.

From Eq.(7), we find $P_{\tilde{\Delta}}(\tilde{A}) = P_{\tilde{\Delta}}(P_{\Delta}(\tilde{A}))$. Eq.(5) implies that an element

$$r\tilde{E}_1 + s(\tilde{E}_2 + \tilde{E}_3) + t\tilde{E}_4 \in \Delta$$

is triangular iff $r \le s \le t$. By Lemma 3.1, we obtain that $P_{\Delta}(\tilde{A}) \notin \tilde{\Delta}$ implies either $r_2 < r_1$ or $r_2 > r_4$, hence the best approximation (triangular approximation)

$$P_{\tilde{\Delta}}(\tilde{A}) := r_1'\tilde{E}_1 + r_2'(\tilde{E}_2 + \tilde{E}_3) + r_4'\tilde{E}_4$$

will satisfy $r_2' = r_1'$ or $r_2' = r_4'$, respectively. For instance, suppose that the fuzzy number $\tilde{A}$ has the approximation

$$P_{\Delta}(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4 \quad \text{with } r_2 < r_1.$$

Then, $P_{\tilde{\Delta}}(\tilde{A})$ will belong to Span $\{\tilde{E}_1 + \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4\}$, since $r_2' = r_1'$. So, we consider the best approximation of $\tilde{A}$ from Span $\{\tilde{E}_1 + \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4\}$. We hence let

$$P_{\tilde{\Delta}}(\tilde{A}) = r_1'(\tilde{E}_1+\tilde{E}_2+\tilde{E}_3)+r_4'\tilde{E}_4 = r_1'[1,\alpha]+r_4'[0,1-\alpha].$$

By applying Eq.(4), we can solve

$$\begin{aligned}\begin{pmatrix} r_1' \\ r_4' \end{pmatrix} &= \begin{pmatrix} a+2b+2c & b \\ b & a \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1,\alpha] \rangle_\lambda \\ \langle \tilde{A}, [0,1-\alpha] \rangle_\lambda \end{pmatrix} \\ &= \frac{1}{\delta} \begin{pmatrix} a & -b \\ -b & a+2b+2c \end{pmatrix} \begin{pmatrix} L_0 + U_1 \\ U_0 - U_1 \end{pmatrix},\end{aligned} \tag{10}$$

where $\delta = (a+b)^2 + 2(ac - b^2)$, and we have substituted by

$$\langle [1,\alpha], [1,\alpha] \rangle_\lambda = \int_0^1 (1+\alpha^2)\lambda(\alpha)d\alpha = a + 2b + 2c.$$

Notice that, the extended trapezoidal fuzzy number

$$\tilde{E}_1 + \tilde{E}_2 + \tilde{E}_3 = [1,\alpha]$$

does not belong to $\tilde{\mathbb{F}}$. But, this does not effect results of our computation. Applying Lemma 3.2, the reader can easily verify $r_1' \le r_4'$ in Eq.(10), so that

$$r_1'[1-\alpha, 0] + r_4'[\alpha, 1] \in \tilde{\mathbb{F}}.$$

Hence, we obtain

$$P_{\tilde{\Delta}}(\tilde{A}) = r_1'[1,\alpha] + r_4'[0,1-\alpha],$$

where r_1' and r_4' are computed by Eq.(10).

On the other hand, if $\tilde{A}$ has the approximation

$$P_{\Delta}(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4 \quad \text{with } r_2 > r_4$$

then its triangular approximation leads to

$$P_{\tilde{\Delta}}(\tilde{A}) = r_1'[1-\alpha, 0] + r_4'[\alpha, 1],$$

where

$$\begin{aligned}\begin{pmatrix} r_1' \\ r_4' \end{pmatrix} &= \begin{pmatrix} a & b \\ b & a+2b+2c \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1-\alpha, 0] \rangle_\lambda \\ \langle \tilde{A}, [\alpha, 1] \rangle_\lambda \end{pmatrix} \\ &= \frac{1}{\delta} \begin{pmatrix} a+2b+2c & -b \\ -b & a \end{pmatrix} \begin{pmatrix} L_0 - L_1 \\ L_1 + U_0 \end{pmatrix}.\end{aligned} \tag{11}$$

where $\delta = (a+b)^2 + 2(ac-b^2)$. Again, applying Lemma 3.3 the reader can easily verify $r_1' \le r_4'$ in Eq.(11).

Theorem 4. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\Delta}(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4,$$

where r_1, r_2 and r_4 are computed by Eq.(9). Then, the triangular approximation $P_{\tilde{\Delta}}(\tilde{A})$ can be determined in the following cases:

1. *If $r_1 \le r_2 \le r_4$, then*
$$P_{\tilde{\Delta}}(\tilde{A}) = [r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_2)\alpha].$$
2. *If $r_2 < r_1$, then $P_{\tilde{\Delta}}(\tilde{A}) = [r_1', r_4' - (r_4' - r_1')\alpha]$, where r_1' and r_4' are computed by Eq.(10).*
3. *If $r_2 > r_4$, then $P_{\tilde{\Delta}}(\tilde{A}) = [r_1' + (r_4' - r_1')\alpha, r_4']$, where r_1' and r_4' are computed by Eq.(11).*

Next, we compute the symmetric triangular approximation $P_{\tilde{\Delta}_s}(\tilde{A})$ which was first proposed by Ma et al. [18]. Recall that, an extended trapezoidal element

$$r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4 \in \Delta$$

is symmetric triangular iff

$$r_2 - r_1 = r_4 - r_2 \ge 0.$$

Hence, by substituting $r_2 = \frac{1}{2}(r_1 + r_4)$ we get

$$r_1\tilde{E}_1+r_2(\tilde{E}_2+\tilde{E}_3)+r_4\tilde{E}_4 = r_1[1-\frac{1}{2}\alpha, \frac{1}{2}\alpha]+r_4[\frac{1}{2}\alpha, 1-\frac{1}{2}\alpha].$$

Let's consider the best approximation $P_{\Delta_s}(\tilde{A})$ of $\tilde{A}$ from the subspace Δ_s, where Δ_s is defined by

$$\Delta_s := \text{Span } \{[1-\frac{1}{2}\alpha, \frac{1}{2}\alpha], [\frac{1}{2}\alpha, 1-\frac{1}{2}\alpha]\}.$$

Suppose that

$$P_{\Delta_s}(\tilde{A}) = r_1[1-\frac{1}{2}\alpha, \frac{1}{2}\alpha] + r_4[\frac{1}{2}\alpha, 1-\frac{1}{2}\alpha].$$

By applying Eq.(4), we can solve

$$\begin{aligned}&\begin{pmatrix} r_1 \\ r_4 \end{pmatrix} \\ &= \begin{pmatrix} a+b+\frac{c}{2} & b+\frac{c}{2} \\ b+\frac{c}{2} & a+b+\frac{c}{2} \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1-\frac{1}{2}\alpha, \frac{1}{2}\alpha] \rangle_\lambda \\ \langle \tilde{A}, [\frac{1}{2}\alpha, 1-\frac{1}{2}\alpha] \rangle_\lambda \end{pmatrix} \\ &= \frac{1}{\delta} \begin{pmatrix} a+b+\frac{c}{2} & -(b+\frac{c}{2}) \\ -(b+\frac{c}{2}) & a+b+\frac{c}{2} \end{pmatrix} \begin{pmatrix} L_0 - \frac{1}{2}L_1 + \frac{1}{2}U_1 \\ \frac{1}{2}L_1 + U_0 - \frac{1}{2}U_1 \end{pmatrix},\end{aligned} \tag{12}$$

where $\delta = a^2 + 2ab + ac$. The reader can verify that $r_1 \le r_4$ for any $\tilde{A} \in \tilde{\mathbb{F}}$. This implies $P_{\Delta_s}(\tilde{A}) \in \tilde{\mathbb{F}}$, so that

$$P_{\tilde{\Delta}_s}(\tilde{A}) = P_{\Delta_s}(\tilde{A}).$$

Theorem 5. *Let $\tilde{A}$ be a fuzzy number. Then, its symmetric approximation is*

$$P_{\tilde{\Delta}_s}(\tilde{A}) = [r_1 + \frac{r_4 - r_1}{2}\alpha, r_4 - \frac{r_4 - r_1}{2}\alpha],$$

where r_1 and r_4 are computed by Eq.(12).

While $\lambda(\alpha) = 1$, we get that, $a = \frac{1}{3}$, $b = \frac{1}{6}$, and $c = \frac{1}{3}$. Substituting into the above equation, we obtain

$$P_{\tilde{\Delta}_s}(\tilde{A}) = [x_0 - \sigma(1-\alpha), x_0 + \sigma(1-\alpha)],$$

where $x_0 = \frac{L_0+U_0}{2}$ and $\sigma = \frac{3}{2}(-L_0 + L_1 + U_0 - U_1)$. This formula coincides with [18, Equations (8) and (9)].

ISBN: 978-989-95079-6-8

6 The trapezoidal approximations

The (extended) trapezoidal approximation was first proposed by Abbasbandy and Asady in 2004 [1]. Afterwards, Grzegorzewski and Mrówka proposed the (extended) trapezoidal approximation preserving the expected interval [12]. Since the expected interval of any fuzzy number $\tilde{A}$ is equal to $P_{\mathbb{I}}(\tilde{A})$ and $\mathbb{I} \supseteq \mathbb{T} \supseteq \tilde{\mathbb{T}}$, by the reduction principle we get that these two (extended) trapezoidal approximations are equal. Now, we start with computing the extended trapezoidal approximation $P_{\mathbb{T}}(\tilde{A})$.

Because that $\mathbb{T} = \text{Span}\,\{\tilde{E}_1, \tilde{E}_2, \tilde{E}_3, \tilde{E}_4\}$, we let

$$P_{\mathbb{T}}(\tilde{A}) = t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4.$$

In the same vein, by applying Eq.(4) we can solve

$$\begin{aligned}
\begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{pmatrix} &= \begin{pmatrix} a & b & 0 & 0 \\ b & c & 0 & 0 \\ 0 & 0 & c & b \\ 0 & 0 & b & a \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, \tilde{E}_1 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_2 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_3 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_4 \rangle_\lambda \end{pmatrix} \\
&= \frac{1}{ac - b^2} \begin{pmatrix} c & -b & 0 & 0 \\ -b & a & 0 & 0 \\ 0 & 0 & a & -b \\ 0 & 0 & -b & c \end{pmatrix} \begin{pmatrix} L_0 - L_1 \\ L_1 \\ U_1 \\ U_0 - U_1 \end{pmatrix} \qquad (13) \\
&= \frac{1}{ac - b^2} \begin{pmatrix} cL_0 - (b+c)L_1 \\ -bL_0 + (a+b)L_1 \\ -bU_0 + (a+b)U_1 \\ cU_0 - (b+c)U_1 \end{pmatrix}.
\end{aligned}$$

While $\lambda(\alpha) = 1$, the above equation coincides with Grzegorzewski's formula [12, Equations (29)-(32)]. Note that, the extended trapezoidal approximation $P_{\mathbb{T}}(\tilde{A})$ may be not in $\tilde{\mathbb{F}}$, refer to [4, 5, 20]. This is because that $P_{\mathbb{T}}(\tilde{A})$ may happen $t_2 > t_3$.

Lemma 6. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\mathbb{T}}(\tilde{A}) = t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4,$$

where t_i, $1 \le i \le 4$, are computed by Eq.(13). Then,

$$t_1 \le t_2 \quad \textit{and} \quad t_3 \le t_4.$$

Proof. Omitted.

From Eq.(8), we find

$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = P_{\tilde{\mathbb{T}}}(P_{\mathbb{T}}(\tilde{A})). \qquad (14)$$

Hence, if $P_{\mathbb{T}}(\tilde{A})$ is in $\tilde{\mathbb{F}}$ (by applying Lemma 6, it is equivalent to $t_2 \le t_3$), we obtain

$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = P_{\mathbb{T}}(\tilde{A}).$$

Otherwise, we have $t_2 > t_3$. Consequently, Eq.(14) implies that the trapezoidal approximation $P_{\tilde{\mathbb{T}}}(\tilde{A})$ will be restricted to triangular fuzzy numbers $\tilde{\Delta}$. This leads to

$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = P_{\tilde{\Delta}}(\tilde{A}).$$

By applying Theorem 4, we prove the following theorem which is a generalization of [21, Theorem 4.4].

Theorem 7. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\mathbb{T}}(\tilde{A}) = t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4,$$

where t_i, $1 \le i \le 4$, are computed by Eq.(13). If $t_2 \le t_3$, then the trapezoidal approximation of $\tilde{A}$ is

$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = [t_1 + (t_2 - t_1)\alpha, t_4 - (t_4 - t_3)\alpha].$$

Otherwise, by Eq.(9) compute

$$P_{\Delta}(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4.$$

Then, the trapezoidal approximation $P_{\tilde{\mathbb{T}}}(\tilde{A})$ can be determined in the following cases:

1. *If $r_1 \le r_2 \le r_4$, then*
$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = [r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_2)\alpha].$$
2. *If $r_2 < r_1$, then $P_{\tilde{\mathbb{T}}}(\tilde{A}) = [r'_1, r'_4 - (r'_4 - r'_1)\alpha]$, where r'_1 and r'_4 are computed by Eq.(10).*
3. *If $r_2 > r_4$, then $P_{\tilde{\mathbb{T}}}(\tilde{A}) = [r'_1 + (r'_4 - r'_1)\alpha, r'_4]$, where r'_1 and r'_4 are computed by Eq.(11).*

Proof. Omitted.

In 1998, Delgado et al. [7] proposed a symmetric trapezoidal approximation of $\tilde{A}$ under the Euclidean distance between the respective 1/2-levels. In the following, we turn to study the symmetric trapezoidal approximation of $\tilde{A}$ under a weighted L^2-distance. Recall that, an extended trapezoidal fuzzy number $t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4$ is symmetric trapezoidal iff

$$t_2 - t_1 = t_4 - t_3 \ge 0 \quad \text{and} \quad t_2 \le t_3.$$

Let $d = t_2 - t_1$. By substituting $t_2 = t_1 + d$ and $t_3 = t_4 - d$, we obtain

$$t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4 = t_1[1,0] + d[\alpha, -\alpha] + t_4[0,1].$$

Let $\mathbb{T}_s := \text{Span}\,\{[1,0], [\alpha, -\alpha], [0,1]\}$. Suppose that the best approximation of $\tilde{A}$ from $\mathbb{T}_s$ is

$$P_{\mathbb{T}_s}(\tilde{A}) := t_1[1,0] + d[\alpha, -\alpha] + t_4[0,1].$$

By applying Eq.(4), we can solve

$$\begin{aligned}
\begin{pmatrix} t_1 \\ d \\ t_4 \end{pmatrix} &= \begin{pmatrix} \lambda_0 & \lambda_1 & 0 \\ \lambda_1 & 2\lambda_2 & -\lambda_1 \\ 0 & -\lambda_1 & \lambda_0 \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1,0] \rangle_\lambda \\ \langle \tilde{A}, [\alpha, -\alpha] \rangle_\lambda \\ \langle \tilde{A}, [0,1] \rangle_\lambda \end{pmatrix} \\
&= \frac{1}{\delta} \begin{pmatrix} 2\lambda_0\lambda_2 - \lambda_1^2 & -\lambda_0\lambda_1 & -\lambda_1^2 \\ -\lambda_0\lambda_1 & \lambda_0^2 & \lambda_0\lambda_1 \\ -\lambda_1^2 & \lambda_0\lambda_1 & 2\lambda_0\lambda_2 - \lambda_1^2 \end{pmatrix} \begin{pmatrix} L_0 \\ L_1 - U_1 \\ U_0 \end{pmatrix} \qquad (15)
\end{aligned}$$

where $\lambda_i := \int_0^1 \alpha^i \lambda(\alpha) d\alpha$, $i = 0, 1, 2$, and

$$\delta := 2\lambda_0(\lambda_0\lambda_2 - \lambda_1^2) > 0.$$

Lemma 8. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\mathbb{T}_s}(\tilde{A}) = t_1[1,0] + d[\alpha, -\alpha] + t_4[0,1],$$

where t_1, d, and t_4 are computed by Eq.(15). Then, $d \ge 0$.

ISBN: 978-989-95079-6-8

Proof. Omitted.

If $P_{\mathbb{T}_s}(\tilde{A}) \in \tilde{\mathbb{F}}$ (by applying Lemma 8, it is equivalent to $t_1 + d \leq t_4 - d$), then it equals the symmetric trapezoidal approximation $P_{\tilde{\mathbb{T}}_s}$. Otherwise, we will have $d > \frac{1}{2}(t_4 - t_1)$. This shows that, in the case the symmetric trapezoidal approximation

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) = t_1'[1,0] + d'[\alpha, -\alpha] + t_4'[0,1]. \qquad (16)$$

will satisfy $d' = \frac{1}{2}(t_4' - t_1')$. Substituting into Eq.(16) by $d' = \frac{1}{2}(t_4' - t_1')$, we get

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) := t_1'[1 - \frac{1}{2}\alpha, \frac{1}{2}\alpha] + t_4'[\frac{1}{2}\alpha, 1 - \frac{1}{2}\alpha].$$

Since $\Delta_s = \text{Span}\,\{[1 - \frac{1}{2}\alpha, \frac{1}{2}\alpha], [\frac{1}{2}\alpha, 1 - \frac{1}{2}\alpha]\}$, we get

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) \in \Delta_s \cap \tilde{\mathbb{F}} = \tilde{\Delta}_s.$$

This implies $P_{\tilde{\mathbb{T}}_s}(\tilde{A})$ is equal to the symmetric triangular approximation $P_{\tilde{\Delta}_s}(\tilde{A})$. Consequently, by applying Theorem 5 we obtain the following theorem.

Theorem 9. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\mathbb{T}_s}(\tilde{A}) = t_1[1,0] + d[\alpha, -\alpha] + t_4[0,1],$$

where t_1, d, and t_4 are computed by Eq.(15). If $d \leq \frac{1}{2}(t_4 - t_1)$, then the symmetric trapezoidal approximation of $\tilde{A}$ is

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) = [t_1 + d\alpha, t_4 - d\alpha].$$

Otherwise, it is

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) = [r_1 + \frac{r_4 - r_1}{2}\alpha, r_4 - \frac{r_4 - r_1}{2}\alpha],$$

where r_1 and r_4 are computed by Eq.(12).

7 Conclusions

Recently, many scholars studied on computation of interval, triangular, and trapezoidal approximations approximations of fuzzy numbers by applying Langrange multiplier method or Karush-Kuhn-Tucker theorem. In the present paper, we propose a new method for computing these approximations under a weighted distance by applying function approximation theory on Hilbert space.We embed fuzzy numbers into the Hilbert space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$. Then, by introducing extended trapezoidal fuzzy numbers and applying the reduction principle (Theorem 1), it suffices to compute the best approximations of an extended trapezoidal fuzzy number. Hence, we can easily determine these approximations by choosing a suitable basis. In fact, the weighted distance can be generalized to more general form, as follows

$$\mathrm{d}_\lambda(\tilde{A}, \tilde{B}) = \int_0^1 |A_L - B_L|^2 d\mu_1 + \int_0^1 |A_U - B_U|^2 d\mu_2,$$

where μ_1 and μ_2 are arbitrary positive measures on [0,1].

Acknowledgment

The author is very grateful to Prof. P. Grzegorzewski and Prof. L. Stefanini for their invitation and suggestions, which have been very helpful in improving the presentation of the paper. The author also wishes to thank my advisor Prof. Chen-Lian Chuang for his assistance. This work was supported by National Science Council of Taiwan.

References

[1] S. Abbasbandy and B. Asady, The nearest trapezoidal fuzzy number to a fuzzy quantity, Applied Mathematics and Computation 156(2004) 381-386.

[2] S. Abbasbandy and M. Amirfakhrian, The nearest trapezoidal form of a generalized left right fuzzy number, International Journal of Approximate Reasoning 43(2006) 166-178.

[3] S. Abbasbandy and M. Amirfakhrian, The nearest approximation of a fuzzy quantity in parametric form, Applied Mathematics and Computation 172(2006) 624-632.

[4] T. Allahviranloo and M.A. Firozja, Note on "Trapezoidal approximation of fuzzy numbers", Fuzzy Sets and Systems 158(2007) 747-754.

[5] A. Ban, Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval, Fuzzy Sets and Systems, 159(2008) 1327-1344.

[6] S. Chanas, On the interval approximation of a fuzzy number, Fuzzy Sets and Systems 122(2001) 353-356.

[7] M. Delgado, M.A. Vila and W. Voxman, On a canonical representation of fuzzy number, Fuzzy Sets and Systems 93(1998) 125-135.

[8] F. Deutsch, Best approximation in inner product space, Springer, New York, 2001.

[9] D. Dubois and H. Prade, The mean value of a fuzzy number, Fuzzy Sets and Systems 24(1987) 279-300.

[10] P. Grzegorzewski, Metrics and orders in space of fuzzy numbers, Fuzzy Sets and Systems 97(1998) 83-94.

[11] P. Grzegorzewski, Nearest interval approximation of a fuzzy number, Fuzzy Sets and Systems 130(2002) 321-330.

[12] P. Grzegorzewski and E. Mrówka, Trapezoidal approximations of fuzzy numbers, Fuzzy Sets and Systems 153(2005) 115-135.

[13] P. Grzegorzewski and E. Mrówka, Trapezoidal approximations of fuzzy numbers-revisited, Fuzzy Sets and Systems 158(2007) 757-768.

[14] P. Grzegorzewski and E. Mrówka, Trapezoidal approximations of fuzzy numbers preserving the expected interval - algorithms and properties, Fuzzy Sets and Systems, 159(2008) 1354-1364.

[15] R. Hassine, F. Karray, A.M. Alimi and M. Selmi, Approximation properties of piece-wise parabolic functions fuzzy logic systems, Fuzzy Sets and Systems 157(2006) 501-515.

[16] S. Heilpern, The expected value of a fuzzy number, Fuzzy Sets and Systems 47(1992) 81-86.

[17] M. Hukuhara, Integration des applications measurables dont la valeur est un compact convexe, Funkcialaj Ekvacioj 10(1967) 205-223.

[18] M. Ma, A. Kandel and M. Friedman, A new approach for defuzzification, Fuzzy Sets and Systems 111(2000) 351-356.

[19] E. N. Nasibov and S. Peker, On the nearest parametric approximation of a fuzzy number, Fuzzy Sets and Systems, 159(2008) 1365-1375.

[20] C.T. Yeh, A note on trapezoidal approximations of fuzzy numbers, Fuzzy Sets and Systems 158(2007) 747-754.

[21] C.T. Yeh, On improving trapezoidal and triangular approximations of fuzzy numbers, International Journal of Approximate Reasoning, 48/1 (2008) 297-313.

[22] C.T. Yeh, Trapezoidal and triangular approximations preserving the expected interval, Fuzzy Sets and Systems, 159(2008) 1345-1353.

[23] W. Zeng and H. Li, Weighted triangular approximation of fuzzy numbers, International Journal of Approximate Reasoning, 46/1 (2007) 137-150.

ISBN: 978-989-95079-6-8

Fuzzy c–Means Herding

Thomas A. Runkler

Siemens Corporate Technology, Munich, Germany
Email: Thomas.Runkler@siemens.com

Abstract— *Herding is the process of bringing individuals (e.g. animals) together into a group. More specifically, we consider self–organized herding as the process of moving a set of individuals to a given number of locations (cluster centers) without any external control. We formally describe the relation between herding and clustering and show that any clustering model can be used to control herding processes. For the specific case of the fuzzy c–means model we derive the equations of the fuzzy c–means herding algorithm using a gradient descent approach with limited step size. Several experiments related to an autonomous mobile robot scenario show that fuzzy c–means herding yields smooth trajectories, well–balanced clusters, and fast convergence.*

Keywords— Fuzzy clustering, herding, robot swarms, swarm intelligence.

1 Introduction

This paper deals with the problem of *online herding* of n individuals specified by the state vectors $x_1, \ldots, x_n \in \mathbb{R}^p$, into c clusters. Application examples for herding are illustrated by the following four scenarios:

1. A swarm of autonomous micro–robots explores the Mars surface. Just before their individual energy supplies are exhausted, the robots move to a couple of individual locations from where they are collected and returned to the spaceship.
2. Museum visitors individually walk through an exhibition. At the beginning of the guided tours, each visitor receives instructions where to go to join one of the tour groups.
3. Mountain hikers have been caught by a severe thunderstorm. They receive instructions on their mobile phones where to go, so that they can be picked off from several evacuation points by helicopter.
4. Wireless sensor devices are attached to individual components of an industrial plant for condition monitoring purposes. At power on, each sensor automatically adapts its communication protocol so that several clusters of networked sensors emerge.

Herding mobile robots (as in our first scenario) is a very active field in the robotics community. A neuro–fuzzy control approach for herding (also called convoying) of multiple robots in presented in [1]. A similar approach to herding very heterogeneous types of robots is presented in [2]. Also herding in evacuation processes (as in our third scenario) is currently actively discussed in the scientific community, for example in [3]. Herding has also been applied to swarm based learning processes in order to balance exploration and exploitation behavior [4]. Some work has also been done concerning herding using outside agents [5], for example the so–called *dog and sheep* herding [6]. In this paper, however, we only consider *self–organizing* herding, where all individuals equally contribute to the herding process, without any external control.

In each of the four scenarios presented above, the individual state vectors $x_1, \ldots, x_n$ (for example the locations of the micro–robots) are successively updated, until a clustering is achieved, i.e. each of the individuals x_k is (fuzzily) assigned to cluster i with a membership $u_{ik} \in [0, 1]$. When the herding process is converged, each state vector is equal to the prototypical state $v_i \in \mathbb{R}^p$, $i = 1, \ldots, c$, of its corresponding cluster, and is crisply assigned to this cluster, $u_{ik} = 1$ and $u_{jk} = 0$ for all $j \neq i$. More formally, herding can be described by a sequence

$$X \rightarrow (U, V, X) \rightarrow (U, V, X) \rightarrow (U, V, X) \rightarrow \ldots \quad (1)$$

which converges to $X = V$ and $U \in M_{hcn}$,

$$M_{hcn} = \left\{ \; U \in \{0,1\}^{c \times n} \;\middle|\; \sum_{i=1}^{c} u_{ik} = 1, \quad k = 1, \ldots, n, \quad \sum_{k=1}^{n} u_{ik} > 0, \quad i = 1, \ldots, c \; \right\} \quad (2)$$

In practical scenarios we will not be able to run this algorithm for an infinite time until convergence, but will terminate after a previously specified number of t steps, or when an appropriate termination criterion holds, for example when the maximum change in the state vectors is lower than a previously specified threshold. Even though the limit of the final partition is crisp and somewhat trivial (because $X = V$), we pursue a fuzzy approach here to *smoothly* control the herding process.

Apparently, herding is closely related to clustering. A discussion of the relation between herding and clustering from a theoretical physics perspective can be found in [7]. Just as herding, clustering also assigns each object to a cluster, and each cluster can be described by a prototype as well, but clustering does not change the data vectors X. More formally, in contrast to (1) clustering is described by a sequence

$$X \rightarrow (U, V) \rightarrow (U, V) \rightarrow (U, V) \rightarrow \ldots \quad (3)$$

where the initial X remains constant. Despite this difference between clustering and herding, the herding process can be performed by optimizing *any* conventional clustering model. The main difference is that in clustering we consider (U, V)

ISBN: 978-989-95079-6-8

as the set of free variables for optimization, and in herding we consider (U, V, X). In this paper we illustrate this for the case of the *fuzzy c–means* [8] clustering model. Other suitable clustering models are, for example, hard c–means [9], possibilistic c–means [10], noise clustering [11], or high–dimensional fuzzy c–means [12].

This paper is structured as follows: Sections 2 and 3 briefly repeat the fuzzy c–means objective function and how this model is applied to clustering. Section 4 shows how the fuzzy c–means model can be used to control a herding process, leading to the fuzzy c–means herding algorithm. Section 5 presents experiments with the fuzzy c–means herding algorithm referring to the Mars robot scenario illustrated above. Section 6 finally summarizes the conclusions and gives an outline for future work in this field.

2 The Fuzzy c–Means Objective Function

The *fuzzy c–means* objective function [8] is a dissimilarity–based measure for the clustering of a data set $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^p$ into $c \in \{2, \ldots, n-1\}$ clusters specified by a partition matrix $U \in M_{fcn}$ where

$$M_{fcn} = \left\{ \; U \in [0,1]^{c \times n} \;\middle|\; \sum_{i=1}^{c} u_{ik} = 1, \quad k = 1, \ldots, n, \quad \sum_{k=1}^{n} u_{ik} > 0, \quad i = 1, \ldots, c \; \right\} \tag{4}$$

and a set of prototypes (cluster centers) $V = \{v_1, \ldots, v_c\} \subset \mathbb{R}^p$. The FCM objective function is the sum of the dissimilarities between data and prototypes, weighted by the (fuzzy) memberships. Using a fuzziness index $m > 1$, it is formally defined as

$$J(U, V, X, c, m) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m \, |x_k - v_i|^2. \tag{5}$$

For $m \to 1$ this objective function converges to the *hard c–means objective function* [9]. For many applications good results were reported for $m = 2$. In this paper we always use the Euclidean distance $|.| = \|.\|$. Lower values of J indicate a better clustering of (U, V, X). The theoretical minimum of J is zero. This minimum is achieved, if for each $k = 1, \ldots, n$ we have an $i \in \{1, \ldots, c\}$ with $x_k = v_i$, $u_{ik} = 1$ and $u_{jk} = 0$ for all $j \neq i$, i.e. $X = V$ and $U \in M_{hcn}$. This means that herding converges to $J = 0$.

3 Fuzzy c–Means Clustering

As pointed out above we assume X to be given and fixed in clustering and only find U and V. In fuzzy c–means clustering we find U and V by minimizing J. In clustering we are usually not able to reduce J to zero. We may use different optimization methods such as genetic algorithms [13, 14], particle swarm optimization [15, 16, 17], ant colony optimization [18], or wasp swarm optimization [19]. In this paper we apply direct minimization using the necessary conditions for local extrema or saddle points of J. An equation to compute the optimal U for a given V under the constraints in (4) is obtained by defining the Langrangian

$$L = J - \sum_{k=1}^{n} \lambda_k \left(1 - \sum_{i=1}^{c} u_{ik} \right) \tag{6}$$

and setting $\partial L / \partial u_{ik} = 0$, from which follows

$$u_{ik} = 1 \Big/ \sum_{j=1}^{c} \left(\frac{|x_k - v_i|}{|x_k - v_j|} \right)^{\frac{2}{m-1}}, \tag{7}$$

$i = 1, \ldots, c$, $k = 1, \ldots, n$. An equation to compute the optimal V for a given U is obtained by setting $\partial J / \partial v_i = 0$, from which follows

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}, \tag{8}$$

$i = 1, \ldots, c$. The equations for V depend on U and the equations for U depend on V, so optimization can be done in an alternating scheme, the so–called alternating optimization. The clustering algorithm terminates when an appropriate termination criterion holds, e.g. when successive changes in V are lower than a threshold ε. The fuzzy c–means clustering algorithm can then be formally defined as follows:

Algorithm 1: Fuzzy c–Means Clustering.

```
input X, c, m
randomly initialize V ∈ R^p
repeat
   update U(V, X, m) by (7)
   update V(U, X, m) by (8)
until max max ΔV < ε
output U, V
```

4 Fuzzy c–Means Herding

As pointed out in the introduction, herding does not only update U and V at each optimization step but also X. If we do the optimization of $J(U, V, X)$ in an alternating scheme, then we also have to determine an appropriate update equation for X. In the scenarios described above we do not arbitrarily change the values of X (for example the position of the robots) in each step but we can only *gradually* change X (for example because the robots can only move with a limited speed). Therefore, in contrast to the *direct* optimization of U and V in clustering, we only apply a *gradient descent* of X with a limited step size $s \in \mathbb{R}^+$ in herding. The step size s divided by the time for one step represents the maximum speed of the individuals, e.g. robots. The direction vectors for the gradient descent on X are obtained as

$$\partial J / \partial x_k = \Delta x_k = -\sum_{i=1}^{c} 2 u_{ik}^m \, (x_k - v_i), \tag{9}$$

$k = 1, \ldots, n$. So, the update equations for X are

$$x_k = x_k + \frac{\min\{s, \|\Delta x_k\|\}}{\|\Delta x_k\|} \cdot \Delta x_k, \tag{10}$$

ISBN: 978-989-95079-6-8

$k = 1, \ldots, n$. This makes sure that the individuals move in the direction of the gradient but that the step length never exceeds the threshold s. The herding algorithm terminates when all elements of X have approached the corresponding elements of V, i.e. when $\max \Delta X < s$. The fuzzy c–means herding algorithm can then be formally defined as follows:

Algorithm 2: Fuzzy c–Means Herding.

```
input X, c, m, s
randomly initialize V ∈ ℝ^p
repeat
   update U(V,X,m) by (7)
   update V(U,X,m) by (8)
   update X(U,V,m) by (10) and (9)
until max‖ΔX‖ < s
output U, V, X
```

5 Experiments

We illustrate the operation of the fuzzy c–means herding algorithm using an artificial data set that is inspired by the autonomous robot scenario presented at the beginning of section 1. We randomly generate a set X of $n = 100$ data vectors of the dimension $p = 2$ using a Gaussian distribution with mean zero and variance one. This data set is illustrated in Fig. 1. The idea is that these data vectors represent the current positions of the exploring robots in the two–dimensional plane just before the herding process begins. Then, we start the fuzzy c–means herding algorithm with a fuzziness index of $m = 2$, a step size $s = 0.05$ (representing the speed of the robots), and different numbers of clusters $c \in \{2, 3, 4, 5\}$. Each experiment is performed for the same initial vectors X and V. Additional experiments with other random initializations of X and V (not presented in detail here) produced very similar results. The trajectories of the vectors V and X are shown in Fig. 2. The solid curves in these figures show the trajectories of X. Each of these trajectories starts at one of the initial points of X (as shown in Fig. 1), smoothly moves through the two–dimensional plane, and terminates in one of the final cluster centers. The data vectors do not directly move towards one of the cluster centers, but are apparently initially fuzzily assigned to several clusters, have a tendency to move towards the mean of the data set and then gradually focus on one of the cluster centers. We counted the number of steps t until all of the data points eventually reached one of the cluster centers and obtained the following table:

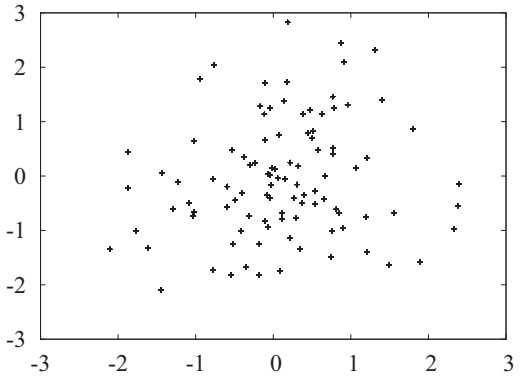


Figure 1: Initial data set (robot positions).

Table 1: Convergence steps for fuzzy c–means herding, $m = 2$.

c	2	3	4	5
t	34	29	25	23

A detailed study of the step size showed that lower values of s produce smoother trajectories and lead to higher numbers of convergence steps t. Correspondingly, higher values of s produce rather coarse trajectories and lead to lower numbers of convergence steps t. Despite the influence of s on the absolute values of t, the relation between the convergence steps for the different experiments are almost independent of s.

The ticks $\times$ in the figures show the trajectories of V. The cluster centers in V are not fixed but also move through the two–dimensional plane with an approximately decreasing step size as the herding process proceeds. The data points that are assigned to each cluster originally covered roughly the same area in the two–dimensional plane. Notice that a similar effect can be observed with the fuzzy c–means clustering algorithm.

The reader might wonder if it wasn't more efficient to arbitrarily pick fixed cluster centers, say c of the initial data vectors in X, assign each data point to the closest data center and let each data point directly move towards its pre–assigned data point. The trajectories obtained by this experiment are shown in Fig. 3. The trajectories of X (solid curves) form stars whose centers are the initially randomly chosen cluster centers. Unlike the clusters obtained by fuzzy c–means herding, the data points in each of these clusters do not represent similar areas, but some clusters obtain only very few, and others obtain many data points, depending on the cluster centers chosen. Apparently, the linear trajectories of some of the data points are considerably shorter than those obtained by fuzzy c–means herding. However, some of the trajectories are considerably longer than their herding equivalents. Also for this experiment we counted the number of steps t until all of the data points reached their closest cluster centers:

Table 2: Convergence steps for pre–assigned cluster centers (one specific run).

c	2	3	4	5
t	35	29	27	26

Notice that fuzzy c–means herding yields the same or even lower values of t for all numbers of clusters. However, this result depends on the particular choices of the cluster centers. Therefore, we repeated each of these experiments with 1000 different random choices of cluster centers and computed the numbers of steps until termination. The histograms of these step numbers are shown in Fig. 4 and the mean numbers of steps $\bar{t}$ are given in the following table:

Table 3: Convergence steps for pre–assigned cluster centers (average of 1000 runs).

c	2	3	4	5
$\bar{t}$	33.9	30.3	28.0	26.4

The average number of steps is almost equal or larger than the number of steps needed for fuzzy c–means herding with

ISBN: 978-989-95079-6-8

$m = 2$. In order to reduce the number of convergence steps in fuzzy c–means herding we reduce the value of the fuzziness parameter m, so the individual data points are less fuzzily assigned to the clusters in the beginning and more directly orientate towards their corresponding cluster centers. To avoid numerical problems that sometimes occur when m is very close to one, we pick $m = 1.1$. For this case, the trajectories of V and X are shown in Fig. 5. Just as in Fig. 2 and in contrast to Fig. 3 each cluster corresponds to approximately the same area in the two–dimensional plane. The trajectories are a little less smooth than in Fig. 2 and rather resemble the star–like trajectories in Fig. 3. This observation is corroborated by the fact that the number of convergence steps is much lower than in Fig. 2, as shown in the following table:

Table 4: Convergence steps for fuzzy c–means herding, $m = 1.1$.

c	2	3	4	5
t	28	22	20	18

A comparison of Table 4 with Table 3 reveals the fact that the number of convergence steps for fuzzy c–means herding is even lower than (on average) for pre–assigned cluster centers. We marked the numbers of convergence steps of fuzzy c–means herding for $m = 2$ and for $m = 1.1$ as vertical lines in Fig. 4. Apparently, for $m = 1.1$ the number of steps is very close to the minimum of the histograms, i.e. it compares with the number of steps when the clusters by chance have been approximately optimally assigned. In other words, fuzzy c–means herding implicitly finds an (almost) optimal cluster assignment, leads to a considerably low number of convergence steps, and yields smoothly herding trajectories at the same time.

6 Conclusions

In this paper we have shown that minimization of clustering models can be used to control herding processes. In particular, we have focussed on herding by minimization of the fuzzy c–means clustering model. Th experiments with a gradient descent of the data vectors with limited step size can be illustrated by a scenario with automous mobile robots that have to move to a certain number of locations (cluster centers) in a self–organizing manner. The experiments with fuzzy c–means herding yield very smooth trajectories. The individuals that are collected at each cluster center approximately represent the same area (or, more generally, hypervolume) of the initial state space. This means that the final clusters are nicely balanced. Compared with herding with pre–defined cluster centers, fuzzy c–means herding on average needs less steps until termination. For low values of the fuzziness index the trajectories become less smoother, but the number of steps is further reduced. For $m = 1.1$ the number of steps is even close to the global minimum. Thus, fuzzy c–means herding is an attractive herding algorithm that yields smooth trajectories, well–balanced clusters, and fast convergence.

References

[1] K. C. Ng and M. M. Trivedi. Multirobot convoying using neuro-fuzzy control. In *International Conference on Pattern Recognition*, pages 417–421, Vienna, Austria, August 1996.

[2] L. E. Parker, B. Kannan, X. Fu, and Y. Tang. Heterogeneous mobile sensor net deployment using robot herding and line-of-sight formations. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2488–2493, Las Vegas, October 2003.

[3] A. Kirchner and A. Schadschneider. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. *Physica A*, 312(1–2):260–276, 2002.

[4] M. R. Akbarzadeh, H. Rezaei, and M. B. Naghibi. A fuzzy adaptive algorithm for expertness based cooperative learning, application to herding problem. In *International Conference of the North American Fuzzy Information Processing Society*, pages 317–322, Chicago, July 2003.

[5] J. M. Lien, O. B. Bayazit, R. T. Sowell, S. Rodriguez, and N. M. Amato. Shepherding behaviors. In *IEEE International Conference on Robotics and Automation*, pages 4159–4164, New Orleans, April 2004.

[6] P. Kachroo, S. A. Shedied, J. S. Bay, and H. Vanlandingham. Dynamic programming solution for a class of pursuit evasion problems: the herding problem. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 31(1):35–41, 2001.

[7] D. Costantini, S. Donadio, U. Garibaldi, and P. Viarengo. Herding and clustering: Ewens vs. Simon-Yule models. *Physica A*, 355(1):224–231, 2005.

[8] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[9] G. B. Ball and D. J. Hall. Isodata, an iterative method of multivariate analysis and pattern classification. In *IFIPS Congress*, 1965.

[10] R. Krishnapuram and J. M. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, May 1993.

[11] R. N. Davé. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12:657–664, 1991.

[12] T. A. Runkler. A single objective fuzzy clustering model for very high–dimensional data. In *GMA/GI Workshop Fuzzy Systems and Computational Intelligence, Dortmund, Germany*, pages 234–243, December 2008.

[13] J. C. Bezdek and R. J. Hathaway. Optimization of fuzzy clustering criteria using genetic algorithms. In *IEEE Conference on Evolutionary Computation, Orlando*, volume 2, pages 589–594, June 1994.

[14] F. Klawonn and A. Keller. Fuzzy clustering with evolutionary algorithms. *International Journal of Intelligent Systems*, 13(10–11):975–991, 1998.

[15] C. Y. Chen and F. Ye. Particle swarm optimization and its application to clustering analysis. In *International Conference on Networking, Sensing and Control*, pages 789–794, Taipeh, Taiwan, March 2004.

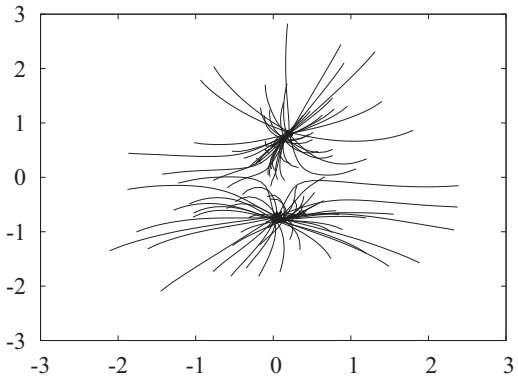
[16] X. Cui, T. E. Potok, and P. Palathingal. Document clustering using particle swarm optimization. In *IEEE Swarm Intelligence Symposium, Pasadena*, pages 185–191, June 2005.

[17] T. A. Runkler and C. Katz. Fuzzy clustering by particle swarm optimization. In *IEEE International Conference on Fuzzy Systems*, pages 3065–3072, Vancouver, July 2006.

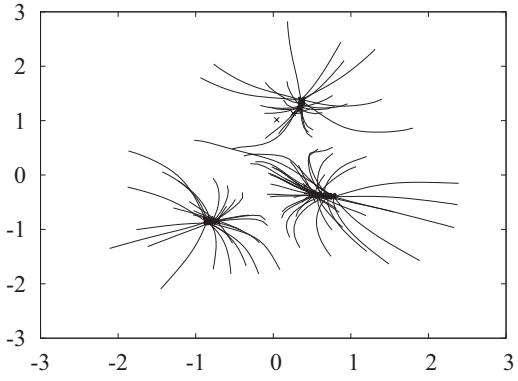
[18] T. A. Runkler. Ant colony optimization of clustering models. *International Journal of Intelligent Systems*, 20(12):1233–1261, December 2005.

[19] T. A. Runkler. Wasp swarm optimization of the c-means clustering model. *International Journal of Intelligent Systems*, 23(3):269–285, February 2008.

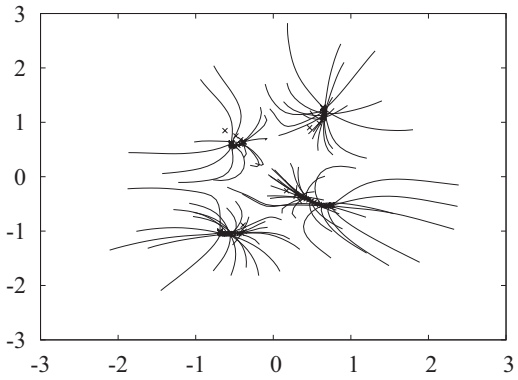
ISBN: 978-989-95079-6-8



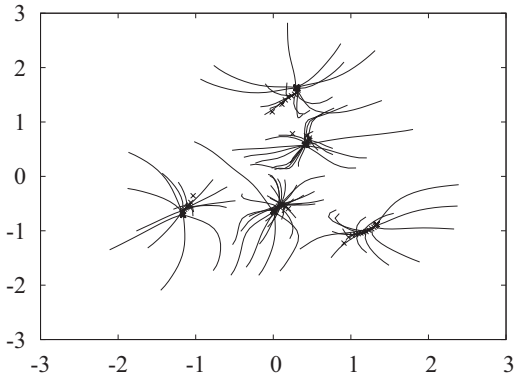
$c = 2$



$c = 3$

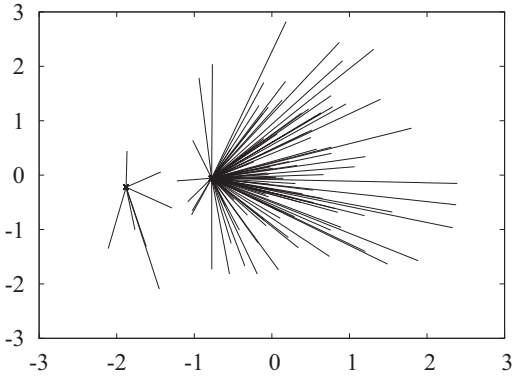


$c = 4$

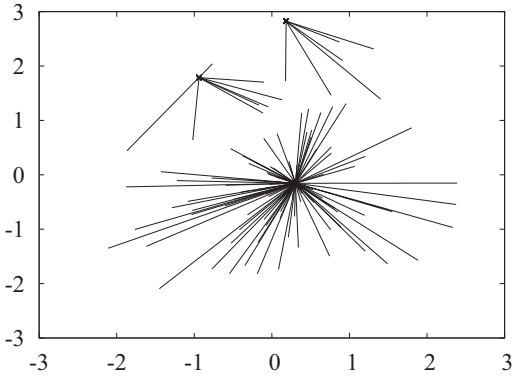


$c = 5$

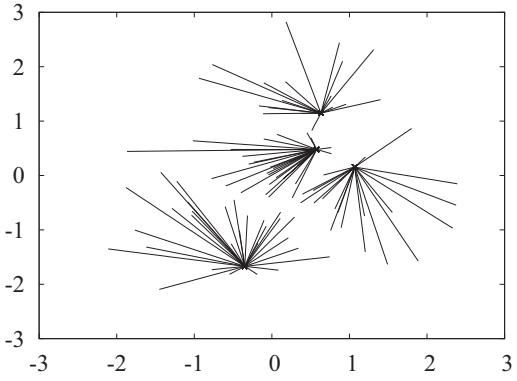
Figure 2: Herding trajectories for $m = 2$.



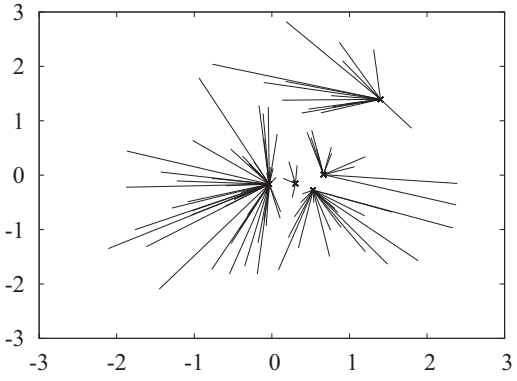
$c = 2$



$c = 3$



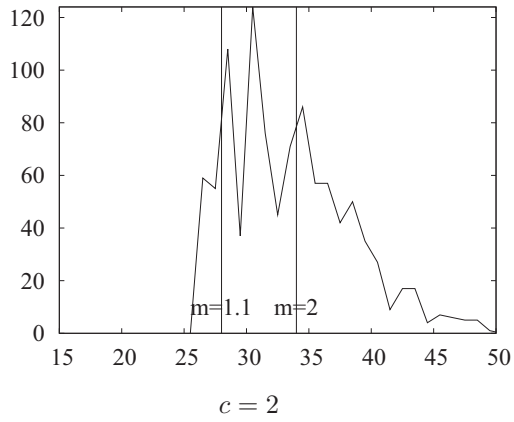
$c = 4$



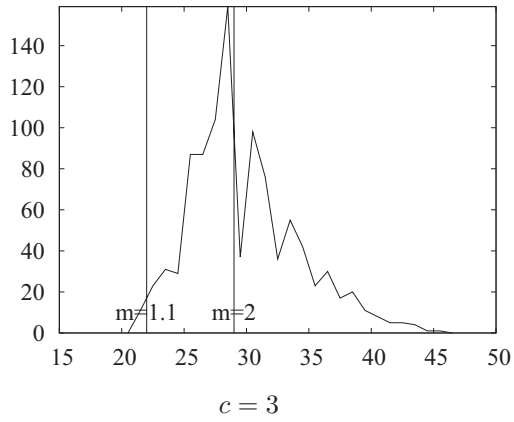
$c = 5$

Figure 3: Trajectories with pre-assigned arbitrary cluster centers.

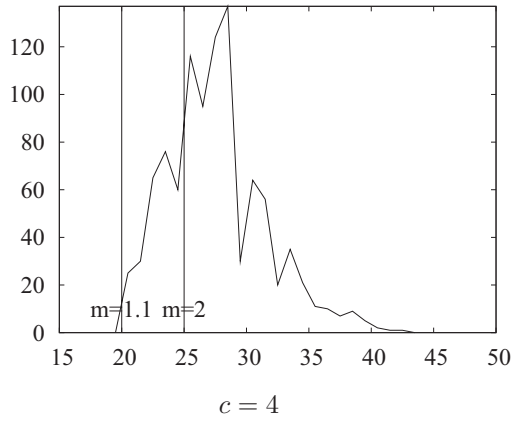
ISBN: 978-989-95079-6-8



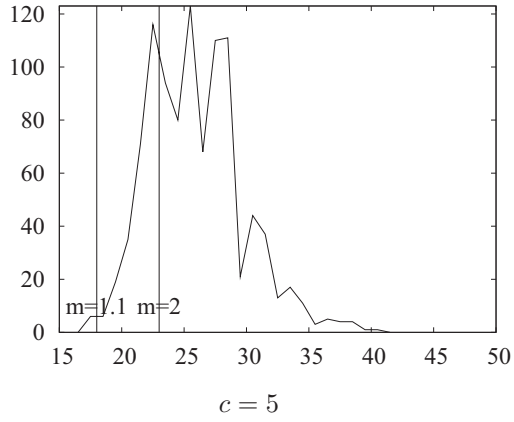
$c = 2$



$c = 3$

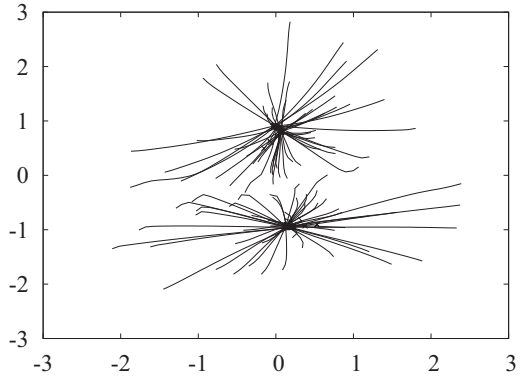


$c = 4$

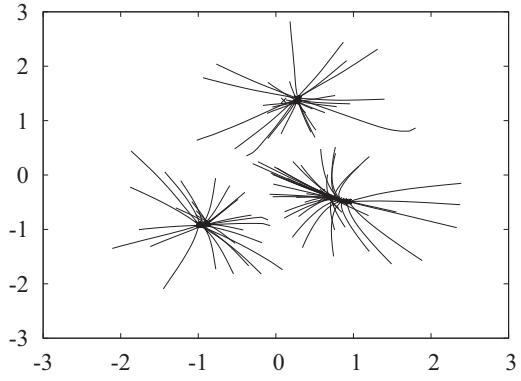


$c = 5$

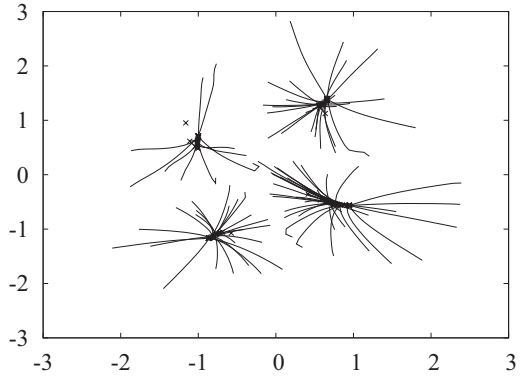
Figure 4: Histograms of the number of steps for preassigned cluster centers.



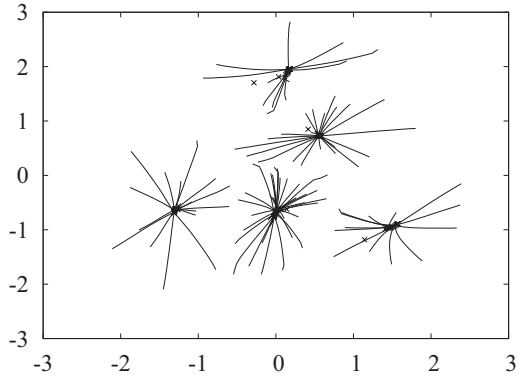
$c = 2$



$c = 3$



$c = 4$



$c = 5$

Figure 5: Herding trajectories for $m = 1.1$.

ISBN: 978-989-95079-6-8

Interval AHP for a group of decision makers

Tomoe Entani[1] 1.Department of Economics and Social Sciences, Kochi University
2-5-1 Akebono Kochi, Japan
Email: entani@cc.kochi-u.ac.jp

Abstract— *This paper considers Interval Analytic Hierarchy Process (Interval AHP) in group decision making for encouraging communication. Interval AHP is suitable method to handle subjective judgments since the induced results are intervals which can include uncertainty of given information. The decision makers' opinions can be aggregated at some stages of decision making process to reach consensus. One of the approaches is to aggregate the given judgments considering outliers and from them the group preferences are obtained. By the other approach, first the individual preferences are obtained from the respective decision maker's judgments and then they are aggregated. In the sense of reducing communication barriers, obtaining individual priority weights of alternatives beforehand helps decision makers realize their own and the others' opinions. The judgments and preferences can be aggregated based on the possibility view or by introducing importance weights of decision makings.*

Keywords— group decision making, importance weight, interval AHP, intuitive judgments, possibility, uncertainty

1 Introduction

In most organizations decision making is made by two or more people, regardless of whether the organization is public or private and the decision problem is local or global. Although all members do not need to be located in the same physical location, they are aware of one another and act as a member of the group which makes a decision. The apparent purpose of the group is to reach a decision, that is, to choose one alternative which seems to be acceptable and agreeable for all members. However, it is sometimes difficult to reach a consensus among group members [1]. Especially when there are various members who are not face-to-face, there exist some barriers to understand one another. These communication barriers make decision making tough. The technological advancements, such as computers and networks, accelerate such situations. In addition, the decision makers have some difficulties to represent and recognize their own opinions which fit their intuitive judgments and experiences, because the problems are often very complicated. It may happen that some members may exaggerate their preferences to influence the group decision. In this sense, it is also important to support the interpersonal information exchange, as well as to find the optimal alternative, in the group decision making. As a preparation for the consensus, it becomes necessary to remove communication barriers by representing individual opinions simply and clearly [1, 2]. This paper concerns on these design phase more than the choice phase in group decision support system by focusing on the aggregation of decision makers' opinions.

The group decision support system is discussed from the scope of AHP (Analytic Hierarchy Process). AHP is a useful method in multi-criteria decision making problems [3]. It is structured hierarchically as criteria and alternatives. The priority or weight for each element of the hierarchy is obtained by eigenvector method [3]. Then, the priority weight for each alternative is calculated with them. The advantages of AHP are the following two points. It helps decision maker structure complex problems hierarchically. The decision makers can give their subjective judgments directly by comparing a pair of attributes.

The group decision making with AHP is discussed in [4, 5, 6, 7]. On this assumption, more than two comparison matrices are given. The decision makers' individual opinions are aggregated in several stages of decision making process by several approaches. It can be done at the beginning stage by searching consensus on the given judgments. It is also possible to take the average of individually induced preferences with importance weights of the group members at the last stage. Between these stages, the decomposed intermediate judgments may be aggregated by voting and so on [2]. This paper concerns on the aggregations at the beginning and last stages based on two concepts such as possibility and nominal average. Furthermore, in order to reflect uncertainty of the given comparisons, Interval AHP [8, 9], where the interval weights, instead of crisp weights in the conventional AHP, are obtained is applied.

This paper is based on possibility concept in dealing with the given data and obtained results and consists as follows. As a preliminary, the definition and properties of the interval probability which are used for normalization of intervals are explained in Section 2. At first, Interval AHP as a tool to represent each decision maker's preference is shown briefly in Section 3. Then, in Section 4, the approaches to aggregate various opinions from several viewpoints are proposed. Finally, the proposed models are tested with a numerical example in case of a group of four decision makers in Section 5.

2 Interval probability as preliminary

The interval probabilities are defined by a set of intervals as follows. This definition is originally proposed in [10] and also is used in [11, 12]. The conventional crisp probabilities are extended into interval ones.

Definition 1 (Interval probability) *The set of intervals denoted as* $\{W_1, ..., W_n\}$ *where* $W_i = [\underline{w}_i, \overline{w}_i]$ *are called interval probabilities if and only if*

$$\begin{array}{ll} 1) & 0 \le \underline{w}_i \le \overline{w}_i \ \forall i \\ 2) & \sum_{i \ne j} \overline{w}_i + \underline{w}_j \ge 1 \forall j \\ 3) & \sum_{i \ne j} \underline{w}_i + \overline{w}_j \le 1 \ \forall j. \end{array} \quad (1)$$

From Definition 1, following two inequalities hold

$$\textstyle\sum_i \underline{w}_i \le 1 \ \text{ and } \ \sum_i \overline{w}_i \ge 1.$$

ISBN: 978-989-95079-6-8

Definition 1 is regarded as a normality condition of intervals corresponding to the conventional one $\sum_i w_i = 1$. It is noted that in interval probabilities there are many combinations of crisp values whose sum is one.

The combination of a pair of interval probability sets is denoted as follows.

Property 1 Combination rule *Assuming a pair of interval probabilities on n elements as $\{W_1^A, ..., W_n^A\}$ and $\{W_1^B, ..., W_n^B\}$ which satisfy Definition 1, their combination is denoted as $\{W_1^{AB}, ..., W_n^{AB}\}$. Each of elements is an interval $W_i^{AB} = [\underline{w}_i^{AB}, \overline{w}_i^{AB}]$ denoted as follows.*

$$\begin{array}{l} \underline{w}_i^{AB} = \min\{\underline{w}_i^A, \underline{w}_i^B\} \\ \overline{w}_i^{AB} = \max\{\overline{w}_i^A, \overline{w}_i^B\} \end{array} \tag{2}$$

The set of combined intervals $\{W_1^{AB}, ..., W_n^{AB}\}$ also satisfies Definition 1 so that it is interval probability.

Property 2 Weighted sum of interval probabilities *Assuming m sets of interval probabilities on n elements as $\{W_1^{A_k}, ..., W_n^{A_k}\}$ $k = 1, ..., m$, and the weights for the sets as $\{p_k \forall k\}$ where $p_k \geq 0$ and $\sum_k p_k = 1$, the weighted sum of interval probabilities is $\{W_1, ..., W_n\}$. Each of elements is an interval $W_i = [\underline{w}_i, \overline{w}_i]$ denoted as follows.*

$$\begin{array}{l} \underline{w}_i = \sum_k p_k \underline{w}_i^{A_k} \\ \overline{w}_i = \sum_k p_k \overline{w}_i^{A_k} \end{array} \tag{3}$$

As far as the sum of weights is 1, the weighted sum of interval probabilities becomes also interval probability. It is verified as follows. They apparently satisfy the condition 1) in Definition 1. The condition 2) holds as follows.

$$\begin{array}{l} \sum_{i \neq j} \sum_k p_k \overline{w}_i^{A_k} + \sum_k p_k \underline{w}_j^{A_k} \\ = \sum_k p_k (\sum_{i \neq j} \overline{w}_i^{A_k} + \underline{w}_j^{A_k}) \text{ [Definition 1-2)]} \\ \geq \sum_k p_k = 1 \end{array} \tag{4}$$

The other condition 3) is satisfied by the similar way.

$$\begin{array}{l} \sum_{i \neq j} \sum_k p_k \underline{w}_i^{A_k} + \sum_k p_k \overline{w}_j^{A_k} \\ = \sum_k p_k (\sum_{i \neq j} \underline{w}_i^{A_k} + \overline{w}_j^{A_k}) \\ \leq \sum_k p_k = 1 \end{array} \tag{5}$$

Then, the weighted sum of interval probabilities satisfy the conditions in Definition 1.

3 Interval AHP

AHP is an approach to multi-criteria decision making problems. The problem is decomposed into hierarchy by criteria and alternatives. The choice or preferences of alternatives are induced as a final decision from the decision maker's judgments given as pariwise comparison matrix. The decision maker compares all pairs of alternatives and gives the pairwise comparison matrix for n alternatives as follows [3].

$$A = \begin{pmatrix} 1 & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{n1} & \cdots & 1 \end{pmatrix} \tag{6}$$

where a_{ij} shows the importance ratio of alternative i comparing to alternative j. The comparison matrix satisfies the following relations so that the number of given comparisons is $n(n-1)/2$.

$$\begin{array}{ll} a_{ii} = 1 & \text{identical} \\ a_{ij} = 1/a_{ji} & \text{reciprocal} \end{array} \tag{7}$$

The decision maker can give his/her judgment intuitively without caring about the relative relations of comparisons. Therefore, the given comparisons are not always consistent each other. The consistent comparisons satisfy the following transitivity relations.

$$a_{ij} = a_{ik} a_{kj} \ \forall (i, j) \tag{8}$$

In the following, inconsistency means that (8) is not satisfied. The proposed models in this paper deals with such inconsistency from the possibility view [13].

In the conventional AHP, crisp priority weights are obtained from the given comparison matrix. They are extended to intervals in Interval AHP [8, 9]. The given comparisons are inconsistent each other, that is, they do not always satisfy (8). In order to reflect such inconsistency, the priority weight of alternative is denoted as the following interval.

$$W_i = [\underline{w}_i, \overline{w}_i] \ \forall i \tag{9}$$

For their normalization, they are represented as interval probabilities so that they satisfy (1) in Definition 1.

The pairwise comparison is an intuitive ratio of two alternatives so that they are approximated by the following interval.

$$\frac{W_i}{W_j} = \left[\frac{\underline{w}_i}{\overline{w}_j}, \frac{\overline{w}_i}{\underline{w}_j} \right] \tag{10}$$

where $0 < \underline{w}_i \forall i$ and the upper and lower bounds of the approximated comparison are defined as the maximum range with respect to the two intervals.

In the approximation model the probabilities are determined so as to include the given pairwise comparisons. Thus, from the possibility view, the obtained interval probabilities satisfy the following inclusion relations.

$$a_{ij} \in \frac{W_i}{W_j} \ \forall (i, j) \tag{11}$$

which leads to the following inequalities.

$$\frac{\underline{w}_i}{\overline{w}_j} \leq a_{ij} \leq \frac{\overline{w}_i}{\underline{w}_j} \Leftrightarrow \left\{ \begin{array}{l} \underline{w}_i \leq a_{ij} \overline{w}_j \ \forall (i, j) \\ \overline{w}_i \geq a_{ij} \underline{w}_j \ \forall (i, j) \end{array} \right. \tag{12}$$

The approximations by the obtaining interval priority weights include the given inconsistent comparisons.

For any inconsistent comparisons, assuming $[\underline{w}_i, \overline{w}_i] = [0, 1] \ \forall i$, the above inclusion relation (12) is apparently satisfied. A decision maker does not need to revise their intuitive judgments so as to be consistent. When a decision maker gives completely inconsistent judgments, the obtained priority weights of all alternatives are equally $[0, 1]$. It represents complete ignorance and fits our natural sense. Inconsistency among the given comparisons is reflected in the uncertainty of interval probabilities.

ISBN: 978-989-95079-6-8

The constraint conditions for determining the interval probabilities are (1) and (12). In order to obtain the least uncertain probabilities, the uncertainty of interval probabilities should be minimized. The uncertainty of interval probabilities can be measured by several indices, such as widths of intervals and entropy [14]. For simplicity, the sum of widths of intervals is used in this paper. The problem to determine the interval priority weights is formulated as follows.

$$\begin{array}{ll} I = & \min \quad \sum_i (\overline{w}_i - \underline{w}_i) \\ \text{s.t.} & \sum_{i \neq j} \overline{w}_i + \underline{w}_j \geq 1 \; \forall j \\ & \sum_{i \neq j} \underline{w}_i + \overline{w}_j \leq 1 \; \forall j \\ & \underline{w}_i \leq a_{ij} \overline{w}_j \; \forall (i,j) \\ & \overline{w}_i \geq a_{ij} \underline{w}_j \; \forall (i,j) \\ & \underline{w}_i \geq \epsilon \; \forall i \end{array} \tag{13}$$

The greater optimal objective function value is, the more uncertain the given interval priority weight becomes.

4 Group of Decision Makers

Interval AHP is introduced to the group decision making by aggregating individual opinions. Each decision maker gives pairwise comparisons for alternatives based on his/her subjective judgments. The comparison matrix given by the decision maker k, where $k = 1, ..., m$, is denoted as follows.

$$A_k = \begin{pmatrix} 1 & \cdots & a_{1nk} \\ \vdots & a_{ijk} & \vdots \\ a_{n1k} & \cdots & 1 \end{pmatrix} \quad \forall k \tag{14}$$

In the following sections, the given judgments and the obtained priority weights are aggregated at the beginning and last stages of the decision making process, respectively. The procedures of these approaches are illustrated in Figure 1.

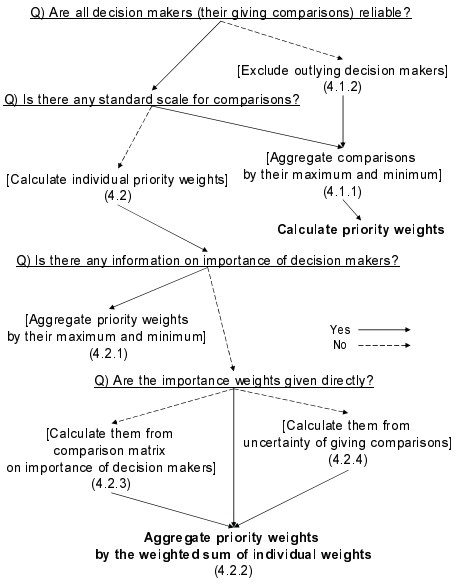


Figure 1: Steps for group decision making

4.1 Aggregation of given judgments

When the judgments which are the comparisons on alternatives in this paper are given, they can be aggregated at the beginning stage of group decision making process. It is usually said that the smaller group is more efficient since all the given judgments are reliable and tend not to be too much diverse. As for the well experienced group, for instance, the standard of judgments might be previously required or satisfied and then it is useful to aggregate the given judgments at the beginning stage. From the view of possibility concept, this is a natural and simple aggregation method. The decision makers can see their differences on their giving comparisons.

4.1.1 Directly aggregation of comparisons

The approach where the geometric mean of comparisons is taken has been proposed in [4, 5, 15, 16]. Since the aggregated comparison matrix satisfies (7), the eigenvector method can be also used to obtain the priority weights. In this section, the comparisons given by m decision makers are aggregated from the possibility view by taking their minimum and maximum.

$$A_{ij} = [\underline{a}_{ij}, \overline{a}_{ij}] = [\min_k a_{ijk}, \max_k a_{ijk}] \;\; \forall (i,j) \tag{15}$$

Since the aggregated comparisons are intervals, the inclusion constraints (12) can be written as follows.

$$A_{ij} \in \frac{W_i}{W_j} \;\Leftrightarrow\; \frac{\underline{w}_i}{\overline{w}_j} \leq \underline{a}_{ij}, \overline{a}_{ij} \leq \frac{\overline{w}_i}{\underline{w}_j} \tag{16}$$

where the aggregated comparisons are interaval and included in the approximated ones by the interval priority weights. The inclusion relation (16) with interval comparisons is an extension of (12) with crisp ones.

By solving the revised problem (13) from constraint (12) into (16), the bounds of the interval priority weights are obtained. The approximated comparison with them surely includes individually given comparisons.

4.1.2 Finding and excluding outliers

The decision making with a big group is also required. Instead of including opinions of all the decision makers in Section 4.1.1, the way to find outliers is proposed in this section. The decision makers whose opinions are apparently different from the others is called outliers. Such outliers are removed one by one till the number of removed decision makers reaches the pre-fixed number. It is noted that the number of appropriate decision makers should be determined beforehand. Then, the following function is used to measure his/her outlierness.

$$J_o = \sum_{(i,j)} \frac{\max_k a_{ijk}}{\max_{k \neq o} a_{ijk}} + \frac{\min_{k \neq o} a_{ijk}}{\min_k a_{ijk}} \geq n(n-1) \tag{17}$$

The evaluation function J_o represents the amount of extension by adding the decision maker o into the group. The increase of their upper bounds is measured as a ratio since the comparisons are ratio measure. As for the lower bounds, the decrease ratio is measured by the same way. The decision maker $\{o^* | J_o^* = \max_o J_o\}$, who gives the maximum of J is removed. If there are more than two decision makers who give the maximum, all of them are excluded. When the number of excluded decision makers reach the pre-fixed number, the comparisons given by the others are aggregated by (15) into the interval comparison matrix. The following process is the same as in Section 4.1.1.

ISBN: 978-989-95079-6-8

4.2 Aggregation of obtained individual preferences

The pariwise comparisons are relative values and each decision maker gives comparisons based on his/her own standard. In order to deal with these variations of comparisons, the decision makers' opinions are aggregated after obtaining individual preferences from respective given comparison matrix. First, the priority weights are obtained from the individually given pairwise comparison matrix. The interval priority weights based on the comparison matrix given by kth decision maker are denoted as $[\underline{w}_{ik}, \overline{w}_{ik}]$. Each decision maker can realize his/her priority weights on the alternatives, as well as others' ones. Then, in order to reach a consensus of the group, the obtained individual priority weights are aggregated.

4.2.1 Directly aggregation of priority weights

When there is not enough information on the importance of the decision makers, they are aggregated simply by taking their maximum and minimum.

$$W_i^1 = [\underline{w}_i, \overline{w}_i] = [\min_k w_{ik}, \max_k w_{ik}] \; \forall i \tag{18}$$

The obtained set of intervals are also interval probabilities by Property 1. The aggregated interval can be said to give an agreeable result, since the priority weights by all members are included in it. This is based on possibility concept and the assumption that all the decision makers give reasonable information. When the individual preferences are very different one another such as in a big group, the widths of the aggregated priority weights become large so that they are uncertain. The diversity of opinions is reflected in uncertainty of the aggregated weights. Based on the individual preferences, it is also possible to exclude outliers similarly in Section 4.1.2.

4.2.2 Aggregation with directly given importance weights

When the importance of decision makers and/or their confidence degrees of the given judgments are different, it is better to reflect such information. Therefore, the weighted sum of priority weights is proposed in the following sections.

When the importance of each decision maker is given by a supervisor, it is introduced as his/her weight directly. Denoting the given importance weight as $p_k \geq 0 \forall k$, where $\sum_k p_k = 1$, the aggregated priority weights are obtained as follows.

$$W_i^2 = [\sum_k p_k \underline{w}_{ik}, \sum_k p_k \overline{w}_{ik}] \; \forall i \tag{19}$$

As far as the weights are normalized so as to make the sum one, the set of the aggregated priorities is also interval probability by Property 2. The situation that there is not enough information is also represented by giving all the decision makers the same importance weights $1/m$. In the next section, it is represented differently from the view of interval probabilities.

4.2.3 Aggregation with importance weights by AHP

It is also possible to use conventional AHP and Interval AHP in order to induce the importance weights of decision makers. It is easier for a supervisor to give pariwise comparisons on importance of decision makers than to give their importance weights directly. The similar idea has been proposed in [6, 16], where a level of decision makers is added to the existing criteria levels. The importance weights of decision makers are determined at the level. Then, they are used as weights in the way of summing the local weights up to the overall priority weights. Using Interval AHP in obtaining weights, the overall weights are obtained as intervals in [17].

The supervisor gives comparisons on the importance of decision makers.

$$B = \begin{pmatrix} 1 & \cdots & b_{1m} \\ \vdots & b_{kl} & \vdots \\ b_{m1} & \cdots & 1 \end{pmatrix} \tag{20}$$

where each comparison represents how important is the decision maker comparing to the other.

Applying the conventional AHP in [3], the importance weights are obtained so as to make the sum one. Then, the following process is the same as in Section 4.2.2.

The Interval AHP formulated by (13) can be used to determine importance weights. The given comparisons on alternatives a_{ij} and the obtained priority weights $W_i = [\underline{w}_i, \overline{w}_i]$ in (13) are replaced into those on decision makers b_{kl} and their interval importance weights $P_k = [\underline{p}_k, \overline{p}_k]$, respectively. By solving the revised problem (13), the interval importance weights of decision makers are obtained. The problems to obtain the upper and lower bounds of the aggregated priority weight of alternative i are formulated as follows, respectively.

$$\begin{aligned} \underline{w}_i = \quad & \min \textstyle\sum_k p_{ik} \underline{w}_{ik} \\ & \text{s.t.} \; \textstyle\sum_k p_{ik} = 1 \\ & \underline{p}_k \leq p_{ik} \leq \overline{p}_k \; \forall k \\ \overline{w}_i = \quad & \max \textstyle\sum_k p_{ik} \overline{w}_{ik} \\ & \text{s.t.} \; \textstyle\sum_k p_{ik} = 1 \\ & \underline{p}_k \leq p_{ik} \leq \overline{p}_k \; \forall k \end{aligned} \tag{21}$$

In both problems, the variables are the importance weights p_{ik} included in the obtained interval importance weights. For each alternative i and its upper and lower bounds, the optimal solutions might be different. Denoting their optimal solutions for the alternative i as p_{ik*} and p_{ik}^*, respectively, the aggregated priority weight is obtained as the following intervals.

$$W_i^3 = [\sum_k p_{ik*} \underline{w}_{ik}, \sum_k p_{ik}^* \overline{w}_{ik}] \; \forall i \tag{22}$$

They are also interval probabilities by Properties 1 and 2.

From the view of interval probability, complete ignorance is represented as giving all decision makers the same interval weights $P_k = [0, 1]$. With them, the priority weights (22) by solving (21) is the same as the directly aggregated ones (18). Apparently, the aggregated priority weights with importance weights are included in the directly aggregated priority weights, that is, $W_i^2, W_i^3 \in W_i^1$. The results with importance weights are less uncertain. If any information on importance of decision makers is available, it is reasonable to introduce them as their importance weights.

4.2.4 Aggregation with importance weights based on uncertainty of information

As shown in Section 3, the sum of widths represents the uncertainty of the interval weights. The uncertainty of the given information may depend on the confidence or expert levels of decision makers on the problem. It is possible to assume such uncertainty degree of judgments as the decision maker's

ISBN: 978-989-95079-6-8

importance among the group. The importance weights of the decision makers are determined by reflecting uncertainty of their information. The decision maker who gives less uncertain comparisons will be evaluated as more important. The importance weight of decision maker l is as follows.

$$p_l = \frac{\sum_{k \neq l} I_k}{(m-1)\sum_k I_k} \tag{23}$$

where I_k is the optimal objective value of (13) with the given information by the kth decision maker and represents his/her uncertainty degree. It is noted that their sum is one, that is, $\sum_l p_l = (\sum_l \sum_{k \neq l} I_k)/((m-1)\sum_k I_k) = 1$. The importance weight of the decision maker who gives interval weights with great widths becomes small.

The aggregated priority weights are denoted as follows and also interval probabilities by Property 2.

$$W_i^4 = [\sum_k p_l \underline{w}_{ik}, \sum_k p_l \overline{w}_{ik}] \; \forall i \tag{24}$$

5 Numerical example

Assuming four decision makers, $k = 1, 2, 3, 4$, each of them gives the pairwise comparison matrix on four alternatives, $i = 1, 2, 3, 4$, shown in Table 1. The interval priority weights W_k and their uncertainty indices I_k are obtained by (13) and its optimal objective function value, respectively.

Table 1: Comparison matrices by four decision makers

A_1				W_1	
1	2	3	4	0.500	J_1=12
	1	2	3	0.250	I^1=0.083
		1	2	[0.125,0.167]	p_1=0.297
			1	[0.083,0.125]	
A_2				W_2	
1	2	4	6	0.545	J_2=13.2
	1	3	4	0.273	I^1=0.109
		1	4	[0.091,0.145]	p_2=0.285
			1	[0.036,0.091]	
A_3				W_3	
1	3	3	4	0.571	J_3=13.5
	1	2	4	0.190	I^1=0.190
		1	1	[0.095,0.190]	p_3=0.250
			1	[0.048,0.143]	
A_4				W_4	
1	1	2	2	0.375	J_4=16.5
	1	3	1	[0.219,0.375]	I^1=0.375
		1	3	[0.125,0.188]	p_4=0.168
			1	[0.063,0.219]	

5.1 Aggregation of comparisons

By (15), the comparisons given by four decision makers are aggregated and shown in Table 2. The interval priority weights are obtained by solving (13) with constraints (16). Each decision maker can accept the results more or less, since they are based on his/her giving comparisons by aggregating them into interval comparisons.

Table 2: Aggregated comparison matrix

A (aggregated)				W
1	[1,3]	[2,4]	[2,6]	[0.333,0.500]
	1	[2,3]	[1,4]	[0.167,0.333]
		1	[1,4]	[0.111,0.167]
			1	[0.042,0.167]

In this example, there are only four decision makers so that there is no need to exclude any of them. However, in order to make sure how the outliers are found, the appropriate number of decision makers is assumed as three so that one of four decision makers will be excluded. The evaluation function values J_k by (17) are shown in the right column of Table 1. They represent the changes of aggregated comparisons' bounds by excluding each decision maker and are measured as the sum of increase and decrease ratios. The fact, $J_1 = 12 = n(n-1)$, tells that A_1 does not effect the bounds of aggregated comparisons. Since J_4 is the maximum of the four, the others, A_1, A_2 and A_3, are aggregated and the comparison matrix is transformed from Table 2 into Table 3. The obtained interval priority weights are also shown in Table 3. Since outlier A_4 is excluded, the interval priority weights in Table 3 is less uncertain than those in Table 2.

Table 3: Aggregated comparison matrix without outlier A_4

A (aggregated)				W
1	[2,3]	[3,4]	[4,6]	0.522
	1	[2,3]	[3,4]	[0.174,0.261]
		1	[1,4]	[0.087,0.174]
			1	[0.043,0.130]

5.2 Aggregation of priority weights

Instead of aggregating comparisons at the beginning stage, it is possible to aggregate the calculated individual preferences at the last stage. The obtained individual priority weights shown in Table 1 help the decision makers realize their own preferences, as well as understand one another.

The aggregated priority weights with and without the importance weights of decision makers are shown in Table 5. Its left column represents the case with no information on the importance of decision makers so that they are aggregated directly by (18). The results equal to those obtained by (21) with $P_k = [0, 1] \forall k$. All decision makers' individual priority weights are included in the aggregated ones, $W_k \in W^1$ $k = 1, 2, 3, 4$. In this sense, the results are easily acceptable for all decision makers. On the other hand, the right two columns represent the cases where the importance weights of decision makers are determined based on Interval AHP and the uncertainty of their giving judgments, respectively. The comparison matrix for importance of decision makers, B, is given by a supervisor and the obtained interval importance weights, P, are shown in Table 4. With the importance weights, the priority weight of each alternative W^3 is obtained as an interval by (21). The uncertainty of the given judgments, I_k, which is the

ISBN: 978-989-95079-6-8

optimal function value of (13) and the respective importance weight, p_k, calculated by (23) are shown in Table 1. The A_1 and A_4 who give the least and most uncertain information are considered as the most and least important, respectively. With the importance weights, the priority weights of alternatives W^4 are obtained as intervals by (24). The uncertainty of W^3 and W^4 is less than that of W^1 which represents complete ignorance on decision makers' importance, since their widths of intervals are smaller. The individually obtained priory weights are reflected depending on their importance among the group.

Table 4: Comparison matrix on importance of decision makers

B				P
1	2	2	4	0.462
	1	3	4	[0.231,0.317]
		1	3	[0.106,0.231]
			1	[0.077,0.115]

Table 5: Aggregated interval priority weights with and without importance weights of decision makers

W^1	Interval AHP W^3	uncertainty W^4
[0.375,0.571]	[0.508,0.517]	0.510
[0.190,0.375]	[0.239,0.265]	[0.236,0.263]
[0.091,0.190]	[0.110,0.169]	[0.108,0.170]
[0.036,0.219]	[0.062,0.131]	[0.058,0.136]

6 Conclusion

The group decision support system based on Interval AHP has been discussed. Interval AHP is particularly suitable for dealing with uncertainty of human intuitive judgments. By Interval AHP the priority weights of alternatives and importance weights of decision makers are obtained as interval from the given pariwise comparison matrix. The obtained interval priority weights reflect all the possibilities in the given information. The approaches of aggregating group members' opinions at two different stages have been proposed. At the beginning stage, the given judgments are aggregated from the possibility view and the outliers are excluded if there are some. From the aggregated comparison matrix whose elements are intervals, the priority weights of alternatives are obtained by Interval AHP as intervals. The other approach is to aggregate individual preferences at the last stage. At first, from each given comparison matrix the interval priority weights are obtained by Interval AHP. They represent the decision makers' individual preferences. Then, they are aggregated based on the possibility view or the average with the importance weights of the group members. The importance weights of decision makers can be given by a supervisor directly, from a pariwise comparison matrix for their importance or based on uncertainty of their giving information. The aggregated preferences by the proposed approaches help the group members to understand one another and reach consensus.

References

[1] G. Desanctis and R. B. Gallupe. A foundation for the study of group decision support systems. *Management Science*, 33(5):589–609, 1987.

[2] A. A. Salo. Interactive decision aiding for group decision support. *European Journal of Operational Research*, 84:134–149, 1994.

[3] T. L. Saaty. *The Analytic Hierarchy Process*. McGraw-Hill, New York, 1980.

[4] M. T. Escobarm, J. Aguaron, and J. M. Moreno-Jimenez. A note on AHP group consistency for the row geometric mean priorization procedure. *European Journal of Operational Research*, 153(2):318–322, 2004.

[5] T. L. Saaty. Group decision making and the AHP. In *B.L. Golden et.al., Editors, The Analytic Hierarchy Process: Applications and Studies*, pages 59–67. McGraw-Hill, New York, 1989.

[6] R. F. Dyer and E. H. Forman. Group decision support with the Analytic Hierarchy Process. *Decision Support Systems*, 8:94–124, 1992.

[7] E. Forman and K. Peniwati. Aggregating individual judgments and priorities with the Analytic Hierarchy Process. *European Journal of Operational Research*, 108(1):165–169, 1998.

[8] K. Sugihara and H. Tanaka. Interval evaluations in the Analytic Hierarchy Process by possibilistic analysis. *Computational Intelligence*, 17(3):567–579, 2001.

[9] K. Sugihara, H. Ishii, and H. Tanaka. Interval priorities in AHP by interval regression analysis. *European Journal of Operational Research*, 158(3):745–754, 2004.

[10] L. M. de Campos, J. F. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty*, 2(2):167–196, 1994.

[11] H. Tanaka, K. Sugihara, and Y. Maeda. Non-additive measures by interval probability functions. *Information Sciences*, 164:209–227, 2004.

[12] H. Tanaka and T. Entani. Properties of evidences based on interval probabilities obtained by pairwise comparisons in AHP. In *Proceedings of Taiwan-Japan Symposium on Fuzzy Systems and Innovational Computing*, pages 35–40, 2006.

[13] H. Tanaka and P. Guo. *Possibilistic Data Analysis for Operations Research*. Physica-Verlag, A Springer-Verlag Company, Heidelberg, New York, 1999.

[14] T. Entani and H. Tanaka. Management of ignorance by interval probabilities. In *Proceedings of 2007 IEEE International Conference on Fuzzy Systems*, pages 841–846, 2007.

[15] J. Brazilia and B. Golany. AHP rank reversal, normalization and aggregation rules. *Information Systems and Operational Research*, 32(2):14–20, 1994.

[16] R. Ramanathan and L. S. Ganesh. Group reference aggregation methods in AHP: An evaluation and an intrinsic process for deriving members' weightages. *European Journal of Operational Research*, 79(2):249–268, 1994.

[17] T. Entani and H. Tanaka. Modified interval global weight in AHP. In *Proceedings of 9th Fuzzy Days Dortmund*, pages 415–424, 2006.

ISBN: 978-989-95079-6-8

A Monge algorithm for computing the Choquet integral on set systems

Ulrich Faigle[1] Michel Grabisch[2]

1. Zentrum für Angewandte Informatik Köln, Universität zu Köln
Weyertal 80, D-50931 Köln, Germany
2. Centre d'Economie de la Sorbonne, Université Paris I 106-112, Bd de l'Hôpital, 75013 Paris, France
Email: faigle@zpr.uni-koeln.de, michel.grabisch@univ-paris1.fr

Abstract— In works dealing with capacities (fuzzy measures) and the Choquet integral on finite spaces, it is usually considered that all subsets of the universe are measureable. Hence, all functions are integrable in the sense of Choquet. We consider the situation where some subsets are not measurable (not feasible), so that there are non-integrable functions. Since this is a severe limitation in applications, we study how to extend the Choquet integral to any function. Our results mainly deal with the case where the set of feasible subsets is a distributive lattice.

Keywords— Choquet integral, fuzzy measure, capacity, Monge algorithm, set system, distributive lattice

1 Introduction

The Choquet integral [1] has become a widely used tool in decision making and its properties are well known. With general spaces, the issue of measurability is closely related to integration, that is, a function is integrable if all subsets involved in the integration are measurable. In the case of the Choquet integral, these sets are the level sets of the integrand. Although these facts are standard when considering infinite spaces, most of the time it is considered that for finite spaces, all subsets are measurable. However, there are many situations where this assumption cannot be always true. For example, if the space N is a set of players in a game, some coalitions may not be feasible. If N is a set of states of nature, some event may be impossible to realize, if it is a set of criteria, some combinations of criteria may not correspond to conceivable objects, etc.

In cooperative game theory, already a great deal of research has been done concerning games with nonfeasible coalitions, making various assumptions on the structure of feasible coalitions (distributive lattices, as in Faigle and Kern [2], and in Grabisch and Lange [3], convex geometries as in Bilbao [4], antimatroids [5], regular set systems [6], etc.).

Computing the Choquet integral over these sets systems can be done simply by taking the usual definition. However, there will be many nonintegrable functions, for which the Choquet integral remains undefined. This could be a serious limitation in applications related to the above fields. To the best of our knowledge, the question of how to define the Choquet integral for any function on these sets systems has not been addressed. This is precisely what we aim for in this paper, supposing most of the time that the set system is a distributive lattice. We introduce a Monge algorithm, whose output coincides with the Choquet integral for every integrable function. We will show that the output of this algorithm, called the Monge-Choquet integral, is the smallest extension of the Choquet integral. Dually, we introduce the greatest extension and its associated dual Monge algorithm.

2 Preliminaries

In the whole paper, $N := \{1, 2, \ldots, n\}$.

2.1 Capacities and the Choquet integral

A *set system* $\mathcal{F}$ is any collection of sets in 2^N containing the empty set and N. For any set $A \subseteq N$, we define $\mathcal{F}(A) := \{F \in \mathcal{F} \mid F \subseteq A\}$. A *game* on $\mathcal{F}$ is any function $v : \mathcal{F} \to \mathbb{R}$ such that $v(\emptyset) = 0$. A game is *monotone* if for any $A, B \in \mathcal{F}$ such that $A \subseteq B$, we have $v(A) \leq v(B)$. A monotone game is usually called a *capacity* [1] or *fuzzy measure* [7].

Consider a function $f : N \to \mathbb{R}_+$, denoting $f(i)$ by f_i. We say that f is *measurable* if the family of sets $\{j \in N \mid f(j) \geq f(i)\}, i = 1, \ldots, n$ belongs to $\mathcal{F}$.

Definition 1. Let v be a game on $\mathcal{F}$, and $f : N \to \mathbb{R}_+$ be an measurable function. The *Choquet integral* of f w.r.t. v is defined by:

$$\int f \, dv := \sum_{i=1}^{n} (f_{\sigma(i)} - f_{\sigma(i-1)}) v(\{\sigma(i), \ldots, \sigma(n)\}) \quad (1)$$

with $f_{\sigma(0)} := 0$, $f_{\sigma(1)} \leq \cdots \leq f_{\sigma(n)}$.

For any $A \subseteq N$, let 1_A denotes the characteristic vector of A. A fundamental property of the Choquet integral is that

$$\int 1_A \, dv = v(A), \quad (2)$$

for any A in $\mathcal{F}$. We give some fundamental results on the Choquet integral, useful for the sequel.

Proposition 1. The Choquet integral is monotonic vs. the game, i.e., if $v \leq v'$ pointwise, then $\int f \, dv \leq \int f \, dv'$ for any measurable f.

Two functions $f, f' \in \mathbb{R}^n_+$ are *comonotonic* if there is no $i, j \in N$ such that $f_i > f_j$ and $f'_i < f'_j$. A functional $I : \mathbb{R}^n_+ \to \mathbb{R}$ is *comonotonic additive* if for any two comonotonic functions $f, f' \in \mathbb{R}^n_+$, we have $I(f + f') = I(f) + I(f')$.

ISBN: 978-989-95079-6-8

Proposition 2. (Characterization theorem of Schmeidler, established for continuous spaces [8]) Let $I : \mathbb{R}^n_+ \to \mathbb{R}$ be a functional. Then I is the Choquet integral w.r.t. a capacity v on 2^N if and only if I is nondecreasing, comonotonic additive, and $I(0) = 0$. Then v is uniquely determined by (2).

Consider the hypercube $[0,1]^n$. The *canonical simplices* of $[0,1]^n$ are those induced by any permutation on N as follows:

$$[0,1]^n_\sigma := \{x \in [0,1]^n \mid x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}\}.$$

Proposition 3. (Interpolation theorem, see Singer [9], and [10, Ch. 5]) The unique functional $I : [0,1]^n \to [0,1]$, whose value is fixed on vertices of $[0,1]^n$, being linear in each canonical simplex, continuous, and vanishing at 0 is the Choquet integral w.r.t. a game v on 2^N, with $v(A) = I(1_A)$, for all $A \subseteq N$.

2.2 Posets, lattices, and set systems

(see, e.g., [11]) A *poset* is any set P endowed with a binary relation $\leq$, being reflexive, antisymmetric and transitive (partial order). A *lattice* $(L, \leq)$ is a poset such that for any $x, y \in L$ their least upper bound, denoted by $x \vee y$, and greatest lower bound $x \wedge y$ always exist. When L is finite, there always exist a greatest element and a least element in L, called the *top* and the *bottom* element.

Let $(P, \leq)$ be any poset. $Q \subseteq P$ is a *downset* of P if $x \in Q$ and $y \leq x$ imply $y \in Q$. The set of all downsets of P is denoted by $\mathcal{O}(P)$. Observe that if Q, Q' are downsets of P, then so are $Q \cup Q'$ and $Q \cap Q'$. For any $x \in P$, we denote by $\downarrow x$ the downset generated by x (principal ideal of x), i.e.:

$$\downarrow x := \{y \in P \mid y \leq x\}.$$

More generally, for any subset Q of P, one can compute $\downarrow Q$, the *downset generated by* Q, defined by $\downarrow Q := \bigcup_{x \in Q} \downarrow x$. It is indeed a downset since downsets are closed under union.

For any two elements $x, y \in P$, x *is covered by* y or y *covers* x (denoted by $x \prec y$ or $y \succ x$) if $x < y$ and there is no z such that $x < z < y$. A sequence of elements such that $x \leq y_1 \leq y_2 \leq \cdots \leq z$ is called a *chain* from x to z. If in addition $x \prec y_1 \prec y_2 \cdots \prec z$, the chain is *maximal*.

A lattice is *distributive* if $\vee, \wedge$ obey distributivity. Any poset $(P, \leq)$ generates a distributive lattice of subsets of P ordered by inclusion, which is $\mathcal{O}(P)$ (and reciprocally). Its bottom and top elements are $\emptyset$ and P, respectively. In the distributive lattice $\mathcal{O}(P)$, any maximal chain C from bottom to top has length $|P|$, and corresponds to a permutation σ on P, such that $C = \{\emptyset, \{\sigma(n)\}, \{\sigma(n), \sigma(n-1)\}, \ldots, P\}$.

Let us consider as in Section 2.1 the set N of n elements. By defining a partial order $\leq$ on N, $(N, \leq)$ is a poset and $\mathcal{O}(N)$ is a distributive lattice with $\emptyset, N$ as bottom and top elements, hence it is a set system, having all its maximal chains from $\emptyset$ to N of length n. Conversely, a set system whose all maximal chains from $\emptyset$ to N are of length n is not necessarily a distributive lattice. It is called a *regular set system* [6]. An *antimatroid* $\mathcal{F}$ is a regular set system which is closed under union. A *convex geometry* is a regular set system closed under intersection.

2.3 The Monge algorithm

The original idea of the Monge algorithm goes back to [12]. Monge studied a geometric transportation problem in which a set of locations $s_1, \ldots, s_n$ of mass points has to be matched optimally (in the sense of minimizing the total cost) with another set of locations $t_1, \ldots, t_n$, and proved that optimality was reached if the transportation lines do not cross. This geometric fact can be expressed as follows: if the costs c_{ij} of matching objects s_i with t_j have the "uncrossing" property:

$$c_{ij} + c_{k\ell} \geq c_{\max(i,k),\max(j,\ell)} + c_{\min(i,k),\min(j,\ell)}$$

then the optimal matching is $(s_1, t_1), \ldots, (s_n, t_n)$. This is also called the "north-west corner rule". Translated into the language of set functions, the uncrossing property is in fact submodularity:

$$v(A) + v(B) \geq v(A \cup B) + v(A \cap B).$$

The following algorithm, which we call "Monge algorithm", is based on the previous ideas (see also the greedy algorithm of Lovász for maximizing a linear function over the core of a submodular game [13]). Let $A = [a_{ij}] \in \{0,1\}^{m \times n}$ be a $(0,1)$-matrix with m rows $A_1, \ldots, A_m$ and n columns $A^1, \ldots, A^n$. We assume throughout the paper that A does not contain the null row. Let us put $N = \{1, \ldots, n\}$ the set of column indices. Hence, any row of the matrix can be interpreted as the characteristic vector of a nonempty subset of N.

Consider an (input) vector $f \in \mathbb{R}^n_+$. The Monge algorithm produces an output line-vector $y \in \mathbb{R}^m_+$ as follows.

MONGE ALGORITHM

(M0) Initialization. Put $f' = f$, $A' = [a'_{ij}] = A$, $N' = N$, $y = 0 \in \mathbb{R}^m$.

(M1) If the current matrix A' is empty, STOP.

(M2) Let s be the smallest row index in A'.

(M3) Find the smallest value α of f'_j for all $j \in N'$ such that $a'_{sj} = 1$. Put t the smallest such j.

(M4) Updating: $y_s \leftarrow \alpha$, $f'_j \leftarrow f'_j - \alpha$ for all $j \in N'$ such that $a'_{sj} = 1$, $N' \leftarrow N' \setminus t$, suppress all rows A'_i in A' such that $a'_{it} = 1$. Go to (M1).

3 The Monge algorithm and the Choquet integral

Consider a set system $\mathcal{F}$ on N. It can be encoded by a $(0,1)$-matrix A as given in Section 2.3. The rows of A order the subsets of $\mathcal{F}$ in a particular way. An interesting case arises when this order is a linear extension of the decreasing inclusion order, that is, for any subsets S, T such that $S \subseteq T$, the corresponding rows A_{i_S}, A_{i_T} are such that $i_S \geq i_T$. We say that in this case the row order *is compatible with decreasing inclusion*. A row order compatible with increasing inclusion can be defined as well.

Consider next a game v on $\mathcal{F}$. It can be encoded as a vector in $\mathbb{R}^m$, with $m = |\mathcal{F}| - 1$. The next result shows that the Choquet integral can be computed by yv, where y is the output of the Monge algorithm, provided rows of A are ordered according to the (decreasing) inclusion order.

ISBN: 978-989-95079-6-8

Proposition 4. Let $\mathcal{F}$ be a set system, and an measurable function $f \in [0,1]^N$. Let v be a game on $\mathcal{F}$.

On the other hand, consider a (0,1)-matrix A encoding the set system $\mathcal{F}$ (except the empty set), such that the row order is compatible with decreasing inclusion. Then the output of the Monge algorithm for f is a vector y, such that yv is the Choquet integral defined in (1).

A question is whether the condition on row ordering is necessary to get this result. The answer is given in the next proposition.

Proposition 5. Assume A is a (0-1)-matrix with no two identical rows. For each nonempty subset S corresponding to a row of A, say A_{i_S}, the output of the algorithm is $y = 1_{i_S}$ (i.e., $y_i = 0$ for all i except for $i = i_S$) when the input is $f = 1_S$, if and only if the row order is compatible with decreasing inclusion.

In other words, for any set system $\mathcal{F}$ and any game v on $\mathcal{F}$, for $f = 1_S$, $S \in \mathcal{F}$, we have $yv = v(S)$, which is by (2) the output of the Choquet integral. This proves the necessity of the condition on the row order in Proposition 4.

4 Extended Choquet integrals for distributive lattices

For any set system $\mathcal{F}$, we define the *Monge-Choquet integral* as the output of the Monge algorithm for any nonnegative input vector. In this section, we restrict to the case of distributive lattices induced by a poset on N, denoted by $(N, \leq)$. We know from Section 3 that the Monge-Choquet integral coincides with the usual Choquet integral for measurable input vectors, provided the rows of A are arranged in an order extending the decreasing inclusion order (otherwise stated, this assumption holds from now on). Hence, the Monge-Choquet integral is an extension of the Choquet integral.

4.1 Computation of the Monge-Choquet integral

We show that the Monge-Choquet integral can be computed in a much simpler way, independent of the order of rows of A, provided this order is compatible with decreasing inclusion. Take f any input vector in $\mathbb{R}^n_+$. We can assume w.l.o.g. (this will be shown later on) that there is a unique permutation σ on N such that $f_{\sigma(1)} < \cdots < f_{\sigma(n)}$.

In the first step of the algorithm, we have $s = 1$, $\alpha = f_{\sigma(1)}$, and $t = \sigma(1)$. In the (initially null) vector y, we put $f_{\sigma(1)}$ at position 1, $\sigma(1)$ is deleted from N, any row in A containing $\sigma(1)$ is deleted, and $f' = f - f_{\sigma(1)}$.

In the second step, s corresponds to the largest subset F in $\mathcal{F}$ not containing $\sigma(1)$, which is uniquely determined by

$$F = \{i \in N \mid i \not\geq \sigma(1)\}$$

where $\geq$ is understood for the poset $(N, \leq)$. $F \in \mathcal{F}$ because F is a downset of $(N, \leq)$, and it is the unique largest such set in $\mathcal{F}$ because $\mathcal{F}$ is closed under union. Since it is the largest one not containing $\sigma(1)$, it is ranked first in A', hence s corresponds to this subset. Next, we look for $\sigma(i) \in F$ such that f' is minimum on F, and we put $t = \sigma(i)$, $\alpha = f_{\sigma(i)} - f_{\sigma(1)}$, and α is put in y at position s. Then $\sigma(i)$ is deleted, and all rows of A' containing $\sigma(i)$. The process continues till A' is empty.

The above shows that the following algorithm computes the Monge-Choquet integral in a simpler way. We recall that for any $A \subseteq N$, $\mathcal{F}(A) := \{F \in \mathcal{F} \mid F \subseteq A\}$.

COMPUTATION OF THE MONGE-CHOQUET INTEGRAL FOR DISTRIBUTIVE LATTICES

(MC0) Initialization: $N' \leftarrow N$.

(MC1) Take F to be the largest subset in $\mathcal{F}(N')$. It is given by $F = \{i \in N \mid i \not\geq j, \forall j \in N \setminus N'\}$. If $F = \emptyset$, goto (MC3).

(MC2) Find the minimum of f over F, call i the smallest index in F for which f is minimum. Put $N' \leftarrow N' \setminus i$. Go to (MC1).

(MC3) Denote by $F_1 = N, F_2, \ldots, F_k$ the sequence of subsets obtained in (MC1), and by $i_1, i_2, \ldots, i_k$ the sequence of indices obtained in (MC2). Then

$$\int f\, dv = \sum_{j=1}^{k} (f_{i_j} - f_{i_{j-1}}) v(F_i), \quad (3)$$

with $f_{i_0} := 0$.

Example 1. Take $n = 5$, and the poset represented on Fig. 1 (left). The corresponding set system $\mathcal{F}$ is on the right. Many

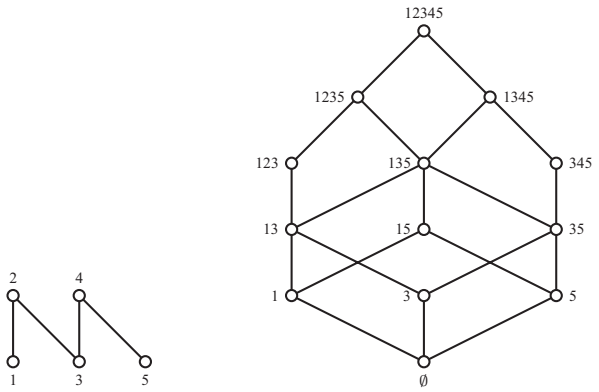


Figure 1: Example of poset with $N = \{1, 2, 3, 4, 5\}$ (left) and the corresponding set system $\mathcal{F}$.

permutations are not compatible with $\mathcal{F}$ (i.e., the induced maximal chain contains sets which are not in $\mathcal{F}$), for example 35241, 12345, 43125, etc. We can anyway compute the corresponding Monge-Choquet integrals:

$$35241 : \int f\, dv = f_3 v(N) + (f_5 - f_3) v(15) + (f_1 - f_5) v(1)$$

$$12345 : \int f\, dv = f_1 v(N) + (f_3 - f_1) v(345) + (f_5 - f_3) v(5)$$

$$43125 : \int f\, dv = f_4 v(N) + (f_3 - f_4) v(1235) + (f_1 - f_3) v(15) + (f_5 - f_1) v(5).$$

ISBN: 978-989-95079-6-8

4.2 Properties of the Monge-Choquet integral

The most important property to check is whether the functional is continuous. Obviously it is for a given permutation, but when passing from a permutation to another one, we have to check continuity. If continuity does not hold, we could even say that the integral is not well defined, because a given f may be rearranged in increasing order by several permutations.

Let us take f with permutation σ, such that

$$f_{\sigma(1)} < f_{\sigma(2)} < \cdots < f_{\sigma(i)} = f_{\sigma(i+1)} < \cdots < f_{\sigma(n)}.$$

Hence the permutation σ' obtained by a switch between $\sigma(i)$ and $\sigma(i+1)$ is also rearranging f in nondecreasing order. Observe that in any case, all "compatible" permutations can be obtained by a sufficient number of switches, hence it is enough to study what happens for one switch.

To avoid a heavy notation, and without loss of generality, we may assume that the two following orders are compatible with f:

$$1, 2, \ldots, i, i+1, \ldots, n$$
$$1, 2, \ldots, i+1, i, \ldots, n.$$

The first and second orders will be used by the algorithm when computing the integral of $f' := f + \epsilon 1_{i+1}$ and $f'' := f + \epsilon 1_i$ respectively. Let us check whether the two expressions coincide when ϵ tends towards 0. Clearly, the sequences $F'_1, \ldots, F'_{k'}$ and $F''_1, \ldots, F''_{k''}$ for f' and f'' coincide as long as both $i, i+1$ belong to the sets, and also from the point where both $i, i+1$ disappear. Let us denote respectively by F_0 and F^0 these sets (note that $F^0 = \emptyset$ is possible). Let i_0 be the index of (MC2) when $F = F_0$. Then several cases can occur.

(i) $i_0 < i$ and $i_0 < i+1$ (in the poset). Then F in the next step (MC1) will contain neither i nor $i+1$ for both f', f'' (hence $F = F^0$). Consequently, there will be no difference between the two expressions.

(ii) $i_0 < i$, $i_0 \not< i+1$. Then, the next terms obtained while processing $i, i+1$ will be:

$$\text{for } f'\text{: } (f'_{i+1} - f'_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\})$$
$$\text{for } f''\text{: } (f''_{i+1} - f''_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}).$$

(iii) $i_0 < i+1$, $i_0 \not< i$. Similarly,

$$\text{for } f'\text{: } (f'_i - f'_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\})$$
$$\text{for } f''\text{: } (f''_i - f''_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}).$$

(iv) $i_0 \not< i$, $i_0 \not< i+1$, and $i < i+1$. Then, the next terms obtained while processing $i, i+1$ will be:

$$\text{for } f'\text{: } (f'_i - f'_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\})$$
$$\text{for } f''\text{: } (f''_{i+1} - f''_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}) +$$
$$(f''_i - f''_{i+1})v(F_0 \cap \{j \in N \mid j \not\geq i_0, j \not\geq i+1\}).$$

(v) $i_0 \not< i$, $i_0 \not< i+1$, and $i+1 < i$. Same as above, permuting the roles of f' and f''.

(vi) $i_0 \not< i$, $i_0 \not< i+1$, and $i || i+1$. Then we obtain

$$\text{for } f'\text{: } (f'_i - f'_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}) +$$
$$(f'_{i+1} - f'_i)v(F_0 \cap \{j \in N \mid j \not\geq i_0\}, j \not\geq i)$$
$$\text{for } f''\text{: } (f''_{i+1} - f''_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}) +$$
$$(f''_i - f''_{i+1})v(F_0 \cap \{j \in N \mid j \not\geq i_0, j \not\geq i+1\}).$$

In all cases, the expressions become identical when ϵ tends towards 0.

In addition, this proves that in the algorithm, when taking the smallest i when several such i's exist, this choice is unimportant.

From continuity and (3), we easily deduce the following:

Proposition 6. The Monge-Choquet integral is nondecreasing w.r.t. the integrand, positively homogeneous, and is comonotonic additive.

4.3 Smallest and largest extended Choquet integrals

From (3) and the algorithm, it is easy to compute the Monge-Choquet integral for any binary function 1_A, $A \subseteq N$. From Proposition 5, we already know that for any $F \in \mathcal{F}$, $\int 1_F \, dv = v(F)$. Let us define $\hat{v}(A) := \int 1_A \, dv$, for any $A \subseteq N$, which could be seen as an extension of v on 2^N.

Proposition 7. For any $A \subseteq N$, $\hat{v}(A)$ is obtained by

$$\hat{v}(A) = v(F), \quad \text{with } F \text{ the largest subset of } \mathcal{F}(A).$$

Moreover, if v is a capacity, then so is $\hat{v}$, and it is the smallest extension of v over 2^N.

From continuity of the Monge-Choquet integral and Proposition 3, we deduce the following fundamental result.

Proposition 8. For any $f \in \mathbb{R}^n_+$,

$$\int f \, dv = \int f \, d\hat{v}$$

where the left integral is the Monge-Choquet integral, and the right one, the classical Choquet integral.

Hence, the Monge-Choquet integral inherits all properties from the Choquet integral w.r.t. the smallest extension of v over 2^N. In particular, by monotonicity of the integral w.r.t. the capacity (see Proposition 1), we get by Proposition 2:

Corollary 1. Let v be a capacity on $\mathcal{F}$. The Monge-Choquet integral w.r.t. v is the smallest functional $I : \mathbb{R}^n_+ \to \mathbb{R}$ being nondecreasing, comonotonic additive, and such that $I(1_F) = v(F)$ for each F in $\mathcal{F}$.

It is easy to define the largest extension of v over 2^N, denoted by $\check{v}$. It is given by

$$\check{v}(A) := v(\downarrow A)$$

where $\downarrow A$ is the downset generated by A in the poset $(N, \leq)$. Indeed, this is the smallest set in $\mathcal{F}$ containing A. It is possible to obtain $\check{v}$ by an algorithm which is dual to the Monge algorithm.

ISBN: 978-989-95079-6-8

DUAL MONGE ALGORITHM. We assume that the row order is compatible with *increasing* inclusion.

(DM0) Init: $A' = [a'_{ij}] = A$, $y = 0 \in \mathbb{R}^m$, $\alpha =$largest value of f, $N' = N$.

(DM1) Choose the smallest index $t \in N'$ such that $f_t = \alpha$. Put $N' \leftarrow N' \setminus t$.

(DM2) Suppress all rows in A' not containing t (i.e., $a'_{s,t} = 0$). Take the smallest row index s in A'.

(DM3) If $N' = \emptyset$ or s is the last row of A' (i.e., it corresponds to N), put $y_s = f_t$ and STOP. Otherwise, look for the highest value α of f_j for all j such that $a'_{s,j} = 0$, put $y_s = f_t - \alpha$, and go to (DM1).

Again, due to the structure of F, a simpler version can be given.

COMPUTATION OF THE DUAL MONGE-CHOQUET INTEGRAL FOR DISTRIBUTIVE LATTICES

(DMC0) Put $N' = \{i_0\}$, where i_0 is the smallest index maximising f.

(DMC1) Take F the smallest set in $\mathcal{F}$ containing N', which is given by $F =\downarrow N'$. If $F = N$, go to (DMC3).

(DMC2) Find the smallest index i maximising f over $N \setminus F$. Put $N' \leftarrow N' \cup \{i\}$. Go to (DMC1).

(DMC3) Denote by $F_1, \ldots, F_k$ and $i_0, i_1, \ldots, i_{k-1}$ the sequence of sets obtained in (DMC1) and the sequence of indices obtained in (DMC0) and (DMC2) respectively. Then

$$\int f\, dv = \sum_{j=1}^{k} (f_{i_{j-1}} - f_{i_j}) v(F_j) \qquad (4)$$

with $f_{i_k} := 0$.

Example 2. Taking the same poset as in Example 1, we obtain:

$$35241 : \int f\, dv = (f_1 - f_4)v(1) + (f_4 - f_2)v(1345) + f_2 v(N)$$

$$12345 : \int f\, dv = (f_5 - f_4)v(5) + (f_4 - f_2)v(345) + f_2 v(N)$$

$$43125 : \int f\, dv = (f_5 - f_2)v(5) + (f_2 - f_4)v(1235) + f_4 v(N).$$

Proposition 9. The dual Monge-Choquet integral is continuous.

The consequence is that, applying the same reasoning as above, we can deduce the following.

Proposition 10. (i) For any $f \in \mathbb{R}^n_+$,

$$\int^* f\, dv = \int f\, d\check{v}$$

where the left integral is the dual Monge-Choquet integral, and the right one, the classical Choquet integral.

(ii) If v is a capacity on $\mathcal{F}$, then the dual Monge-Choquet integral is the greatest functional $I : \mathbb{R}^n_+ \to \mathbb{R}_+$ being nondecreasing, comonotonic additive, and such that $I(1_F) = v(F)$ for each $F \in \mathcal{F}$.

(iii) For any measurable f, the Monge-Choquet integral and the dual Monge-Choquet integral coincide.

5 The case of regular set systems, antimatroids and convex geometries

5.1 Regular set systems

We know from Section 2.1 that a function is measurable if it exists a permutation σ on N ordering f such that the induced sets $\{\sigma(i), \ldots, \sigma(n)\}$, $i = 1, \ldots, n$, form a maximal chain from $\emptyset$ to N in $\mathcal{F}$, of length n. Hence if $\mathcal{F}$ does not contain maximal chains of length n, there will be no measurable function. This has motivated the definition of regular set systems, whose maximal chains from $\emptyset$ to N are all of length n. In this respect, it would be interesting to know what happens for nonmeasurable functions over regular set systems. It is not difficult to see that, unfortunately, the output of the Monge algorithm depends on the row order, even if it is compatible with inclusion (see example below). The reason is that a regular set system need not be closed under union (nor under intersection), and this property is fundamental for proving results in Section 4. Therefore, the Monge-Choquet integral is not well-defined in this case.

Example 3. Consider $n = 5$ and the regular set system of Fig. 3. Remark that it is not closed under union nor intersection. Hence it is neither an antimatroid nor a convex geometry. Take any function f such that $f_4 < f_3 < f_2 < f_1 < f_5$. Then,

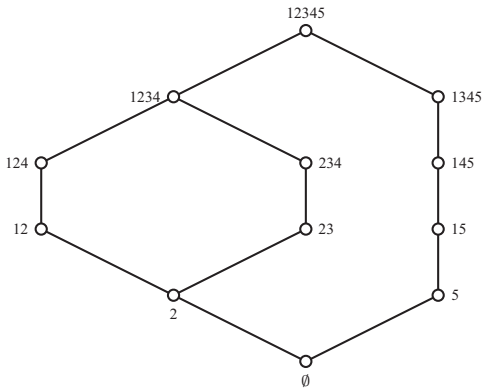


Figure 2: Example of regular set system

in the 1st iteration, $s = 1$, $\alpha = f_4$, $t = 4$, and $f' = f - f_4$. At second iteration, 4 has been discarded, so we have 3 candidates for the next subset: 12, 23, and 15. If 12 is ranked first in the matrix A', we get $\alpha = f_2$, $t = 2$, and 2 is discarded, so only 1 remains. Finally, we get:

$$\int d\, dv = f_4 v(N) + (f_2 - f_4)v(12) + (f_1 - f_2)v(1).$$

ISBN: 978-989-95079-6-8

Now, if 23 is ranked first, $\alpha = f_3$, $t = 3$, and we will get:

$$\int f\, dv = f_4 v(N) + (f_3 - f_4)v(23) + (f_2 - f_3)v(2).$$

Hence, the result depends on the ordering.

5.2 Antimatroids and convex geometries

Antimatroids are regular set systems closed under union. This crucial property (as exemplified above) makes most of previous results to hold. More precisely, there will be always a unique largest subset in $\mathcal{F}(N')$ for all $N' \subseteq N$. Consequently, the algorithm of computation of the Monge-Choquet integral and especially Equation (3) given in Section 4.1 remain valid, except for Step (MC1), where F is still the largest subset in $\mathcal{F}(N')$, but now it is no more possible to give it explicitly.

An important question is whether continuity still holds. Using a similar argument, one can prove that continuity still holds in the case of antimatroids. A sketch of the proof goes as follows. Keeping the same notation as in Section 4.2, consider functions f', f'', whose orderings differ only on i and $i+1$. There is a common chain from N to F_0 for both functions f' and f''. From F_0 which contains both i and $i+1$, functions f' and f'' may have different chains, but these chains will necessarily rejoin in F^0 (because there is a unique largest subset contained in $\{i+2, \ldots, n\}$). At F_0, the new term in the integral is $(f'_i - f'_k)v(F_0)$ for f' and $(f''_{i+1} - f''_k)v(F_0)$ for f'' for some k among $1, \ldots, i-1$. At F^0, the new term is $(f'_k - f'_{i+1})v(F^0)$ for f' and $(f''_k - f''_i)v(F^0)$ for f'' for some k in F^0, assuming $i, i+1$ are present in the preceding step (other cases work similarly). Between F_0 and F^0, there is at most one set F on each chain, where the new term will be $(f'_{i+1} - f'_i)v(F)$, and similarly for f''. When ϵ tends towards 0, both integrals coincide.

Consequently, Propositions 6, 7, 8, and Corollary 1 still hold.

On the contrary, it is no more possible to define the largest extension $\check{v}$ and consequently the dual Monge-Choquet integral, because since an antimatroid is not closed under intersection, it may happen that there is no smallest set of $\mathcal{F}$ containing in a given subset of N.

In a dual way, since convex geometries are closed under intersection but not under union, the dual Monge-Choquet integral exists and possesses all properties given in Section 4.3, while the Monge-Choquet integral is not well-defined.

References

[1] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[2] U. Faigle and W. Kern. The Shapley value for cooperative games under precedence constraints. *Int. J. of Game Theory*, 21:249–266, 1992.

[3] M. Grabisch and F. Lange. Games on lattices, multichoice games and the Shapley value: a new approach. *Mathematical Methods of Operations Research*, 65:153–167, 2007.

[4] J.M. Bilbao and P.H. Edelman. The Shapley value on convex geometries. *Discrete Applied Mathematics*, 103:33–40, 2000.

[5] E. Algaba, J. M. Bilbao, R. van den Brink, and A. Jiménez-Losada. Cooperative games on antimatroids. *Discrete Mathematics*, 282:1–15, 2004.

[6] A. Honda and M. Grabisch. An axiomatization of entropy of capacities on set systems. *Eur. J. of Operational Research*, 190:526–538, 2008.

[7] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[8] D. Schmeidler. Integral representation without additivity. *Proc. of the Amer. Math. Soc.*, 97(2):255–261, 1986.

[9] I. Singer. Extensions of functions of 0-1 variables and applications to combinatorial optimization. *Numerical Functional Analysis and Optimization*, 7(1):23–62, 1984.

[10] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*. Number 127 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2009.

[11] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Orders*. Cambridge University Press, 1990.

[12] G. Monge. Déblai et remblai. Mém. de l'Académie des Sciences, Paris, 1781.

[13] L. Lovász. Submodular functions and convexity. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical programming. The state of the art*, pages 235–257. Springer Verlag, 1983.

ISBN: 978-989-95079-6-8

Decomposition of Interval-Valued Fuzzy Morphological Operations by Weak $[\alpha_1, \alpha_2]$-cuts

Tom Mélange, Mike Nachtegael, Etienne Kerre[1] Peter Sussner[2]

1.Ghent University, Dept. of Applied Math. and Computer Science
Krijgslaan 281-S9, 9000 Ghent, Belgium
2.University of Campinas, Dept. of Applied Mathematics
Campinas, SP, 13083-859, Brazil
Email: {tom.melange,mike.nachtegael,etienne.kerre}@ugent.be, sussner@ime.unicamp.br

Abstract— *One of the extensions of binary morphology to greyscale images is given by the classical fuzzy mathematical morphology. Interval-valued fuzzy mathematical morphology further extends the latter theory by now also allowing uncertainty in the grey values of the image. In this paper, the decomposition of the interval-valued fuzzy morphological operations into their binary counterparts is studied both in a general continuous framework and a discrete framework. It will be shown that some properties that do not hold in the continuous framework, do hold in the discrete framework, which is the framework that is used in practice.*

Keywords— Decomposition, interval-valued fuzzy sets, mathematical morphology.

1 Introduction

The aim of many image processing theories is to extract specific information out of an image. Mathematical morphology is one of those theories. The basic morphological operations dilation, erosion, opening and closing transform a given image into another image by the help of a structuring element. In the original binary morphology [1], both the image and the structuring element needed to be binary images (i.e., black-white). The threshold approach [1] extended the binary morphology by allowing greyscale images. Another extension, the umbra approach [2], even allowed both greyscale images and structuring elements. Some time later, one more greyscale approach was introduced, based on fuzzy set theory [3, 4]. Greyscale images and fuzzy sets can namely both be modelled as a mapping from a universe $\mathcal{U}$ into the unit interval $[0, 1]$. Remark however that fuzzy set theory is only used as a tool here and not to deal with uncertainty. Recently, also extensions of the fuzzy mathematical morphology have arised [5, 6]. The interval-valued fuzzy extension that we consider in this paper, now allows uncertainty regarding the grey values in the image, since a coordinate in the image domain is no longer mapped onto a specific grey value belonging to the unit interval $[0, 1]$, but onto an interval of grey values ($\subseteq [0, 1]$) to which the uncertain grey value is expected to belong. The uncertainty in grey values can have several causes. (i) In practice, we have to deal with the fact that any device rounds off a captured grey value to an element of a finite set of allowed values; (ii) Taking several shots of the same image may result in different grey values for some of the image pixels. This can sometimes happen under identical recording circumstances and is very likely to be the case under changing circumstances (e.g. illumination changes due to clouds passing the sun). The difference in grey value can also be the result of a slight shift in position of the camera or an object in the image between takes. This will especially cause (mostly large) uncertainty at pixels belonging to the edge of an object. (iii) There might also exist uncertainty regarding the grey values in the structuring element that is used. This structuring element can be chosen by the user and sometimes the user might doubt which weight (importance) to give to a certain pixel in the structuring element. In all of the above cases, it might be advisible to work with intervals of likely values instead of one specific value. The interval-valued fuzzy set theory now thus also serves as a model and not only as a tool [6].

In this paper, the relationships between the weak $[\alpha_1, \alpha_2]$-cuts of the interval-valued fuzzy dilation, erosion, opening and closing and the corresponding binary operations are studied. Such decomposition properties are useful (i) to compute interval-valued fuzzy morphological operations by applying several binary operations and combining the obtained results and (ii) to approximate interval-valued fuzzy morphological operations by only considering a finite number of $[\alpha_1, \alpha_2]$-cuts.

The remainder of the paper is organized as follows: the basic notions on interval-valued fuzzy mathematical morphology are introduced in section 2; section 3 studies the decomposition of the interval-valued fuzzy morphological operations by weak $[\alpha_1, \alpha_2]$-cuts in a continuous framework and investigates whether a discrete framework affects those results. Section 4 finally, concludes the paper.

2 Interval-Valued Fuzzy Mathematical Morphology

2.1 Interval-valued Fuzzy Set Theory

A fuzzy set [7] F defined over a universe $\mathcal{U}$, maps every element $u \in \mathcal{U}$ onto its membership degree $F(u) \in [0, 1]$ in that set F and allows in this way a gradual transition between belonging to (membership degree 1) and not belonging to (membership degree 0). Interval-valued fuzzy sets [8] extend the classical fuzzy sets by allowing uncertainty concerning the membership degree. They map an element of the universe onto an interval of values instead of one specific membership degree. In other words, they are modelled by mappings from the universe $\mathcal{U}$ into the class of closed intervals $L^I = \{[x_1, x_2] | [x_1, x_2] \subseteq [0, 1]\}$. This means that for an interval-valued fuzzy set G in a universe $\mathcal{U}$, $G(u) = [G_1(u), G_2(u)] \subseteq [0, 1]$, $\forall u \in \mathcal{U}$. In this paper, the class of

ISBN: 978-989-95079-6-8

interval-valued sets over the universe $\mathcal{U}$ will be denoted by $\mathcal{IVFS}(\mathcal{U})$. The lower and upper bound of an element x of L^I is denoted by respectively x_1 and x_2, i.e., $x = [x_1, x_2]$. Further, we will restrict the universe $\mathcal{U}$ in the sequel to $\mathbb{R}^n$, corresponding to the coordinates of an n-dimensional image.

Consider now the following partial ordering $\leq_{L^I}$ on L^I:

$$x \leq_{L^I} y \Leftrightarrow x_1 \leq y_1 \text{ and } x_2 \leq y_2,\ \forall x, y \in L^I.$$

It can be shown that $(L^I, \leq_{L^I})$ forms a complete lattice [9]. The infimum and supremum of a subset S of L^I are then given by $\inf S = [\inf_{x \in S} x_1, \inf_{x \in S} x_2]$ and $\sup S = [\sup_{x \in S} x_1, \sup_{x \in S} x_2]$ respectively. In the remainder, we will use the notations 0_{L^I} for $\inf L^I = [0, 0]$ and 1_{L^I} for $\sup L^I = [1, 1]$ and we will also use the following orderings on L^I ($\forall x, y \in L^I$):

$$\begin{aligned} x \ll_{L^I} y &\Leftrightarrow x_1 < y_1 \text{ and } x_2 < y_2, \\ x \geq_{L^I} y &\Leftrightarrow y \leq_{L^I} x, \\ x \gg_{L^I} y &\Leftrightarrow y \ll_{L^I} x. \end{aligned}$$

2.2 Binary and Interval-Valued Fuzzy Mathematical Morphological Operations

In this paper, the decomposition of the interval-valued fuzzy morphological operations into their binary counterparts is studied. Therefore, we first refresh the binary morphological operations.

Definition 1. [*I*] *Let $A, B \subseteq \mathbb{R}^n$. The binary dilation $D(A, B)$, erosion $E(A, B)$, closing $C(A, B)$ and opening $O(A, B)$ are the sets given by:*

$$\begin{aligned} D(A, B) &= \{y | T_y(B) \cap A \neq \emptyset\}, \\ E(A, B) &= \{y | T_y(B) \subseteq A\}, \\ C(A, B) &= E(D(A, B), -B), \\ O(A, B) &= D(E(A, B), -B), \end{aligned}$$

with $T_y(B) = \{x \in \mathbb{R}^n | x - y \in B\}$ and $-B = \{-b | b \in B\}$.

It is clear that the notions of intersection and inclusion are quite important in the above definitions. Hence, to extend the binary morphological operations to interval-valued fuzzy operators, the underlying Boolean conjunction and implication are extended by conjunctors and implicators on L^I. First, however, we extend the Boolean negation by a negator on L^I.

A *negator* $\mathcal{N}$ on L^I is a decreasing $L^I - L^I$ mapping that coincides with the Boolean negation on $\{0, 1\}$ ($\mathcal{N}(0_{L^I}) = 1_{L^I}$ and $\mathcal{N}(1_{L^I}) = 0_{L^I}$). If $(\forall x \in L^I)(\mathcal{N}(\mathcal{N}(x)) = x)$, then the negator $\mathcal{N}$ is called *involutive*. The standard negator $\mathcal{N}_s$, given by $\mathcal{N}_s([x_1, x_2]) = [1-x_2, 1-x_1]$, for all $x = [x_1, x_2] \in L^I$, is an example of an involutive negator on L^I.

A *conjunctor* $\mathcal{C}$ on L^I is an increasing $(L^I)^2 - L^I$ mapping that coincides with the Boolean conjunction on $\{0, 1\}^2$ ($\mathcal{C}(0_{L^I}, 0_{L^I}) = \mathcal{C}(0_{L^I}, 1_{L^I}) = \mathcal{C}(1_{L^I}, 0_{L^I}) = 0_{L^I}$ and $\mathcal{C}(1_{L^I}, 1_{L^I}) = 1_{L^I}$). If a conjunctor $\mathcal{C}$ satisfies $(\forall x \in L^I)(\mathcal{C}(1_{L^I}, x) = \mathcal{C}(x, 1_{L^I}) = x)$, then it is called a *semi-norm* on L^I. Furthermore, if a semi-norm $\mathcal{C}$ is commutative and associative, then we call it a *t-norm* on L^I. The conjunctor $\mathcal{C}_{\min}$, defined by $\mathcal{C}_{\min}(x, y) = [\min(x_1, y_1), \min(x_2, y_2)]$, for all $(x, y) \in (L^I)^2$, is an example of a t-norm on L^I.

Finally, an *implicator* $\mathcal{I}$ on L^I is a hybrid monotonic $(L^I)^2 - L^I$ mapping (i.e., decreasing in the first argument and increasing in the second argument) that coincides with the Boolean implication on $\{0, 1\}^2$ ($\mathcal{I}(0_{L^I}, 0_{L^I}) = \mathcal{I}(0_{L^I}, 1_{L^I}) = \mathcal{I}(1_{L^I}, 1_{L^I}) = 1_{L^I}$ and $\mathcal{I}(1_{L^I}, 0_{L^I}) = 0_{L^I}$). It can be checked that for every implicator $\mathcal{I}$, the operation $\mathcal{N}_\mathcal{I}$ defined by $\mathcal{N}_\mathcal{I}(x) = \mathcal{I}(x, 0_{L^I}), \forall x \in L^I$, is a negator on L^I. If an implicator $\mathcal{I}$ satisfies $(\forall x \in L^I)(\mathcal{I}(1_{L^I}, x) = x)$, then it is called a *border implicator* on L^I. Furthermore, if a border implicator $\mathcal{I}$ is contrapositive w.r.t. its induced negator, i.e., $(\forall (x, y) \in (L^I)^2)(\mathcal{I}(x, y) = \mathcal{I}(\mathcal{N}_\mathcal{I}(y), \mathcal{N}_\mathcal{I}(x)))$, and if it fulfills the exchange principle, i.e., $(\forall (x, y, z) \in (L^I)^3)(\mathcal{I}(x, \mathcal{I}(y, z)) = \mathcal{I}(y, \mathcal{I}(x, z)))$, then we call it a *model implicator* on L^I. The implicator $\mathcal{I}_{\max, \mathcal{N}_s}$ given by $\mathcal{I}_{\max, \mathcal{N}_s}(x, y) = [\max(1 - x_2, y_1), \max(1 - x_1, y_2)]$, for all $(x, y) \in (L^I)^2$, is an example of a model implicator on L^I.

Using the above concepts, we can extend the binary morphological operations to the interval-valued fuzzy case.

Definition 2. *Let $\mathcal{C}$ be a conjunctor on L^I, let $\mathcal{I}$ be an implicator on L^I, and let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$. The interval-valued fuzzy dilation $D^I_\mathcal{C}(A, B)$, erosion $E^I_\mathcal{I}(A, B)$, closing $C^I_{\mathcal{C},\mathcal{I}}(A, B)$ and opening $O^I_{\mathcal{C},\mathcal{I}}(A, B)$ are the interval-valued fuzzy sets in $\mathbb{R}^n$ defined for all $y \in \mathbb{R}^n$ by:*

$$\begin{aligned} D^I_\mathcal{C}(A, B)(y) &= \sup_{x \in T_y(d_B) \cap d_A} \mathcal{C}(B(x - y), A(x)), \\ E^I_\mathcal{I}(A, B)(y) &= \inf_{x \in T_y(d_B)} \mathcal{I}(B(x - y), A(x)), \\ C^I_{\mathcal{C},\mathcal{I}}(A, B)(y) &= E^I_\mathcal{I}(D^I_\mathcal{C}(A, B), -B)(y), \\ O^I_{\mathcal{C},\mathcal{I}}(A, B)(y) &= D^I_\mathcal{C}(E^I_\mathcal{I}(A, B), -B)(y), \end{aligned}$$

with $d_A = \{x | x \in \mathbb{R}^n$ and $A(x) \neq 0_{L^I}\}$, $d_B = \{x | x \in \mathbb{R}^n$ and $B(x) \neq 0_{L^I}\}$ and $(-B)(x) = B(-x), \forall x \in \mathbb{R}^n$.

Remark that if $y \notin D(d_A, d_B)$, then $D^I_\mathcal{C}(A, B)(y) = 0_{L^I}$.

In [11] it is shown that fuzzy mathematical morphology is compatible with binary morphology and if we restrict ourselves to semi-norms and border implicators it is also compatible with greyscale morphology based on the threshold approach. Since the interval-valued fuzzy morphology is compatible with the fuzzy morphology and because we want to preserve also the compatibility with the threshold approach, we will restrict ourselves in the remainder to semi-norms and border implicators on L^I.

2.3 The Discrete Framework

For the practical processing of an image, one has to deal with the technical limitations of a computer. To store an image, the domain is downsampled from the continuous space $\mathbb{R}^n$ to the discrete space $\mathbb{Z}^n$, and the image is represented by a matrix with a given number of rows and columns. Also the possible grey values are sampled. Grey values do not longer belong to the complete continuous unit interval $[0, 1]$, but are downsampled to a finite subchain of it. In the interval-valued fuzzy mathematical morphology, the used intervals of possible grey values are now thus subsets of the finite subchain $L^I_{r,s}$ of L^I, with $L^I_{r,s} = \{[\frac{r-k}{r-1}, \frac{s-l}{s-1}] | k, l \in \mathbb{Z}$ and $1 \leq k \leq r$ and $1 \leq l \leq s\}$ for given integers r and s. We denote the class of all interval-valued fuzzy sets in $\mathbb{Z}^n$ with membership intervals in $L^I_{r,s}$ as $\mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$. Further, remark that for an interval-valued fuzzy set $A \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n), \forall x \in \mathbb{Z}^n, A_1(x) \in I_r = \{\frac{r-k}{r-1} | k \in \mathbb{Z}$ and $1 \leq k \leq r\}$ and $A_2(x) \in I_s\{\frac{s-l}{s-1} | l \in \mathbb{Z}$ and $1 \leq l \leq s\}$.

ISBN: 978-989-95079-6-8

The concepts of negators, conjunctors and implicators on the chain $L^I_{r,s}$ can be adopted from the previous subsection by replacing L^I by $L^I_{r,s}$.

The definitions of the discrete interval-valued fuzzy dilation and erosion can now be written as follows:

Definition 3. *Let $\mathcal{C}$ be a conjunctor on $L^I_{r,s}$, let $\mathcal{I}$ be an implicator on $L^I_{r,s}$, and let $A, B \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$. The discrete interval-valued fuzzy dilation $D^I_{\mathcal{C}}(A,B) \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$ and erosion $E_{\mathcal{I}}(A,B) \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$ are defined by:*

$$\begin{aligned} D^I_{\mathcal{C}}(A,B)(y) &= [\max_{x \in T_y(d_B) \cap d_A} \mathcal{C}(B(x-y), A(x))_1, \\ &\quad \max_{x \in T_y(d_B) \cap d_A} \mathcal{C}(B(x-y), A(x))_2], \\ E^I_{\mathcal{I}}(A,B)(y) &= [\min_{x \in T_y(d_B)} \mathcal{I}(B(x-y), A(x))_1, \\ &\quad \min_{x \in T_y(d_B)} \mathcal{I}(B(x-y), A(x))_2]. \end{aligned}$$

3 Decomposition of Interval-valued Fuzzy Morphological Operations

3.1 Weak $[\alpha_1, \alpha_2]$-cuts

We first introduce the different weak $[\alpha_1, \alpha_2]$-cuts of an interval-valued fuzzy set [10].

Definition 4. *Let $A \in \mathcal{IVFS}(\mathbb{R}^n)$. For $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$, the weak $[\alpha_1, \alpha_2]$-cut $A^{\alpha_2}_{\alpha_1}$ of A is given by:*

$$\begin{aligned} A^{\alpha_2}_{\alpha_1} &= \{x | x \in \mathbb{R}^n,\ A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2\} \\ &= \{x | x \in \mathbb{R}^n \text{ and } A(x) \geq_{L^I} [\alpha_1, \alpha_2]\}. \end{aligned}$$

For $\alpha_1 \in]0,1]$, the weak α_1-subcut A_{α_1} of A is given by:

$$A_{\alpha_1} = \{x | x \in \mathbb{R}^n \text{ and } A_1(x) \geq \alpha_1\}.$$

For $\alpha_2 \in]0,1]$, the weak α_2-supercut A^{α_2} of A is given by:

$$A^{\alpha_2} = \{x | x \in \mathbb{R}^n \text{ and } A_2(x) \geq \alpha_2\}.$$

The cases $[\alpha_1, \alpha_2] = 0_{L^I}$, $\alpha_1 = 0$ and $\alpha_2 = 0$ are excluded for respectively the weak $[\alpha_1, \alpha_2]$-cut, α_1-subcut and α_2-supercut. Since $\{x | x \in \mathbb{R}^n,\ A_1(x) \geq 0 \text{ and } A_2(x) \geq 0\} = \{x | x \in \mathbb{R}^n \text{ and } A_1(x) \geq 0\} = \{x | x \in \mathbb{R}^n \text{ and } A_2(x) \geq 0\} = \mathbb{R}^n$, these cases don't yield new information.

3.2 Decomposition of the Interval-valued Fuzzy Dilation

Lemma 1. *If $\mathcal{C}$ is a semi-norm on L^I, then it holds that $\mathcal{C} \leq \mathcal{C}_{\min}$, i.e.: $(\forall (x,y) \in (L^I)^2)(\mathcal{C}(x,y) \leq_{L^I} \mathcal{C}_{\min}(x,y))$.*

Proof. Let $\mathcal{C}$ be a semi-norm on L^I, then it holds for all $(x,y) \in (L^I)^2$ that on the one hand $\mathcal{C}(x,y) \leq_{L^I} \mathcal{C}(x, 1_{L^I})$ and $\mathcal{C}(x, 1_{L^I}) = x$, and on the other hand $\mathcal{C}(x,y) \leq_{L^I} \mathcal{C}(1_{L^I}, y)$ and $\mathcal{C}(1_{L^I}, y) = y$, from which it can be concluded that $\mathcal{C}(x,y) \leq_{L^I} \mathcal{C}_{\min}(x,y)$. □

3.2.1 Decomposition by weak $[\alpha_1, \alpha_2]$-cuts

There is no relationship between the weak $[\alpha_1, \alpha_2]$-cut $D^I_{\mathcal{C}}(A,B)^{\alpha_2}_{\alpha_1}$ and the binary dilation $D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1})$ that holds in general for an arbitrary semi-norm $\mathcal{C}$.

Example 1. Let $[\alpha_1, \alpha_2] = [1/4, 1]$, $A(x) = [x/2, x]$ for all $x \in [0,1[$, $A(x) = 0_{L^I}$ for all $x \in \mathbb{R} \backslash [0,1[$, $B(x) = 1_{L^I}$ for all $x \in [0,1]$ and $B(x) = 0_{L^I}$ for all $x \in \mathbb{R} \backslash [0,1]$. So $d_A =]0,1[$, $d_B = [0,1]$ and $D(d_A, d_B) =]-1, 1[$ and consequently $0 \in D(d_A, d_B)$. Let $\mathcal{C}$ be an arbitrary semi-norm.

It then holds that

$$D^I_{\mathcal{C}}(A,B)(0) = \sup_{x \in T_0(d_B) \cap d_A} \mathcal{C}(B(x), A(x)) = \sup_{x \in]0,1[} \mathcal{C}(1_{L^I}, [x/2, x]) = \sup_{x \in]0,1[} [x/2, x] = [1/2, 1],$$

which means that $0 \in D^I_{\mathcal{C}}(A,B)^1_{0.25}$.

On the other hand, however, since $A^1_{0.25} = \emptyset$ also $D(A^1_{0.25}, B^1_{0.25}) = \emptyset$ and thus $0 \notin D(A^1_{0.25}, B^1_{0.25})$. As a consequence $D^I_{\mathcal{C}}(A,B)^1_{0.25} \not\subseteq D(A^1_{0.25}, B^1_{0.25})$.

Neither does it hold in general that $D^I_{\mathcal{C}}(A,B)^{\alpha_2}_{\alpha_1} \supseteq D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1})$. Take for example $[\alpha_1, \alpha_2] = [1/4, 1/2]$, $\mathcal{C}(r,s) = [r_1 \cdot s_1, r_2 \cdot s_2]$ for all $r, s \in L^I$, $A(x) = [0.3, 0.6]$ for all $x \in [0,1]$, $A(x) = 0_{L^I}$ for all $x \in \mathbb{R} \backslash [0,1]$, $B(x) = [0.4, 0.7]$ for all $x \in [0,1]$ and $B(x) = 0_{L^I}$ for all $x \in \mathbb{R} \backslash [0,1])$.

Then on the one hand $0 \in D(A^{0.5}_{0.25}, B^{0.5}_{0.25}) = D(d_A, d_B) = [-1, 1]$.

On the other hand however:

$$D^I_{\mathcal{C}}(A,B)(0) = \sup_{x \in T_0(d_B) \cap d_A} \mathcal{C}(B(x), A(x)) = \sup_{x \in [0,1]} [0.3 \cdot 0.4, 0.6 \cdot 0.7] = [0.12, 0.42] \not\geq_{L^I} [0.25, 0.5],$$

or thus $0 \notin D^I_{\mathcal{C}}(A,B)^{0.5}_{0.25}$. As a consequence $D^I_{\mathcal{C}}(A,B)^{0.5}_{0.25} \not\supseteq D(A^{0.5}_{0.25}, B^{0.5}_{0.25})$.

♢

For the semi-norm $\mathcal{C} = \mathcal{C}_{\min}$, we have the following partial result.

Proposition 1. *Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, then it holds for all $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$ that:*

$$D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}_{\alpha_1} \supseteq D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1}).$$

Proof. Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, and let $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$. Then we have:

$$\begin{aligned} & y \in D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1}) \\ \Leftrightarrow\ & T_y(B^{\alpha_2}_{\alpha_1}) \cap A^{\alpha_2}_{\alpha_1} \neq \emptyset \\ \Leftrightarrow\ & (\exists x \in T_y(d_B) \cap d_A)(x \in T_y(B^{\alpha_2}_{\alpha_1}) \text{ and } x \in A^{\alpha_2}_{\alpha_1}) \\ \Leftrightarrow\ & (\exists x \in T_y(d_B) \cap d_A) \\ & (B(x-y) \geq_{L^I} [\alpha_1, \alpha_2] \text{ and } A(x) \geq_{L^I} [\alpha_1, \alpha_2]) \\ \Leftrightarrow\ & (\exists x \in T_y(d_B) \cap d_A)([\min(B_1(x-y), A_1(x)), \\ & \min(B_2(x-y), A_2(x))] \geq_{L^I} [\alpha_1, \alpha_2]) \\ \Leftrightarrow\ & (\exists x \in T_y(d_B) \cap d_A) \\ & (\mathcal{C}_{\min}(B(x-y), A(x)) \geq_{L^I} [\alpha_1, \alpha_2]) \\ \Rightarrow\ & \sup_{x \in T_y(d_B) \cap d_A} \mathcal{C}_{\min}(B(x-y), A(x)) \geq_{L^I} [\alpha_1, \alpha_2] \\ \Leftrightarrow\ & D^I_{\mathcal{C}_{\min}}(A,B)(y) \geq_{L^I} [\alpha_1, \alpha_2] \\ \Leftrightarrow\ & y \in D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}_{\alpha_1}. \end{aligned}$$

□

The reverse inclusion $D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}_{\alpha_1} \subseteq D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1})$ does not hold in general as already illustrated in Example 1.

Remark that the above decomposition property for weak $[\alpha_1, \alpha_2]$-cuts remains valid in the discrete framework.

ISBN: 978-989-95079-6-8

3.2.2 Decomposition by weak sub- and supercuts

There is also no relationship between the weak sub- and supercut $D^I_{\mathcal{C}}(A,B)_{\alpha_1}$ and $D^I_{\mathcal{C}}(A,B)^{\alpha_2}$ and the binary dilations $D(A_{\alpha_1},B_{\alpha_1})$ and $D(A^{\alpha_2},B^{\alpha_2})$ that holds in general for an arbitrary semi-norm $\mathcal{C}$. To illustrate this, we can use Example 1 again, where the weak $[\alpha_1,\alpha_2]$-cuts and the weak super- and subcuts of A and B coincide and the results thus remain valid for weak sub- and supercuts.

For the semi-norm $\mathcal{C}=\mathcal{C}_{\min}$, we have the following partial result.

Proposition 2. *Let $A,B\in\mathcal{IVFS}(\mathbb{R}^n)$, then it holds that:*

(i) $(\forall\alpha_1\in]0,1])(D^I_{\mathcal{C}_{\min}}(A,B)_{\alpha_1}\supseteq D(A_{\alpha_1},B_{\alpha_1}))$,
(ii) $(\forall\alpha_2\in]0,1])(D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}\supseteq D(A^{\alpha_2},B^{\alpha_2}))$.

Proof. Analogous to the proof of Proposition 1. □

The reverse inclusion does not hold, as already illustrated in Example 1, where replacing the weak $[\alpha_1,\alpha_2]$-cuts of A and B by the coinciding weak sub- and supercuts doesn't affect the results.

In the discrete framework, not only does Proposition 2 still remain valid, but the result now also holds for arbitrary semi-norms. Further, for $\mathcal{C}_{\min}$ also the reverse inclusion now holds.

Proposition 3. *Let $A,B\in\mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds that:*

(i) $(\forall\alpha_1\in]0,1]\cap I_r)(D^I_{\mathcal{C}_{\min}}(A,B)_{\alpha_1}=D(A_{\alpha_1},B_{\alpha_1}))$,
(ii) $(\forall\alpha_2\in]0,1]\cap I_s)(D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}=D(A^{\alpha_2},B^{\alpha_2}))$.

Proof. Analogous to the proof of Proposition 2, where now in the discrete case also

$$(\exists x\in T_y(d_B)\cap d_A)(\mathcal{C}_{\min}(B(x-y),A(x))_1\geq\alpha_1)$$
$$\Updownarrow$$
$$\sup_{x\in T_y(d_B)\cap d_A}\mathcal{C}_{\min}(B(x-y),A(x))_1\geq\alpha_1.$$

□

Proposition 4. *Let $A,B\in\mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds that:*

(i) $(\forall\alpha_1\in]0,1]\cap I_r)(D^I_{\mathcal{C}}(A,B)_{\alpha_1}\subseteq D(A_{\alpha_1},B_{\alpha_1}))$,
(ii) $(\forall\alpha_2\in]0,1]\cap I_s)(D^I_{\mathcal{C}}(A,B)^{\alpha_2}\subseteq D(A^{\alpha_2},B^{\alpha_2}))$.

Proof. Analogous to the proof of Proposition 3, but for an arbitrary semi-norm $\mathcal{C}$, so that

$$D^I_{\mathcal{C}_{\min}}(A,B)(y)_1\geq\alpha_1$$
$$\Uparrow$$
$$D^I_{\mathcal{C}}(A,B)(y)_1\geq\alpha_1$$
$$\Updownarrow$$
$$y\in D^I_{\mathcal{C}}(A,B)_{\alpha_1}.$$

□

3.3 Decomposition of the Interval-valued Fuzzy Erosion

Based on its induced negator $\mathcal{N}_{\mathcal{I}}$, a border implicator $\mathcal{I}$ can be classified as an upper or a lower border implicator as follows. A border implicator on L^I is called an upper border implicator (respectively lower border implicator) if $\mathcal{N}_{\mathcal{I}}\geq\mathcal{N}_s$ (respectively $\mathcal{N}_{\mathcal{I}}\leq\mathcal{N}_s$).

Lemma 2. *If $\mathcal{I}$ is an upper border implicator on L^I, then it holds that $\mathcal{I}\geq\mathcal{I}_{\max,\mathcal{N}_s}$, i.e.: $(\forall(x,y)\in(L^I)^2)(\mathcal{I}(x,y)\geq_{L^I}\mathcal{I}_{\max,\mathcal{N}_s}(x,y))$.*

Proof. Let $\mathcal{I}$ be an upper border implicator on L^I. For all $(x,y)\in(L^I)^2$ it holds that on the one hand $\mathcal{I}(x,y)\geq_{L^I}\mathcal{I}(1_{L^I},y)$ and $\mathcal{I}(1_{L^I},y)=y$, and on the other hand $\mathcal{I}(x,y)\geq_{L^I}\mathcal{I}(x,0_{L^I})$ and $\mathcal{I}(x,0_{L^I})\geq_{L^I}\mathcal{N}_s(x)$, from which it follows that $\mathcal{I}(x,y)\geq_{L^I}\mathcal{I}_{\max,\mathcal{N}_s}(x,y)$. □

3.3.1 Decomposition by weak $[\alpha_1,\alpha_2]$-cuts

As illustrated below, there is no relationship between the weak $[\alpha_1,\alpha_2]$-cut $E^I_{\mathcal{I}}(A,B)^{\alpha_2}_{\alpha_1}$ and the binary erosion $E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}})$ that holds in general for an arbitrary upper border implicator $\mathcal{I}$.

Example 2. Let $[\alpha_1,\alpha_2]=[0.4,0.6]$, $A(x)=[0.3,0.5]$ for all $x\in[0,1]$, $B(x)=[0.5,0.7]$ for all $x\in[0,1]$ and $A(x)=B(x)=0_{L^I}$ for all $x\in\mathbb{R}\backslash[0,1]$. Let $\mathcal{I}$ be the upper border implicator given by $\mathcal{I}_L(x,y)=[\min(1,1-x_2+y_1),\min(1,1-x_1+y_2)]$, $\forall(x,y)\in(L^I)^2$, which is a generalisation of the Łukasiewicz implicator on $[0,1]$.

On the one hand, we have that

$$E^I_{\mathcal{I}_L}(A,B)(0)=\inf_{x\in T_0(d_B)}\mathcal{I}_L(B(x),A(x))=$$
$$\inf_{x\in[0,1]}\mathcal{I}_L([0.5,0.7],[0.3,0.5])=[0.6,1],$$

or thus $0\in E^I_{\mathcal{I}_L}(A,B)^{0.6}_{0.4}$.

On the other hand, $E(A^{0.6}_{0.4},B^{\overline{0.6}}_{\overline{0.4}})=E(\emptyset,[0,1])=\emptyset$ and thus $0\notin E(A^{0.6}_{0.4},B^{\overline{0.6}}_{\overline{0.4}})$, which implies that $E^I_{\mathcal{I}_L}(A,B)^{0.6}_{0.4}\not\subseteq E(A^{0.6}_{0.4},B^{\overline{0.6}}_{\overline{0.4}})$.

Neither does it hold in general that $E^I_{\mathcal{I}}(A,B)^{\alpha_2}_{\alpha_1}\not\supseteq E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}})$. Consider for example the upper border implicator $\mathcal{I}=\mathcal{I}_{\max,\mathcal{N}_s}$, $[\alpha_1,\alpha_2]=[0.3,0.4]$, $A(x)=[0.4,0.5]$ for all $x\in[0,0.5]$ and $A(x)=[0.2,0.3]$ for all $x\in]0.5,1]$, $B(x)=[0.7,0.8]$ for all $x\in[0,0.5]$ and $B(x)=[0.4,0.8]$ for all $x\in]0.5,1]$ and $A(x)=B(x)=0_{L^I}$ for all $x\in\mathbb{R}\backslash[0,1]$.

The binary erosion $E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}})$ is then equal to the set $E([0,0.5],[0,0.5])=\{0\}$.

Further, it also holds that:

$$E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)(0)=\inf_{x\in T_0(d_B)}\mathcal{I}_{\max,\mathcal{N}_s}(B(x),A(x))=$$
$$\inf(\inf_{x\in[0,0.5]}\mathcal{I}_{\max,\mathcal{N}_s}(B(x),A(x)),$$
$$\inf_{x\in]0.5,1]}\mathcal{I}_{\max,\mathcal{N}_s}(B(x),A(x)))=\inf([0.4,0.5],[0.2,0.6])=$$
$$[0.2,0.5]\not\geq_{L^I}[\alpha_1,\alpha_2].$$

We conclude that $E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2}_{\alpha_1}\not\supseteq E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}})$. ♢

For the upper border implicator $\mathcal{I}=\mathcal{I}_{\max,\mathcal{N}_s}$, we have the following partial result.

Proposition 5. *Let $A,B\in\mathcal{IVFS}(\mathbb{R}^n)$, then it holds for all $[\alpha_1,\alpha_2]\in L^I\backslash\{0_{L^I}\}$ that:*

$$E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2}_{\alpha_1}\subseteq E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}}).$$

ISBN: 978-989-95079-6-8

Proof. Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, and let $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$. It holds that:

$$
\begin{aligned}
& y \in E(A^{\alpha_2}_{\alpha_1}, B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}}) \\
\Leftrightarrow \quad & T_y(B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}}) \subseteq A^{\alpha_2}_{\alpha_1} \\
\Leftrightarrow \quad & (\forall x \in T_y(d_B)) \\
& ((B_1(x-y) > 1-\alpha_2 \text{ and } B_2(x-y) > 1-\alpha_1) \\
& \Rightarrow (A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2)) \\
\Leftrightarrow \quad & (\forall x \in T_y(d_B))((B_1(x-y) \leq 1-\alpha_2 \text{ or } B_2(x-y) \\
& \leq 1-\alpha_1) \text{ or } (A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2)) \\
\Leftrightarrow \quad & (\forall x \in T_y(d_B)) \\
& ((1-B_1(x-y) \geq \alpha_2 \text{ or } 1-B_2(x-y) \geq \alpha_1) \text{ or} \\
& (A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2)) \\
\Leftarrow \quad & (\forall x \in T_y(d_B))(\max(1-B_2(x-y), A_1(x)) \geq \alpha_1 \\
& \text{and } \max(1-B_1(x-y), A_2(x)) \geq \alpha_2) \\
\Leftrightarrow \quad & (\forall x \in T_y(d_B)) \\
& (\mathcal{I}_{\max,\mathcal{N}_s}(B(x-y), A(x)) \geq_{L^I} [\alpha_1, \alpha_2]) \\
\Leftrightarrow \quad & \inf_{x \in T_y(d_B)} \mathcal{I}_{\max,\mathcal{N}_s}(B(x-y), A(x)) \geq_{L^I} [\alpha_1, \alpha_2] \\
\Leftrightarrow \quad & E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A, B)(y) \geq_{L^I} [\alpha_1, \alpha_2] \\
\Leftrightarrow \quad & y \in E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A, B)^{\alpha_2}_{\alpha_1}
\end{aligned}
$$

□

As already illustrated in Example 2, the reverse inclusion does not hold.

Remark that Proposition 5 remains valid in the discrete framework.

3.3.2 Decomposition by weak sub- and supercuts

Proposition 6. *Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, then it holds that*

(i) $(\forall \alpha_1 \in]0,1])(E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A, B)_{\alpha_1} = E(A_{\alpha_1}, B^{\overline{1-\alpha_1}}))$,

(ii) $(\forall \alpha_2 \in]0,1])(E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A, B)^{\alpha_2} = E(A^{\alpha_2}, B_{\overline{1-\alpha_2}}))$.

Proof. Analogous to the proof of Proposition 5. Only now it also holds that

$$
\begin{gathered}
(\forall x \in T_y(d_B))(1-B_2(x-y) \geq \alpha_1 \text{ or } A_1(x) \geq \alpha_1) \\
\Updownarrow \\
(\forall x \in T_y(d_B))(\max(1-B_2(x-y), A_1(x)) \geq \alpha_1)
\end{gathered}
$$

Analogously for the decomposition by weak supercuts. □

Proposition 7. *Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$ and let $\mathcal{I}$ be an upper border implicator on L^I, then it holds that*

(i) $(\forall \alpha_1 \in]0,1])(E^I_{\mathcal{I}}(A, B)_{\alpha_1} \supseteq E(A_{\alpha_1}, B_{\overline{1-\alpha_1}}))$,

(ii) $(\forall \alpha_2 \in]0,1])(E^I_{\mathcal{I}}(A, B)^{\alpha_2} \supseteq E(A^{\alpha_2}, B_{\overline{1-\alpha_2}}))$.

Proof. The proof is completely analogous to the one from proposition 6. We only have that due to lemma 2

$$
\begin{gathered}
\inf_{x \in T_y(d_B)} \mathcal{I}_{\max,\mathcal{N}_s}(B_2(x-y), A_1(x))_1 \geq \alpha_1 \\
\Downarrow \\
\inf_{x \in T_y(d_B)} \mathcal{I}(B_2(x-y), A_1(x))_1 \geq \alpha_1
\end{gathered}
$$

only holds in one direction for an arbitrary upper border implicator $\mathcal{I}$ on L^I. Analogously for the decomposition by weak supercuts. □

The reverse inclusion does not hold in general, as illustrated in the first part of Example 2, where replacing the weak $[\alpha_1, \alpha_2]$-cuts by the coinciding weak sub- or supercuts does not affect the results.

Further, the two above properties remain valid in the discrete framework.

3.4 Decomposition of the Interval-valued Fuzzy Closing and Opening

In what follows, we will need the following lemma:

Lemma 3. *Let $A \in \mathcal{IVFS}(\mathbb{R}^n)$ and let $[\alpha_1, \alpha_2] \in L^I$, then it holds that:*

(i) $\alpha_2 \in]0, 0.5] \quad \Rightarrow \quad A^{\alpha_2} \supseteq A^{\overline{\alpha_2}} \supseteq A_{\overline{1-\alpha_2}}$,

(ii) $\alpha_1 \in]0.5, 1] \quad \Rightarrow \quad A_{\alpha_1} \subseteq A^{\overline{1-\alpha_1}}$.

Proof. As an example we prove (i). Suppose that $x \in A_{\overline{1-\alpha_2}}$ and $\alpha_2 \in]0, 0.5]$. The latter implies that $1-\alpha_2 \geq \alpha_2$. Further, since $x \in A_{\overline{1-\alpha_2}}$, we also have that $A_1(x) > 1-\alpha_2$. If we combine the above with the fact that $A_2(x) \geq A_1(x)$, then we find that $A_2(x) > \alpha_2$ and consequently also $A_2(x) \geq \alpha_2$.

(ii) follows in an analogous way. □

3.4.1 Decomposition by weak sub- and supercuts

Proposition 8. *Let $\mathcal{I}$ be an upper border implicator on L^I and let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, then it holds for all $\alpha_1 \in]0, 1]$ that*

(i) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)_{\alpha_1} \supseteq E(D(A_{\alpha_1}, B_{\alpha_1}), -B^{\overline{1-\alpha_1}})$,

(ii) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)_{\alpha_1} \supseteq D(E(A_{\alpha_1}, B^{\overline{1-\alpha_1}}), -B_{\alpha_1})$,

and for all $\alpha_2 \in]0, 1]$ that

(iii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq E(D(A^{\alpha_2}, B^{\alpha_2}), -B_{\overline{1-\alpha_2}})$,

(iv) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq D(E(A^{\alpha_2}, B_{\overline{1-\alpha_2}}), -B^{\alpha_2})$.

Proof. As an example we prove (i). Let $\mathcal{I}$ be an upper border implicator on L^I, let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$ and let $\alpha_1, \alpha_2 \in]0, 1]$. From respectively Proposition 6, Proposition 2, and because the binary erosion is increasing in its first argument, we have that:

$$
\begin{aligned}
C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)_{\alpha_1} &= E^I_{\mathcal{I}}(D^I_{\mathcal{C}_{\min}}(A, B), -B)_{\alpha_1} \\
&\supseteq E(D^I_{\mathcal{C}_{\min}}(A, B)_{\alpha_1}, -B^{\overline{1-\alpha_1}}) \\
&\supseteq E(D(A_{\alpha_1}, B_{\alpha_1}), -B^{\overline{1-\alpha_1}}).
\end{aligned}
$$

(ii), (iii) and (iv) follow in an analogous way. □

Under the restriction of $\alpha_2 \in]0, 0.5]$, the above result leads to the following relationships between the weak α_2-supercut of the interval-valued fuzzy closing and opening and the binary counterparts.

Proposition 9. *Let $\mathcal{I}$ be an upper border implicator on L^I and let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, then it holds for $\alpha_2 \in]0, 0.5]$ that*

(i) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq C(A^{\alpha_2}, B^{\alpha_2})$,

(ii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq C(A^{\alpha_2}, B_{\overline{1-\alpha_2}})$,

and

(iii) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq O(A^{\alpha_2}, B^{\alpha_2})$,

(iv) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq O(A^{\alpha_2}, B_{\overline{1-\alpha_2}})$.

Proof. As an example, we prove (i). Let $\mathcal{I}$ be an upper border implicator on L^I, let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$ and let $\alpha_2 \in]0, 0.5]$.

ISBN: 978-989-95079-6-8

From proposition 8, lemma 3 and the fact that the binary erosion is decreasing in its second argument, it follows that:

$$\begin{aligned} C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} &\supseteq E(D(A^{\alpha_2},B^{\alpha_2}),-B_{\overline{1-\alpha_2}}) \\ &\supseteq E(D(A^{\alpha_2},B^{\alpha_2}),-B^{\alpha_2}) \\ &= C(A^{\alpha_2},B^{\alpha_2}). \end{aligned}$$

(ii), (iii) and (iv) follow in an analogous way. □

The above results for weak sub- and supercuts remain valid in the discrete framework. Because of the new decomposition properties of the interval-valued fuzzy dilation in the discrete framework compared to the continuous framework (Proposition 3 and 4), also a new property holds for the discrete interval-valued fuzzy closing and opening.

Proposition 10. *Let $\mathcal{C}$ be a semi-norm on $L^I_{r,s}$ and $\mathcal{I}$ an upper border implicator on $L^I_{r,s}$ and let $A, B \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds for all $\alpha_1 \in]0,1] \cap I_r$ that*

(i) $C^I_{\mathcal{C}_{\min},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} = E(D(A_{\alpha_1},B_{\alpha_1}),-B^{\overline{1-\alpha_1}})$,
(ii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)_{\alpha_1} \supseteq E(D(A_{\alpha_1},B_{\alpha_1}),-B^{\overline{1-\alpha_1}})$,
(iii) $C^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq E(D(A_{\alpha_1},B_{\alpha_1}),-B^{\overline{1-\alpha_1}})$,
(iv) $O^I_{\mathcal{C}_{\min},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} = D(E(A_{\alpha_1},B^{\overline{1-\alpha_1}}),-B_{\alpha_1})$,
(v) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)_{\alpha_1} \supseteq D(E(A_{\alpha_1},B^{\overline{1-\alpha_1}}),-B_{\alpha_1})$,
(vi) $O^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq D(E(A_{\alpha_1},B^{\overline{1-\alpha_1}}),-B_{\alpha_1})$,

and for all $\alpha_2 \in]0,1] \cap I_s$ that

(vii) $C^I_{\mathcal{C}_{\min},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2} = E(D(A^{\alpha_2},B^{\alpha_2}),-B_{\overline{1-\alpha_2}})$,
(viii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq E(D(A^{\alpha_2},B^{\alpha_2}),-B_{\overline{1-\alpha_2}})$,
(ix) $C^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2} \subseteq E(D(A^{\alpha_2},B^{\alpha_2}),-B_{\overline{1-\alpha_2}})$,
(x) $O^I_{\mathcal{C}_{\min},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2} = D(E(A^{\alpha_2},B_{\overline{1-\alpha_2}}),-B^{\alpha_2})$,
(xi) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq D(E(A^{\alpha_2},B_{\overline{1-\alpha_2}}),-B^{\alpha_2})$,
(xii) $O^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2} \subseteq D(E(A^{\alpha_2},B_{\overline{1-\alpha_2}}),-B^{\alpha_2})$.

Proof. Follows in an analogous way as in the proof of Proposition 8. □

Under the restriction of $\alpha_1 \in]0.5,1] \cap I_r$, the above result leads to the following relationships between the weak α_1-subcut of the interval-valued fuzzy closing and opening and the binary counterparts.

Proposition 11. *Let $\mathcal{C}$ be a semi-norm on $L^I_{r,s}$ and let $A, B \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds for all $\alpha_1 \in]0.5,1] \cap I_r$ that*

(i) $C^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq C(A_{\alpha_1},B_{\alpha_1})$,
(ii) $C^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq C(A_{\alpha_1},B^{\overline{1-\alpha_1}})$,
(iii) $O^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq O(A_{\alpha_1},B_{\alpha_1})$,
(iv) $O^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq O(A_{\alpha_1},B^{\overline{1-\alpha_1}})$,

Proof. Follows in an analogous way as in the proof of Proposition 9. □

Under the restriction of $\alpha_2 \in]0,0.5] \cap I_s$, we come to the same results as in the continuous case (Proposition 9).

Proposition 12. *Let $\mathcal{I}$ be an upper border implicator on $L^I_{r,s}$ and let $A, B \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds for all $\alpha_2 \in]0,0.5] \cap I_s$ that*

(i) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq C(A^{\alpha_2},B^{\alpha_2})$,
(ii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq C(A^{\alpha_2},B_{\overline{1-\alpha_2}})$,
(iii) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq O(A^{\alpha_2},B^{\alpha_2})$,
(iv) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq O(A^{\alpha_2},B_{\overline{1-\alpha_2}})$.

Proof. Follows in an analogous way as in the proof of Proposition 9. □

4 Conclusion

In this paper we have introduced the basic interval-valued fuzzy morphological operations and we have studied their decomposition by weak $[\alpha_1,\alpha_2]$-cuts both in the general continuous case and the practical discrete case. We found out that a few properties that do not hold in the continuous case do hold in the practical discrete case. Finally, we would like to mention that similar, but not completely analogous results hold for the decomposition by strict $[\alpha_1,\alpha_2]$-cuts. Those results will be published elsewhere.

Acknowledgment

This work was financially supported by the GOA-project B/04138/01 IV 1 of Ghent University.

References

[1] J. Serra. *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.

[2] R.M. Haralick, R.S. Sternberg and X. Zhuang. Image analysis using mathematical morphology, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):532-550, 1987.

[3] B. De Baets. Fuzzy morphology: a logical approach, *Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach* (Ayyub, B.M. & Gupta, M.M., editors), Kluwer Academic Publishers, Boston, pp. 53-67, 1997.

[4] P. Sussner and M.E. Valle: Classification of fuzzy mathematical morphologies based on concepts of inclusion measure and duality, *Journal of Mathematical Imaging and Vision* , 32(2):139-159, October 2008.

[5] I. Bloch. Mathematical Morphology on Bipolar Fuzzy Sets, *Proceedings of ISMM 2007 (International Symposium on Mathematical Morphology)*, pp. 3-4, 2007.

[6] M. Nachtegael, P. Sussner, T. Mélange and E.E. Kerre. Modelling numerical and spatial uncertainty in grayscale image capture using fuzzy set theory, *Proceedings of NASTEC 2008*, 2008.

[7] L. Zadeh. Fuzzy Sets, *Information and Control*, 8:338-353, 1965.

[8] R. Sambuc. *Fonctions Φ-floues. Application à l'aide au diagnostic en pathologie thyroidienne*, Ph.D. thesis, Univ. Marseille, France, 1975.

[9] G. Deschrijver and E.E. Kerre. On the relationship between some extensions of fuzzy set theory, *Fuzzy Sets and Systems*, 133:227-235, 2003.

[10] G. Deschrijver and E.E. Kerre. On the cuts of intuitionistic fuzzy compositions, *Kuwait Journal of Science & Engineering*, 32(1):17-38, 2005.

[11] M. Nachtegael and E.E. Kerre. Connections between binary, gray-scale and fuzzy mathematical morphology, *Fuzzy Sets and Systems*, 124:73-86, 2001.

An alternative operationalization of fuzzy consideration set. Application to tourism

Luisa L. Lazzari Patricia I. Mouliá Mariano Eriz

Faculty of Economics Science, University of Buenos Aires
Buenos Aires, Argentina
Email: ilazzari@econ.uba.ar, pimoulia@hotmail.com, erizmariano@hotmail.com

Abstract—*Consideration sets provide a standard tool for the exploration of consumer choice. It is appropriate to analyze it as a fuzzy set because not all the brands of a certain category of products or services that belong to it are considered of the same importance.*

In this work we propose an alternative operationalization of the aggregate fuzzy consideration set. It is based on a new approach for linguistic framework and linguistic aggregation operators which compute with words directly, defined by Xu. The model of consideration presented assigns to each brand or service the degree of membership to the evoked set through weighed aggregate interest.

The aggregate fuzzy consideration set of tourist destinations of a given of individuals segment is obtained as illustrative example of the proposed approach.

Keywords— aggregation operators, computing with words, fuzzy consideration set, linguistic model, tourist destinations.

1 Introduction

Several variables take place in the study of the consumer's behaviour, which is a complex activity. To face it in a more suitable manner, flexible models are appropriate; these will consider the uncertainty implied by its treatment, particularly those using the theory of the fuzzy sets. This theory provides an important background for the representation of preferences and the uncertainty contained in decision making. This approach can be applied to the analysis of the consideration set or evoked set.

Though the individual has a certain number of possibilities to choose among, it is possible that he does not take into account all the available alternatives. These alternatives considered during the decision making form what is known as the *consideration set* or *evoked set*.

The consideration set is a subset of all alternatives available for the consumer in a category of products or services. It is made up by his preferred brands, with which he is acquainted, those he remembers and those he positively evaluates for purchasing and consumption [1, 2].

Howard [3] was the first to introduce the term evoked set in the literature of marketing to talk collectively about the set of brands that the consumer considers in his process of purchases decision. Several authors elaborated this concept later [4, 5, 6].

Consideration sets have been the subject of much research since 1986. The level of interest in this study can be attributed to several factors [7]. Among others, a consideration set represents a task-simplifying heuristic that consumers use to cope with complex choice problems [8]; models that ignore consideration set information may have biased parameter estimates [9] and when a brand enters the consideration set of a consumer, the chances that the consumer will choose that brand increase even if it is not the most preferred. Its exclusion prevents the selection of the brand even if it is likely to be preferred [10].

Lilien *et al.* [9] states that studying the composition of the consideration set is important for two reasons: the first one because it will help the company that does not belong to the set to know its own lacks and the characteristics of those companies or brands that belong to it; the second because it is an important part of the general study of the consumption processes. Some research have shown that to include the consideration set in the stages of decision model allows to obtain better prognoses and a more suitable handling of diagnoses [11].

Fotheringham [12] suggests that the consideration set is not binary, but it might be treated as a fuzzy set, maybe because the consideration is not a discreet process of the consumers or because the investigators are not capable of measuring it. He expresses that not all the brands taken into account in the phases of evaluation and election are given the same importance.

In previous works [13, 14] we have developed a linguistic model to obtain the membership function of the aggregate fuzzy consideration set, for a category of products or services, by means of the usual linguistic framework and linguistic aggregation operators based on the Extension Principle [15, 16, 17, 18].

In this paper an alternative linguistic model is proposed, to obtain the aggregate fuzzy consideration set which is based on a new approach for linguistic framework and on the linguistic weighted arithmetic averaging (LWAA) operator, which enables to compute with words directly, defined by Xu [18, 19, 20, 21].

Then, this methodology is applied to obtain the aggregate fuzzy consideration set of tourist destinations (TDs) of the young, with the data obtained in a survey carried out on students of the University of Buenos Aires.

The cardinality of the evoked set and its support is obtained, in order to find its magnitudes. Besides, a measure of fuzziness is calculated to know the entropy that affects the fuzzy consideration set.

The paper is organized as follows. In Section 2 we introduce a linguistic framework presented by Xu [18, 20] to define some linguistic aggregation operators and the LWAA operator definition. In Section 3 we give an alternative operationalization of the aggregate fuzzy consideration set.

ISBN: 978-989-95079-6-8

In Section 4 the aggregate fuzzy consideration set of TDs is obtained and in Section 5 some concluding remarks are pointed out.

2 Fuzzy linguistic approach

The fuzzy linguistic approach is an appropriate technique to deal with qualitative aspects of problems [22].

Following Xu [18, 19, 20, 21] we consider a finite and totally ordered label set $S=\{s_\alpha/\alpha=-t,...,-1,0,1,...,t\}$, which cardinality value is odd, and t is a positive integer. Each term s_α represents a possible value for a linguistic variable [18, 19, 20, 21] and it must have the following characteristics:

i) $s_\alpha > s_\beta$ iff $\alpha > \beta$.

ii) There is the negation operator: $\text{neg}(s_\alpha)=s_{-\alpha}$; mainly $\text{neg}(s_0)=s_0$.

The mid linguistic label s_0 represents an assessment of "indifference" and the rest of labels are defined around it symmetrically.

To preserve all the given information, Xu [18, 19] extended the discrete linguistic label set S to a continuous linguistic label set $\overline{S}=\{s_\alpha/\alpha\in[-q,q]\}$, where q $(q>t)$ is a sufficiently large positive integer.

If $s_\alpha \in S$, then s_α is called an original linguistic label, otherwise, s_α is called a virtual linguistic label. Generally, the decision maker uses the original linguistic terms to evaluate attributes and alternatives, and the virtual linguistic labels can only appear in calculations [20, 21].

Considering any two linguistic terms $s_\alpha, s_\beta \in \overline{S}$, and $\lambda\in[0,1]$, Xu [19] introduces two operational laws of linguistic variables as follows:

- $s_\alpha \oplus s_\beta = s_\beta \oplus s_\alpha = s_{\alpha+\beta}$. (1)
- $\lambda . s_\alpha = s_{\lambda\alpha}$. (2)

Based on (1) and (2) Xu [18, 19, 21] developed various linguistic aggregation operators, which compute with words directly.

In this section we present the linguistic weighted arithmetic averaging (LWAA) operator due to the fact that it will be the linguistic operator used in our aggregate fuzzy consideration sets model.

Definition [18]: Let $LWAA: \overline{S}^n \to \overline{S}$. If

$$LWAA_w(s_{\alpha_1}, s_{\alpha_2}, ..., s_{\alpha_n}) = w_1 s_{\alpha_1} \oplus w_2 s_{\alpha_2} \oplus ... \oplus w_n s_{\alpha_n} = s_{\overline{\alpha}} \quad (3)$$

Where $\overline{\alpha}=\sum_{j=1}^{n} w_j.\alpha_j$, $w=(w_1, w_2, ..., w_n)$ is the weight vector of the linguistic label s_{α_j}, $j=1,...,n$, $w_j\in[0,1]$ and $\sum_{j=1}^{n} w_j = 1$, then *LWAA* is called the linguistic weighted arithmetic averaging operator.

The fundamental aspect of the *LWAA* operator is that it computes the linguistic labels taking into account the importance of the information [18].

3 The fuzzy consideration set

3.1 Preliminaries

When forming the evoked set of a certain category of products or services, the consumers do not show the same interest in their purchase, due to the variety of attributes these products or services have and to other factors that in most cases are subjective, thus, it is appropriate to analyze it as a fuzzy set [7, 9, 12, 13, 14].

When developing its model of consideration, Fotheringham [12] states that the evoked set can be considered fuzzy because uncertainty about its composition exists. Although the author affirms that the consideration sets are fuzzy, his formulation does not derive from the theory of the fuzzy sets. Besides, it does not explain which the fuzziness source is or how it obtains the membership function.

Wu and Rangaswamy [7] developed a model of consideration and election, which they call fuzzy. His work suggests that a fuzzy approach is useful to represent the fuzziness of the consideration set, their model is derived from probability theory and they do not explain how to obtain the membership function.

The model of consideration presented in this paper assigns to each brand or service the degree of membership to the evoked set through weighed aggregate interest, calculated by means of the *LWAA* operator. To obtain the value of the membership function of a brand or service, the aggregate degree of interest is weighed in agreement with the quantity of individuals who chose the above mentioned brand or service.

3.2 Building aggregate fuzzy consideration sets

Surveys to potential clients interested in the purchase of a type of goods or services are made to obtain the aggregate fuzzy consideration set of a category of products or services of a given of individuals segment. In the questionnaire we ask about:

- The names of the well-known brands.
- The names of the brands that they would consider buying.
- The degree of interest with which these are considered for their purchase.

For the design of questionnaires for the surveys a set of linguistic labels is chosen which fulfils the characteristics given in 2., so that the consulted individuals express their degree of interest in the purchase of the goods or the services.

First of all, the referential set of the evoked set (E) is obtained with the brands that appear in the survey. It is named set of well-known brands or awareness set.

To obtain the value of the membership function of each element the aggregate fuzzy consideration set, the following steps are involved:

Step 1. *The aggregate degree of interest.* If m is the quantity of proposed brands in the surveys and n is the cardinality of the label set S, the aggregate degree of interest

ISBN: 978-989-95079-6-8

(q_i) of each brand or service (c_i) considered is obtained by means of the application of (1).

$$q_i = LWAA_{c_i}\left(s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_n}\right) = s_{\overline{\alpha}_i} \quad i = 1,\ldots, m \qquad (4)$$

Where $\overline{\alpha}_i = \sum_{j=1}^{n} w_j \alpha_j$ and the weighting vector of the linguistic level s_{α_j} is $w = (w_1, w_2, \ldots, w_n)$. If r_{ij} is the quantity of individuals that selected the brand or service c_i with the level s_{α_j} and v_i is the quantity of all individuals that selected this brand, then:

$$w_j = \frac{r_{ij}}{v_i}, j = 1,\ldots, n;\ w_j \in [0,1],\ \sum_{j=1}^{n} w_j = 1 \qquad (5)$$

Step 2. *The value of the membership function, without weighing:* We define it on the basis of the aggregate degree of interest, displacing the scale of order $\frac{n-1}{2}$, later standardized.

$$\kappa_i = \frac{2\overline{\alpha}_i + n - 1}{2(n-1)}, \quad i = 1,\ldots, m,\ 0 \le \kappa_i \le 1 \qquad (6)$$

Step 3. *The weighting vector*: $p = (p_1, \ldots, p_m)$ is obtained to reflect the number of times each brand is chosen in the survey. Specifically, weights p_i are calculated as follows:

$$p_i = \frac{F_i}{\max F_i};\ 0 < p_i \le 1; \forall i = 1,\ldots, m \qquad (7)$$

where F_i is the frequency of occurrence for each brand.

Step 4. *The aggregate fuzzy consideration set*: The value of the membership function for each brand or service (c_i) is calculated as follows:

$$\mu_C(c_i) = p_i . \kappa_i \qquad (8)$$

which represent the intentions of purchase of the individuals selecting the brand or service c_i.

Finally, the aggregate fuzzy consideration set is:

$$C: E \to [0,1] / \mu_C(c_i) = p_i . \kappa_i \qquad (9)$$

4 Application

In this section, the aggregate fuzzy consideration set of TDs of a given of individuals segment is obtained as an illustrative example of the proposed approach.

The consideration set of TDs is a subset of the set of all consumers' available alternatives and it is constituted by those TDs with which the consumer is acquainted, which he remembers or which he considers would be interesting to visit, and which were evaluated positively to spend his holidays [23, 24].

As not all the TDs taken into account in the phases of evaluation and election are considered to have the same importance, it is possible to state that they belong to the evoked set with a different degree [13, 14]. For this reason we propose to treat the consideration set of TDs as a fuzzy set on the basis of the set of the places known by the consumers.

Since mid 1900s, tourism has turned into one of the economic activities of major growth in the world. The tourist activity mobilises public and private funds, affecting the economic and social conditions of the population, and it can be an essential tool to achieve a sustainable development as it is a generator of employment and currency. This justifies the systematical study of the tourist activity in a region or country.

Young tourism is an increasing world trend. The characteristic of this segment is the performing of trips, different from those of the traditional market. In Argentina, the young tourism represents almost 20 % of the tourist entire market.

Then, the methodology presented in Section 3.2 is applied to obtain the aggregate fuzzy consideration set of TDs of young students, with the data obtained in a survey.

A random closed survey to 940 students of the University of Buenos Aires (UBA) between 19 and 29 years of age was done in September 2008. We obtained information about the tourist places that students would like to visit in the next summer vacation (January, February, 2009) and the degree of interest to visit particular locations.

A linguistic term set of five values was used, and every student expressed his degree of interest in each TD considered by one of this label of the set S.

$S = \{ s_{-2} = \text{very low (VL)},\ s_{-1} = \text{low (L)},\ s_0 = \text{middle (M)},\ s_1 = \text{high (H)},\ s_2 = \text{very high (VH)} \}$

The data obtained in the survey were processed. The TDs of the support of aggregate fuzzy consideration set are:

$c_1 = \text{Bariloche}$
$c_2 = \text{Cataratas del Iguazú}$
$c_3 = \text{Córdoba}$
$c_4 = \text{El Calafate}$
$c_5 = \text{Jujuy}$
$c_6 = \text{Mar del Plata}$
$c_7 = \text{Mendoza}$
$c_8 = \text{Pinamar}$
$c_9 = \text{Puerto Madryn}$
$c_{10} = \text{Salta}$
$c_{11} = \text{San Martín de los Andes}$
$c_{12} = \text{Ushuaia}$

As an example of the calculation of the value of the membership function, we will obtain it for $c_1 = \text{Bariloche}$.

Step 1: We assume $s_{\alpha_1} = s_{-2}$; $s_{\alpha_2} = s_{-1}$; $s_{\alpha_3} = s_0$; $s_{\alpha_4} = s_1$; $s_{\alpha_5} = s_2$.

By (4) and (5), we have $w = (\frac{2}{273}, \frac{9}{273}, \frac{59}{273}, \frac{103}{273}, \frac{100}{273})$, thus,

$$\overline{\alpha}_1 = \frac{2}{273} \times (-2) \oplus \frac{9}{273} \times (-1) \oplus \frac{59}{273} \times 0 \oplus \frac{103}{273} \times 1 \oplus \frac{100}{273} \times 2$$

therefore, $\overline{\alpha}_1 = 1.06$ and $q_1 = s_{1.06}$.

Step 2: By means of (6) $\kappa_1 = \frac{2(1.06) + 4}{8} = 0.77$.

Step 3: Using (7) $p_1 = \frac{273}{306} = 0.89$.

Step 4: Finally, $\mu(c_1) = 0.69$.

ISBN: 978-989-95079-6-8

The following results are contained in Table 1.

In column 1: The TDs of the support of aggregate fuzzy consideration set.

In column 2: The frequency of each TD (F_i).

In column 3: The aggregate degree of interest of each TD (q_i).

In column 4: The value of membership function of each TD, without weighing (κ_i).

In column 5: The weight of each TD (p_i).

In column 6: The value of membership function of each TD ($\mu(c_i)$).

Table 1: Results of 1st survey - UBA 2008

TDs	F_i	q_i	κ_i	p_i	$\mu(c_i)$
c_1	273	$s_{1.06}$	0.77	0.89	0.69
c_2	306	$s_{0.97}$	0.74	1	0.74
c_3	231	$s_{0.78}$	0.70	0.75	0.53
c_4	152	$s_{1.43}$	0.86	0.50	0.43
c_5	161	$s_{1.07}$	0.77	0.53	0.41
c_6	182	$s_{0.45}$	0.61	0.59	0.36
c_7	247	$s_{1.02}$	0.75	0.81	0.61
c_8	100	$s_{0.67}$	0.67	0.33	0.22
c_9	91	$s_{1.09}$	0.77	0.30	0.23
c_{10}	269	$s_{1.12}$	0.78	0.88	0.69
c_{11}	91	$s_{1.24}$	0.81	0.30	0.24
c_{12}	162	$s_{1.29}$	0.82	0.53	0.43

It is possible to express the linguistic aggregate degree of interest (LADI) of each TD using a label of set S. Approximating the subindex of the virtual label to an integer value by means of the usual round operation, i.e., round $(\overline{\alpha}_i)$. The results can be observed in Table 2. This information is useful to the decision maker.

Table 2: Linguistic aggregate degree of interest

TD	q_i	$s_\alpha \in S$	LADI
c_1	$\alpha_{1.06}$	α_1	High
c_2	$\alpha_{0.97}$	α_1	High
c_3	$\alpha_{0.78}$	α_1	High
c_4	$\alpha_{1.43}$	α_1	High
c_5	$\alpha_{1.07}$	α_1	High
c_6	$\alpha_{0.45}$	α_0	Middle
c_7	$\alpha_{1.02}$	α_1	High
c_8	$\alpha_{0.67}$	α_1	High
c_9	$\alpha_{1.09}$	α_1	High
c_{10}	$\alpha_{1.12}$	α_1	High
c_{11}	$\alpha_{1.24}$	α_1	High
c_{12}	$\alpha_{1.29}$	α_1	High

The destinations that belong to the aggregate fuzzy consideration set have obtained high levels of aggregate interest, except for Mar del Plata which obtained middle level. The predominant destinations are "Cataratas del Iguazú", "Salta" and "Bariloche", with the biggest values of membership (Table 1).

To have an idea of the magnitude of the obtained sets, their cardinality has been calculated.

Awareness set: $|E| = 62$.

Support of aggregate fuzzy consideration set: $|S(C)| = 12$.

Aggregate fuzzy consideration set: $\|C\| = 5.22$.

To determine the degree of uncertainty contained in the information provided by the aggregate fuzzy consideration set, we obtained the Yager's measure of fuzziness [25].

$$e_Y(C) = 1 - \frac{\left[\sum_{i=1}^{n} \left|\mu_C(x_i) - \mu_{\overline{C}}(x_i)\right|^p\right]^{\frac{1}{p}}}{|\mathrm{supp}(C)|}, \; p \in N - \{0\} \qquad (10)$$

$0 \le e_Y(C) \le 1$.

If $p = 1$, then $e_Y(C) = 0.68$.

We can states that the fuzziness of the obtained set is moderate. This value indicates a measure of the information quality about the uncertainty of the group's consideration.

The utility of the entropy has its roots in the possibility of comparison as well as in the study of fuzziness' evolution existing in a system.

Empirical available evidence of similar studies shows that the fuzziness of the aggregated fuzzy consideration set is moderate in the first instance of its evaluation and diminishes its value as you approach the moment to make the decision about your holidays.

Reference [14] shows that if $p = 1$, then (10) is equal to Kaufmann's linear index of fuzziness.

5 Conclusions

This paper offers a new approach to build aggregate fuzzy consideration set.

We verified that it enables to find more easily results similar to those obtained with the method that uses linguistic operators based on the Extension Principle.

The operator used in the model presented compute with words directly and weighs the linguistic argument itself, instead of the ordered position. It allows for a continuous representation of the linguistic information on its domain, and thus, they can represent any counting of information obtained in an aggregation process without any loss of information.

In this work we have presented preliminary results of a research on young tourism begun in 2008. In this project we will analyze young people behaviour in choosing a TD in Argentina. National University population will be the aim of this study, since attending students have heterogeneous characteristics.

With the information obtained through longitudinal exploratory studies, we will determine and analyze the aggregate fuzzy consideration sets of tourist destinations in Argentina at local and regional level, in winter and summer. We will study the dynamics and the fuzziness of the aggregate sets and the attributes the young have in mind when choosing possible destinations to spend their vacations. Obtained results will be compared considering tourist season and geographic areas.

ISBN: 978-989-95079-6-8

Finally, after each vacation period, the same group of students will be asked about which TD was visited, to check the forecast of the model.

This study will allow us to obtain relevant information, to make a diagnosis of the situation of the young tourism in the considered regions and to plan actions tending to increase this activity.

At this stage of the research we obtained and analyzed the aggregated awareness sets and consideration set of tourist destinations of Argentina for the young people in 2009 summer vacation.

The results allow to infer that the hypothesis referred to consideration set's fuzziness is valid as the consulted young people stated that not all the TDs considered to spend their next vacation have aroused the same interest of visit.

We observed that the destinations considered are traditional and well-known. They are distributed all around the whole country and present a variety of sceneries and attractions, such as the sea, the mountains, lakes, glaciers, rivers, waterfalls and cities with great cultural activity.

The favorites TDs are Bariloche, Mar del Plata, Iguazú Waterfalls and Córdoba, because they present the largest degree of consideration.

Acknowledgement

This work has been supported by the Argentinian Agencia Nacional de Promoción Científica, Tecnológica y de Innovación (PICT 383) and by the University of Buenos Aires (UBACyT E018).

References

[1] D.L. Loudon and A.J. Della Bitta, *Consumer Behavior: Concepts and Applications*. New Jersey: McGraw-Hill, Inc., 2001.

[2] L.G. Schiffman and L.L. Kanuk, *Consumer Behavior*. New Jersey: Prentice Hall Inc., 2000.

[3] J.A. Howard, *Marketing Management: Analysis and Planning*. Illinois: Richard D. Irwin Co., 1963.

[4] J.E. Brisoux and M. La Roche, Evoked set formation and composition: An empirical investigation under a routinized response behavior situation. *Advances in Consumer Research*, 8: 357-361, 1980.

[5] J.A. Howard and J.N. Sheth, *The Theory of Buyer Behavior*. New York: Wiley, 1969.

[6] C.L. Narayana and R.J. Markin, Consumer behavior and product performance: An alternative conceptualization. *Journal of Marketing*, 39: 1-6, 1975.

[7] J. Wu and A. Rangaswamy, A fuzzy set model search and consideration with an application to an online market. *Marketing Science*, 22: 411-434, 2003.

[8] B.J. Bronnenberg and W.R. Vanhonacker, Limited choice sets, local price response and implied measures of price competition. *Journal of Marketing Research*, 33: 163-173, 1996.

[9] G.L. Lilien *et. al.*, *Marketing Models*. New Jersey: Prentice Hall International, Inc., 1992.

[10] R.L. Andrews and T.C. Srinivasan, Studying consideration effects in empirical choice models using scanner panel data *Journal of Marketing Research*, 32: 30-41, 1995.

[11] D.H. Gensch and J.A. Svestka, A maximum likelihood hierarchical disaggregate model for predicting choice of individuals. *Journal of Mathematical Psychology*, 28: 160-178, 1984.

[12] A.S. Fotheringham, Consumer store choice and choice set definition. *Marketing Science*, 7(3): 299-310, 1988.

[13] M. Eriz *et al.* Consideration set in tourist destinations as a fuzzy set. *Proceedings of International Conference on Modelling & Simulation* (ICMS'04), Valladolid, 2004.

[14] L.L. Lazzari, *El comportamiento del consumidor desde una perspectiva fuzzy*. Tesis doctoral. Universidad de Valladolid, Valladolid, 2006.

[15] M. Delgado *et al.* Aggregation of linguistic information based on a symbolic approach. In: L.A. Zadeh and J. Kacprzyk (edits.). *Computing with Words in Information/Intelligent Systems* 1. Heidelberg: Physica-Verlag, 1999.

[16] F. Herrera and E. Herrera-Viedma, Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, 115: 67-82, 2000.

[17] E. Herrera-Viedma, *et al.*, Evaluation the information quality of web sites: A methodology based on fuzzy computing with words. *Journal of American Society for Information Science and Technology*, 57(4): 538-549, 2006.

[18] Z. Xu, Linguistic aggregation operators: An overview, in: H. Bustince *et al.* (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*. Berlin: Springer-Verlag, 2008.

[19] Z. Xu, EOWA and EOWG operators for aggregating linguistic labels based on linguistic preference relations. *International Journal of Uncertainty Fuzziness and Knowledge-Based- Systems*, 12: 791-810, 2004.

[20] Z. Xu, Deviation measures of linguistic preference relations in group decision making. *Omega*, 33: 249-254, 2005.

[21] Z. Xu, A note on linguistic hybrid arithmetic averaging operator in multiple attribute group decision making with linguistic information. *Group Decision and Negotiation*, 15: 593-604, 2006.

[22] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. Part I, *Information Sciences*, 8 : 199-249. Part II, *Information Sciences*, 8: 301-357. Part III, *Information Sciences*, 9: 43-80, 1975.

[23] J.L. Crompton and P.K. Ankomah, Choice set propositions in destination decisions. *Annals of Tourism Research*, 20: 461-476, 1993.

[24] H.R. Seddighi and A.L. Theocharous, A model of tourism destination choice: a theoretical and empirical analysis. *Tourism Management*, 23: 475-487, 2002.

[25] R.R. Yager, On the measure of fuzziness and negation. Part I: Membership in the unit interval. *International Journal of General Systems*, 5: 221-229, 1979.

ISBN: 978-989-95079-6-8

Ant Colony Optimization for Distributed Routing Path Optimization in Optical Burst-Switched Networks

João Pedro[1, 2], João Pires[2], and Joao Paulo Carvalho[3]

1 Nokia Siemens Networks Portugal S.A., R. Irmãos Siemens 1, 2720-093 Amadora, Portugal
2 Instituto de Telecomunicações, Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
3 INESC-id, Instituto Superior Técnico, R. Alves Redol 9, 1000-029 Lisboa, Portugal
Email: joao.pedro@nsn.com, jpires@lx.it.pt, joao.carvalho@inesc-id.pt

Abstract—*This work describes a distributed framework for routing path optimization in Optical Burst-Switched (OBS) networks that loosely mimics the foraging behaviour of ants observed in nature, which in the past has originated the Ant Colony Optimization (ACO) metaheuristic. The framework consists of additional data structures stored at the nodes and of special control packets that traverse the network, estimate the goodness of their paths and update accordingly the data structures of the nodes. The performance of the framework, which has been implemented on an event-driven OBS network simulator, is evaluated on several network topologies and compared with that obtained with centralized routing path optimization. The simulation results show that the distributed routing path selection framework significantly improves the performance of OBS networks by reducing data loss and that it attains a performance comparable to that of the centralized algorithm. The results also suggest that the framework is robust to changes in its parameters.*

Keywords—Ant colony optimization, optical burst switching, routing path optimization, swarm intelligence.

1 Introduction

The foraging behaviour of some ant species has inspired the ant colony optimization metaheuristic [1], which has become one of the most successful swarm intelligence methods for problem solving [2]. Ant colonies are distributed systems that take advantage of their structured social organization to accomplish complex tasks while relying only on very simple individuals. This property, where a colony is able to perform tasks whose complexity far exceeds that of an individual, has made ant colonies a source of inspiration to develop artificial distributed systems and algorithms [3].

It has been shown that ACO algorithms can efficiently solve routing problems in some types of telecommunication networks, namely in circuit-switched [4] and packet-switched [5] networks. Recently, ACO algorithms were also devised for the routing and wavelength assignment problem that arises in Wavelength Division Multiplexing (WDM) optical networks [6], [7].

Circuit-switched WDM optical networks are worldwide used as backbone and metropolitan networks, due to their high capacity and low transmission losses. However, their coarse granularity limits their effectiveness to support bursty Internet Protocol (IP) packet traffic. Research efforts have been directed to finer-grained optical networking paradigms, such as optical burst switching [8]. At the edge nodes of OBS networks, multiple IP packets are assembled into bursts of data, which are then transmitted towards their egress nodes. A one-way resource reservation mechanism without acknowledgement of successful reservation in the burst path is used to allocate wavelength channels to the upcoming data bursts. Hence, there is the possibility that two or more bursts contend for the same wavelength channel at the core nodes, which is further augmented by the lack of efficient optical buffers (limiting the capability to delay data bursts). Thus, contention is a major issue in OBS networks and minimizing it is of paramount importance. One approach to mitigate contention consists of optimizing the routing paths of data bursts as to balance the traffic load across the network links.

This paper describes a distributed framework for routing path optimization in OBS networks inspired on ACO and assesses its performance using network simulation. The simulation results show that this framework can efficiently reduce contention, thus improving network performance, without requiring a fine tuning of its main parameters.

The remainder of the paper is organized as follows. Section 2 describes the ant foraging behaviour observed in nature and its utilization for network routing optimization, whereas section 3 introduces the OBS network architecture and the relevance of optimizing the paths used to route data bursts. The proposed distributed routing path optimization framework for OBS networks, embedding collective intelligence and problem solving capabilities observed in colonies of ants is described in section 4. The performance of the proposed framework is evaluated through network simulation and compared to that obtained with shortest path routing and a centralized routing path optimization algorithm in section 5. Finally, section 6 presents the conclusions.

2 Ant colony optimization and its applications

2.1 Ants in Nature

Individual ants are very simple insects whose behaviour appears to have a large random component. Despite their individual simplicity, when acting collectively they are able to execute complicated tasks with significant consistency and reliability. For instance, some ant species can collectively cooperate in regulating the nest temperature, raiding particular areas for food, building and protecting the nest, carrying large items, preferentially exploiting the richest available food source and, more related to this work, finding the shortest routes from the nest to a food source [4]. These

ISBN: 978-989-95079-6-8

relatively complex tasks seem to emerge from interactions between large numbers of ants and their environment. In some cases, individual ants coordinate their activities using a principle known as *stigmergy* [1], whose main characteristics are: (i) indirect and non-symbolic form of communication mediated by the environment, that is, ants exchange information by modifying their environment; (ii) stigmergic information is local, which means that it can only be accessed by ants that visit the immediate vicinity of the place where it was released.

The communication between individual ants is based on the use of specific chemicals, called pheromones. Particularly important for the social life of some ant species is the *trail pheromone*, which is a specific type of pheromone they use for marking paths on the ground. While walking between food sources and the nest, ants deposit pheromones on the ground, forming a pheromone trail. Subsequent ants sense the pheromone and tend to choose, probabilistically, paths with a stronger pheromone concentration, further reinforcing it. After some time, ants converge to use a single shortest path. A compendium of seminal works on the pheromone trail-laying and trail-following behaviour of some ant species can be found in [3].

2.2 Ant Colony Optimization for Network Routing

The basic idea underlying ant colony-based algorithms and systems is to use a form of *artificial stigmergy* to coordinate the activities of *artificial agents*. The majority of ACO applications have been for solving $\mathcal{NP}$-hard problems, for which the best known algorithms that guarantee an optimal solution have an exponential time worst case complexity. In some of these problems, ACO algorithms have been shown to determine high-quality, although not guaranteed to be optimal, solutions in relatively short time.

The routing problems in telecommunication networks are particularly interesting for ACO, because the intrinsically distributed nature of ACO can fit well with them. However, these problems pose additional challenges, because the routing algorithm should also be able to react rapidly to changing network and traffic conditions and to maintain exploration capabilities so that it can effectively evaluate alternative paths which, in view of the problem dynamics, can become the best ones at some point in time [1].

The different ACO algorithms for network routing have in common the fact that pheromone tables are stored at the network nodes and artificial agents, typically in the form of control packets, are employed to collect node/link state information and update the pheromone tables. ACO was first applied to network routing in circuit-switched networks, like the Plain Old Telephone Service (POTS), in [3]. The main purpose of Ant-Based Control (ABC) is to relieve congestion on the network by preferably routing calls (circuits) via parts of the network with larger spare capacity, consequently reducing call blocking probability. Later, ACO was applied to routing in packet-switched networks, such as IP networks, in [5]. These networks use store-and-forward nodes, which contain buffers to store the incoming packets and delay them until there is free capacity on the preferred output link. With AntNet [5], an artificial ant is launched at regular intervals of time from a network node towards another node to discover a low-cost path between the nodes and to investigate the load status of the network. The selection of the next node by an ant is made by assigning to the neighbouring nodes a probability of being selected that depends not only on the correspondent pheromone value but also on the occupancy of the queue associated to the link between the current node and the neighbouring node. Hence, the chances of selecting a neighbouring node increase with its pheromone value but decrease with the queue waiting time associated to it.

ACO has also found application on routing in optical WDM networks, both in Optical Circuit-Switched (OCS) [6], [7] and Optical Packet-Switched (OPS) networks [9]. AntNet was adapted in [9] for load balancing in OPS networks. The nodes of these networks use Fibre Delay Line (FDL) buffers, which have a very limited contention resolution capability, when compared to electronic buffers. The proposal modifies the solution construction phase of the original AntNet implementation [5] using a suitable expression for computing the buffer occupancy, since FDL buffers behave differently from electronic buffers.

Despite the recent research activity aiming to use ACO for routing in optical WDM networks, both in OCS and OPS, to the best of the author's knowledge ACO has not yet been adapted to the case of OBS networks.

3 Optical burst-switched networks

In current optical transport networks, circuits are established between nodes during long periods of time (weeks, months) using wavelength channels with a capacity of several Gbit/s each. However, the dominant IP packet traffic carried over these networks significantly fluctuates over relatively small time scales. Hence, current circuit-switched optical networks are inefficient to support this type of traffic due to both their coarse granularity and slow circuit setup time. OPS networks would provide the best adaptation between the IP traffic and the optical network resources used, but they cannot be implemented with state-of-the-art technology, namely due to the already referred lack of efficient optical buffers. OBS networks have attracted considerable interest because their switching granularity and optical technology complexity are in between those of OCS and OPS [10].

3.1 Network Architecture

At the ingress edge nodes of OBS networks, IP packets are assembled into large chunks of data, called bursts. Each data burst is transmitted using a wavelength channel and must be switched at the core nodes until it reaches the egress edge node, where it is disassembled and its packets delivered to the local IP network. Resource reservation for each burst is made using a control message, known as Burst Header Packet (BHP), sent ahead of the data burst on a separate wavelength channel. Each node along the routing path of the burst processes the BHP and configures its switching matrix to direct the upcoming data burst to an available wavelength channel on the appropriate output fibre link. The interested reader can find more information on OBS networks in [10].

ISBN: 978-989-95079-6-8

3.2 Routing Path Optimization

In OBS networks, multiple data bursts, overlapping in time, may contend at a node for the same wavelengths of an output fibre link. Unresolved contention leads to dropping some of the data bursts, degrading network performance. Therefore, it is important to minimize the probability that the number of overlapping data bursts directed to a fibre link exceeds the number of wavelength channels. This can be achieved by optimizing the routing paths in order to balance the traffic load offered to the network links, reducing the traffic load on the most congested links at the expense of increasing it on the less congested ones.

Routing path optimization can either be centralized, when all the paths are computed at a central point of the network using global information of the offered traffic load [11], or distributed [12]. Although centralized strategies can reduce contention more efficiently, by exploiting global knowledge of the network state, distributed strategies are often preferred because they are more scalable and robust and faster to adapt the routing paths to changes in the traffic pattern.

4 Distributed Ant-inspired Burst Routing

The framework for distributed routing path optimization in OBS networks proposed in this work is based on principles of ACO. Although it inherits some concepts from previous ACO algorithms for network routing, such as AntNet and ABC, the proposed Distributed Ant-inspired Burst Routing (DABR) framework was specifically designed for bufferless OBS networks and incorporates functions that account for the particular characteristics of these networks.

Common to all applications of ACO for network routing, the DABR framework uses artificial ants, in the form of control packets, and data structures at the network nodes to support stigmergic communication between the artificial ants. The key features of DABR are:

- Forward ants carry a fictitious data burst reservation and, at each node, the instantaneous congestion of the output fibre links is evaluated based on the number of wavelengths available to transmit the fictitious burst.
- Backward ants update the pheromone tables using the congestion information collected by the forward ants as input to a loss performance model that estimates the burst blocking probability on the path traversed by the forward ant.
- In order to mitigate path stagnation, three mechanisms are used: the routing of forward ants employs both a minimum pheromone concentration per adjacent node (noise) and pheromone-heuristic control, whereas the backward ants use privileged pheromone laying [13].
- The nodes keep separate routing and pheromone tables because an additional path integrity mechanism is used to completely avoid circular path formation.

Consider an OBS network modelled as a directed graph $G = (V, E)$, where $N = |V|$ is the number of nodes and $L = |E|$ is the number of unidirectional fibre links. Assume each fibre link supports a set of W wavelengths, $\{\lambda_1, \lambda_2, \ldots, \lambda_{W-1}, \lambda_W\}$. The following sub-sections describe the DABR framework.

4.1 Control Packets and Local Data Structures

In DABR, artificial ants are transmitted on the control channel as modified BHPs. Two types of artificial ants are used: Explorer Ant Packet (EAP) and Referee Ant Packet (RAP). The former ant is used to collect information on the resource availability along a given routing path and to update the pheromone tables, loosely mimicking the behaviour of real ants, whereas the latter one is used to avoid circular path formation. Both EAP and RAP have two modes of operation: forward (F), from the ingress to the egress node, and backward (B), in the reverse path. Two data structures are stored at each node: one pheromone table and one routing table. The pheromone table is used for routing the F-EAPs and it is updated in their return trips as B-EAPs. The routing table can only be updated by RAPs.

Unlike previous ACO routing algorithms, routing of artificial ants and data bursts in DABR is based not only on the egress node but also on the ingress node, providing an additional degree of freedom for load balancing at the expense of increasing the size of the tables. Let A_i denote the set of adjacent nodes of $i \in V$ (nodes connected to i by a fibre link) and τ_{ij}^{sd} denote the pheromone value associated to using node $j \in A_i$ to route at node i a burst generated at s and directed to d. The sum of the pheromone values related with the same ingress and egress node pair must equal one,

$$\sum_{j \in A_i} \tau_{ij}^{sd} = 1, \quad i, s, d \in V, s \neq d . \tag{1}$$

Let h_{ij} denote the number of links of the shortest path between i and j. The pheromone tables are initialized as to be slightly biased towards the shortest paths,

$$\tau_{ij}^{sd} = \begin{cases} \dfrac{1/(1+h_{jd})}{\sum_{k \in A_i} 1/(1+h_{kd})}, & j \in A_i, i \neq d, s \neq d \\ 0, \text{ otherwise} \end{cases} . \tag{2}$$

4.2 Routing of Forward Explorer Ants

Each time node s assembles a data burst directed to d, it generates with probability p_{ant}, where $0 < p_{\text{ant}} \leq 1$, an F-EAP that is routed from s to d to collect information on the resource availability of a particular path. The F-EAP carries a fictitious burst reservation request whose duration equals that of the burst that triggered the generation of the artificial ant and a list V_{EAP} with the nodes it has already visited.

At each node $i \in V$ visited by an F-EAP, the next node of its routing path is selected among the adjacent nodes not visited yet. In case $A_i \subseteq V_{\text{EAP}}$, the F-EAP is dropped to avoid looping. Otherwise, the adjacent node j is selected with a probability p_{ij}^{sd} that is the normalized sum of the pheromone value and a heuristic value η_{ij} that accounts for the resource availability on the fibre link between nodes i and j [5],

$$p_{ij}^{sd} = \begin{cases} \dfrac{\tau_{ij}^{sd}}{(1+\alpha) \sum_{\substack{k \in A_i \\ k \notin V_{\text{EAP}}}} \tau_{ik}^{sd}} + \alpha \dfrac{\eta_{ij}}{(1+\alpha) \sum_{\substack{k \in A_i \\ k \notin V_{\text{EAP}}}} \eta_{ik}}, & j \in A_i, j \notin V_{\text{EAP}} \\ 0, \quad \text{otherwise} \end{cases} . \tag{3}$$

ISBN: 978-989-95079-6-8

The pheromone value is the result of pheromone laying by previous artificial ants and represents the long-term goodness of selecting the fibre link from *i* to *j* to route the F-EAP, whereas the heuristic value corresponds to the instantaneous goodness of selecting this fibre link. In order to compute η_{ij}, the wavelength availability of the fibre link *ij* is assessed by determining if the fictitious data burst could be scheduled on each of the *W* wavelengths of the fibre link. Let $\hat{W}_{ij}$ denote the number of available wavelengths to schedule the data burst on this fibre link. The heuristic value η_{ij} is given by

$$\eta_{ij} = \hat{W}_{ij} / W . \quad (4)$$

The link state weight α determines the importance of the instantaneous resource availability of the fibre link over its long-term pheromone concentration.

The node selected *j** is added to the F-EAP, as well as the number of available wavelengths $\hat{W}_{ij^*}$. When it reaches the egress node, the artificial ant is switched to B-EAP.

4.3 Pheromone Table Update by Backward Explorer Ants

The B-EAP follows the reverse path of the F-EAP, updating the pheromone tables in each node. In an OBS network, the average burst blocking probability on a fibre link can be estimated with the Erlang-B formula, assuming an average offered traffic load of *x* and a Poisson burst arrival process,

$$\mathrm{Erl}(x, W) = \left(x^W / W! \right) \cdot \left(\sum_{k=0}^{W} x^k / k! \right)^{-1} . \quad (5)$$

The wavelength availability $\hat{W}_{ij}$, collected by the F-EAP, is used to estimate the traffic load offered to the fibre link between *i* and *j* as $x = W - \hat{W}_{ij}$. Under the link independence assumption [14], the average burst blocking probability in the routing path between nodes *s* and *d* can be estimated by

$$B^{sd} = 1 - \prod_{k=1}^{|V_{\mathrm{EAP}}|-1} (1 - \hat{B}_{v_k v_{k+1}}) , \quad (6)$$

where $\{v_1, v_2, ..., v_k, ..., v_{|V_{\mathrm{EAP}}|}\}$ denotes the ordered set of the nodes visited by the F-EAP and $\hat{B}_{v_k v_{k+1}}$ the average burst blocking probability in the fibre link between v_k and v_{k+1}. The goodness of a routing path is quantified in DABR as

$$q^{sd} = 1 / B^{sd} . \quad (7)$$

Each node *i* maintains the goodness of the best solution q^{sd}_{best} estimated by the previous *Q* artificial ants between *s* and *d* and the pheromone reinforcement value *r* is given by

$$r = \begin{cases} q^{sd} / q^{sd}_{\mathrm{best}}, & q^{sd} < q^{sd}_{\mathrm{best}} \\ 1, & \text{otherwise} \end{cases} . \quad (8)$$

In order to enhance the rewarding of good solutions (high values of *r*), while saturating the rewards for low values of *r*, the following squash function *s*(*x*) is used [5],

$$s(x) = \left(1 + \exp\left(\frac{1}{r \mid V_{\mathrm{EAP}} \mid} \right) \right)^{-1} , \quad (9)$$

$$r \leftarrow \tau_{\max} \cdot s(r) / s(1) , \quad (10)$$

where $\tau_{\max}$ is the maximum pheromone value that a single artificial ant can deposit. This parameter is used to regulate the impact of a single artificial ant on the pheromone tables.

The pheromone update has to ensure a minimum amount of pheromone $\tau_{\min}$ uniformly divided by the adjacent nodes,

$$\tau^{sd}_{ik} \geq \tau_{\min} / | A_i | , \quad k \in A_i , \quad (11)$$

and it also has to guarantee that equation (1) holds. Hence, the following expressions have been devised for updating the pheromone value associated to the adjacent node *j* the F-EAP has used and that associated to the remaining adjacent nodes,

$$\tau^{sd}_{ij} \leftarrow \tau^{sd}_{ij} + r \cdot \left(1 - \tau^{sd}_{ij} - \tau_{\min} \cdot (| A_i | - 1) / | A_i | \right) , \quad (12)$$

$$\tau^{sd}_{ik} \leftarrow \tau^{sd}_{ik} - r \cdot \left(\tau^{sd}_{ik} - \tau_{\min} / | A_i | \right) , \quad k \in A_i , k \neq j . \quad (13)$$

In case the update has modified the adjacent node that has the highest pheromone value, designated as the dominant adjacent node, a specific flag of the B-EAP is activated.

4.4 Routing Table Update and Circular Path Avoidance

In the initial implementation of DABR, data bursts were routed through the path obtained from the adjacent nodes with the largest pheromone concentration. However, it was found that circular paths were occasionally formed, which resulted in data bursts getting trapped on a loop. Although these paths are only temporary, the resulting communication disruption is not tolerable in a high capacity optical network. Hence, it is imperative to include in DABR a mechanism to avoid the formation of circular paths. The novel mechanism devised for this purpose separates the pheromone and routing tables, guaranteeing that the update of the routing paths never allows the formation of loops without introducing additional restrictions to the update of the pheromone tables.

Let v^{sd}_i denote the adjacent node of node *i* used for routing data bursts from node *s* to node *d*. The update of the routing path between *s* and *d* is triggered by the arrival at node *s* of a B-EAP with the flag that signals a new dominant adjacent node activated. As a result, node *s* generates a F-RAP directed to node *d*. This control packet includes a list to record the nodes it visits. At each node *i* the F-RAP visits, the dominant adjacent node *j** is determined, that is,

$$j^* = \arg\max_{j \in A_i} \left\{ \tau^{sd}_{ij} \right\} , \quad (14)$$

and compared with the nodes previously visited. In case node *j** has already been visited, a loop is detected and the F-RAP is discarded, keeping the routing path unchanged. Otherwise, the F-RAP adds the dominant adjacent node *j** to the list of nodes visited and moves to it. In case *j** = *d*, the new routing path is non-circular. Thus, the F-RAP is switched to B-RAP, follows the inverse path and, at each node *i*, the adjacent node used to route data bursts, v^{sd}_i, is updated to be the same as the node visited by the F-RAP immediately after *i*.

The routing of data bursts in DABR is determined by the routing tables of the nodes. In particular, when a BHP, preceding the correspondent data burst, arrives at node *i*, the node determines the next node of the data burst from v^{sd}_i.

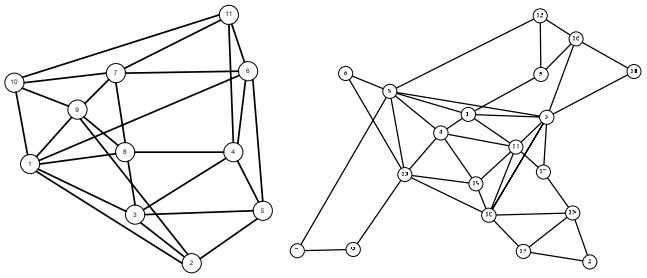
ISBN: 978-989-95079-6-8

5 Simulation Results

In OBS networks, the most relevant performance metric is the average burst blocking probability. The loss performance of OBS networks is evaluated using the network simulator of [15] modified to support DABR. All the simulations use the JET resource reservation mechanism [8], each fibre link has W=32 wavelength channels, each with a capacity μ=10 Gb/s, the time to process one BHP or artificial packet is 100 μs and the switch fabric configuration time is 160 μs. The burst size and interarrival time are exponentially distributed and the average burst size is 10 MB. The average offered traffic load normalized to the network capacity is given by

$$\Gamma = \frac{\sum_{s,d \in V} \gamma^{sd} \cdot h_{sd}}{L \cdot W \cdot \mu}, \qquad (15)$$

where γ^{sd} is the average offered traffic load between nodes s and d. The performance evaluation is made with the network topologies of Fig. 1, assuming a uniform traffic pattern for EON and a non-uniform traffic pattern for COST 239. The average burst blocking probability and a 95% confidence interval on this value are obtained from 10 independent simulations with 5×10^6 data bursts each.

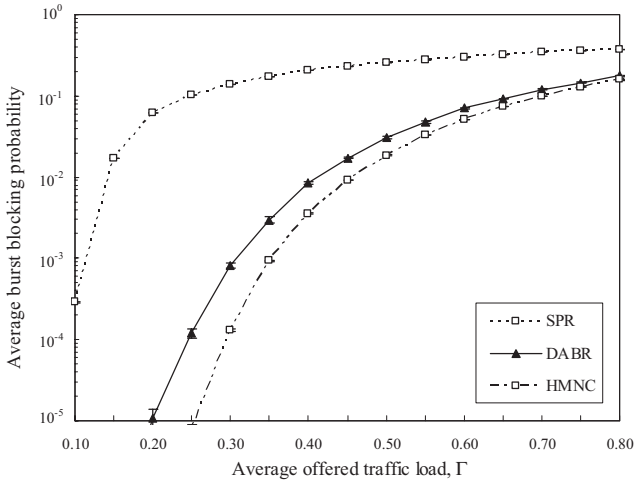


(a) COST 239 (b) EON

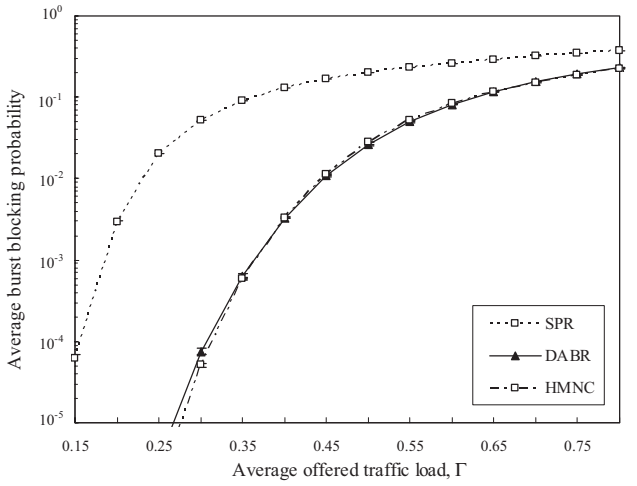
Figure 1: Network topologies.

Figure 2 plots the average burst blocking probability as a function of the average offered traffic load for the COST 239 and EON networks, respectively. The parameters used for DABR are p_{ant}=0.05, α=0.25, τ_{min}=0.2, τ_{max}=0.1, and Q=50, aiming a compromise between performance and adaptability to changing conditions. For comparison purposes the plots also include the performance obtained with Shortest Path Routing (SPR) and the centralized HMNC algorithm of [11].

The simulation results show that the performance obtained with DABR lies between that of SPR and HMNC, although always closer to the latter. This is an expected outcome. On one hand, the simpler SPR does not optimize the traffic load distribution over the fiber links, which can significantly increase contention in some of the links. On the other hand, the HMNC algorithm is more effective because it exploits the global knowledge of accurate traffic statistics. However, in more dynamic traffic scenarios, where traffic statistics are less reliable, HMNC may become less effective than the proposed DABR. The simulation results also show that larger confidence intervals are obtained with DABR than that with SPR and HMNC. This is a consequence of the online modification of the routing paths observed when using DABR, while with SPR and HMNC the same set of routing paths is used during the entire simulation.



(a) 11-node COST 239 network



(b) 19-node EON network

Figure 2: Network performance of DABR, SPR and HMNC.

The performance impact of varying some of the parameters of DABR is assessed in the COST 239 network. In all cases, the parameter values given above are used with the exception of the parameter being varied. The average burst blocking probability as a function of the minimum pheromone value, maximum pheromone increment per EAP and size of sliding window used to record the best EAP solution are plotted in Figures 3, 4 and 5, respectively.

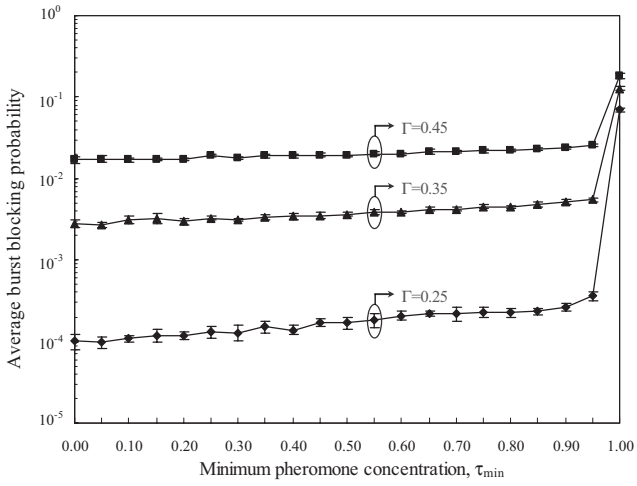


Figure 3: Performance impact of τ_{min}.

ISBN: 978-989-95079-6-8

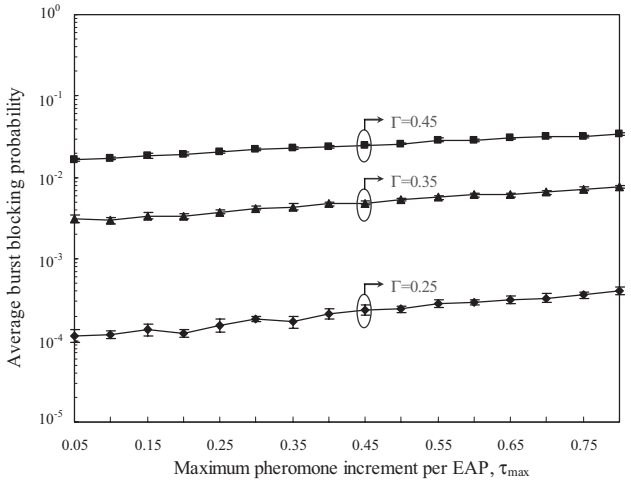


Figure 4: Performance impact of τ_{max}.

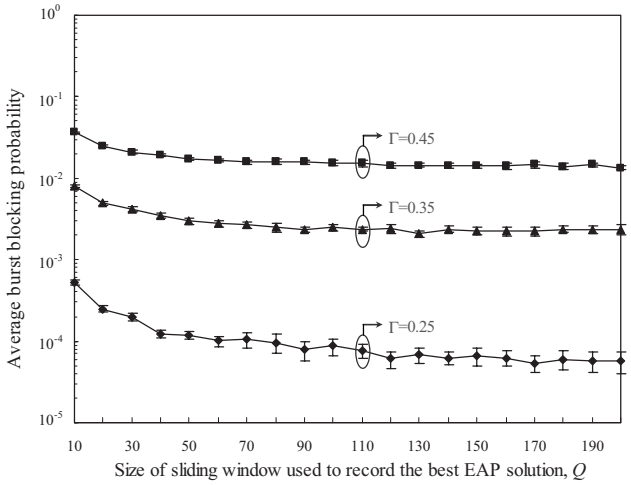


Figure 5: Performance impact of Q.

The results above consider three values of average offered traffic load in order to confirm that the behaviour observed is independent of the traffic load. According to the simulation results of Fig. 3, the loss performance slightly deteriorates as the minimum pheromone concentration on the adjacent nodes is increased. This is because the random component of routing the EAPs becomes larger, increasing the chances of modifications in the routing paths. In the limit case $\tau_{min}=1$, the network cannot optimize the set of routing paths.

From Fig. 4, it can be seen that the average burst blocking probability is slightly increased as the EAPs are allowed to deposit more pheromone in a single trip. This behaviour is attributed to more frequent modifications of the routing paths as a consequence of increased fluctuations of the pheromone values. A smaller pheromone increment makes the routing paths less sensitive to the goodness of the solution found by a single artificial ant, which means they are changed only when a significant number of EAPs have followed a given routing path and confirmed its superior quality.

As shown in Fig. 5, the use of a large Q can improve the network performance because it enhances the accuracy of the pheromone update process.

Overall, given the relatively small range of variation of the average burst blocking probability, the results suggest that DABR is quite robust to changes in its main parameters. The results also suggest that a larger Q and a smaller τ_{min} and τ_{max} slightly improve loss performance. However, it should be stressed that the simulations consider a single traffic pattern (in terms of average offered traffic load), whereas in more dynamic traffic scenarios a very large Q and very small τ_{min} and τ_{max} should be avoided because they can slow down the convergence to a new set of optimized routing paths whenever significant changes in the traffic pattern take place.

6 Conclusions

This paper has proposed a novel distributed routing path optimization framework for OBS networks based on ACO concepts. Simulation results have shown that the proposed framework can significantly improve network performance without needing to fine tune its parameters.

References

[1] M. Dorigo, M. Birattari and T. Stützle, Ant colony optimization: Artificial ants as a computational intelligence technique, *IEEE Computational Intelligence Magazine*, 1 (4): 28-39, Nov 2006.

[2] E. Bonabeau, M. Dorigo and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press, 1999.

[3] M. Dorigo and T. Stützle, *Ant Colony Optimization*, Cambridge: MIT Press, 2004.

[4] R. Schoonderwoerd, O. Holland, J. Bruten and L. Rothkrantz, Ant-based load balancing in telecommunications networks, *Adaptive Behavior*, 5 (2): 169-207, 1996.

[5] G. Di Caro, and M. Dorigo, AntNet: Distributed stigmergetic control for communications networks, *Journal of Artificial Intelligence Research*, 9: 317-365, 1998.

[6] R. Garlick and R. Barr, Dynamic wavelength routing in WDM networks via ant colony optimization, in *Proc. of International Workshop on Ant Algorithms* (*ANTS*) *2002*, 250-255, Brussels, Belgium.

[7] S. Ngo, X. Jiang and S. Horiguchi, An ant-based approach to dynamic RWA in optical WDM networks, *Photonic Network Communications*, 11 (1): 39-48, January 2006.

[8] C. Qiao, and M. Yoo, Optical Burst Switching (OBS) – A new paradigm for an optical Internet, *Journal of High Speed Networks*, 8 (1): 69-84, January 1999.

[9] G. Pavani and H. Waldman, Traffic engineering and restoration in optical packet switching networks by means of ant colony optimization, in *Proc. of BROADNETS 2006*, San Jose, USA.

[10] Y. Chen, C. Qiao and X. Yu, Optical burst switching: A new area in optical network research, *IEEE Network*, 18 (3): 16-23, May 2004.

[11] J. Castro, J. Pedro and P. Monteiro, Burst loss reduction in OBS networks by minimizing network congestion, in *Proc. of ConfTele 2005*, Tomar, Portugal.

[12] G. Thodine, V. Vokkarane and J. Jue, Dynamic congestion-based load balanced routing in optical burst-switched networks, in *Proc. of IEEE GLOBECOM 2003*, 5: 2628-2632, San Francisco, USA.

[13] K. Sim and W. Sun, Ant colony optimization for routing and load-balancing: Survey and new directions, *IEEE Transactions on Systems, Man, and Cybernetics*, 33 (5): 560-572, September 2003.

[14] F. Kelly, Loss networks, *The Annals of Applied Probability*, 1 (3): 319-378, 1991.

[15] J. Pedro, J. Castro, P. Monteiro and J. Pires, On the modelling and performance evaluation of optical burst-switched networks, in *Proc. of IEEE CAMAD 2006*, 30-37, Trento, Italy.

ISBN: 978-989-95079-6-8

Non-conjunctive and non-disjunctive uninorms in Atanassov's intuitionistic fuzzy set theory

Glad Deschrijver

Fuzziness and Uncertainty Modelling Research Unit,
Department of Applied Mathematics and Computer Science, Ghent University,
Krijgslaan 281 (S9), B–9000 Gent, Belgium
Email: Glad.Deschrijver@UGent.be

***Abstract*— *Uninorms are a generalization of t-norms and t-conorms for which the neutral element is an element of* $[0,1]$ *which is not necessarily equal to* 0 *(as for t-norms) or* 1 *(as for t-conorms). Uninorms on the unit interval are either conjunctive or disjunctive, i.e. they aggregate the pair* $(0,1)$ *to either* 0 *or* 1. *In real-life applications, this kind of aggregation may be counter-intuitive. Atanassov's intuitionistic fuzzy set theory is an extension of fuzzy set theory which allows to model uncertainty about the membership degrees. In Atanassov's intuitionistic fuzzy set theory there exist uninorms which are neither conjunctive nor disjunctive. In this paper we study such uninorms more deeply and we investigate the structure of these uninorms. We also give several examples of uninorms which are neither conjunctive nor disjunctive.*

***Keywords*—** Conjunctive, disjunctive, interval-valued fuzzy set, intuitionistic fuzzy set, uninorm.

1 Introduction

Interval-valued fuzzy set theory [1, 2] is an extension of fuzzy set theory in which to each element of the universe a closed subinterval of the unit interval is assigned which approximates the unknown membership degree (using interval-valued fuzzy sets is not always the best approach to deal with uncertainty, see [3] for more information). Another extension of fuzzy set theory is intuitionistic fuzzy set theory introduced by Atanassov [4]. In [5] it is shown that intuitionistic fuzzy set theory is equivalent to interval-valued fuzzy set theory and that both are equivalent to L-fuzzy set theory in the sense of Goguen [6] w.r.t. a special lattice $\mathcal{L}^I$.

Uninorms are an important generalization of t-norms and t-conorms introduced by Yager and Rybalov [7]. Uninorms allow for a neutral element lying anywhere in the unit interval rather than at one or zero as is the case for t-norms and t-conorms. Uninorms on the unit interval are either conjunctive or disjunctive, i.e. they aggregate the pair $(0,1)$ to either 0 or 1. In real-life applications, this kind of aggregation may be counter-intuitive, e.g. in customer satisfaction modelling, if an aspect of the product receives a negative evaluation and another aspect a positive evaluation, then in general the global evaluation will neither be very negative or very positive, but rather be quite uncertain. This situation can be modelled by using uninorms in Atanassov's intuitionistic fuzzy set theory, which can be neither conjunctive nor disjunctive (see [8]). In this paper we therefore investigate such uninorms more deeply.

Definition 1.1 *We define* $\mathcal{L}^I = (L^I, \leq_{L^I})$, *where*

$$L^I = \{[x_1,x_2] \mid (x_1,x_2) \in [0,1]^2 \text{ and } x_1 \leq x_2\}, \tag{1}$$

$$[x_1,x_2] \leq_{L^I} [y_1,y_2] \iff (x_1 \leq y_1 \text{ and } x_2 \leq y_2), \tag{2}$$

for all $[x_1,x_2],[y_1,y_2]$ *in* L^I.

Similarly as Lemma 2.1 in [5] it can be shown that $\mathcal{L}^I$ is a complete lattice.

Definition 1.2 [1, 2] *An interval-valued fuzzy set on* U *is a mapping* $A : U \to L^I$.

Definition 1.3 [4] *An intuitionistic fuzzy set on* U *is a set*

$$A = \{(u, \mu_A(u), \nu_A(u)) \mid u \in U\}, \tag{3}$$

where $\mu_A(u) \in [0,1]$ *denotes the membership degree and* $\nu_A(u) \in [0,1]$ *the non-membership degree of* u *in* A *and where for all* $u \in U$, $\mu_A(u) + \nu_A(u) \leq 1$.

An intuitionistic fuzzy set A on U can be represented by the $\mathcal{L}^I$-fuzzy set A given by

$$\begin{array}{rcl} A : U & \to & L^I : \\ u & \mapsto & [\mu_A(u), 1-\nu_A(u)], \end{array} \tag{4}$$

In Figure 1 the set L^I is shown. Note that each $x = [x_1,x_2] \in L^I$ is represented by the point $(x_1,x_2) \in \mathbb{R}^2$.

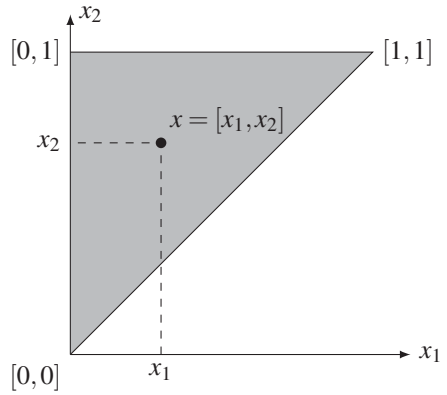


Figure 1: The grey area is L^I.

In the sequel, if $x \in L^I$, then we denote its bounds by x_1 and x_2, i.e. $x = [x_1,x_2]$. The smallest and the largest element of $\mathcal{L}^I$

ISBN: 978-989-95079-6-8

are given by $0_{\mathcal{L}^I} = [0,0]$ and $1_{\mathcal{L}^I} = [1,1]$. Note that, for x, y in L^I, $x <_{L^I} y$ is equivalent to $x \leq_{L^I} y$ and $x \neq y$, i.e. either $x_1 < y_1$ and $x_2 \leq y_2$, or $x_1 \leq y_1$ and $x_2 < y_2$. If for x, y in L^I it holds that either $x_1 < y_1$ and $x_2 > y_2$, or $x_1 > y_1$ and $x_2 < y_2$, then x and y are incomparable w.r.t. $\leq_{L^I}$, denoted as $x \|_{L^I} y$. We define for further usage the set $D = \{[x_1, x_1] \mid x_1 \in [0,1]\}$.

Definition 1.4 *A t-norm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{T} : (L^I)^2 \to L^I$ which satisfies $\mathcal{T}(1_{\mathcal{L}^I}, x) = x$, for all $x \in L^I$.*

A t-conorm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{S} : (L^I)^2 \to L^I$ which satisfies $\mathcal{S}(0_{\mathcal{L}^I}, x) = x$, for all $x \in L^I$.

Definition 1.5 *A negation on $\mathcal{L}^I$ is a decreasing mapping $\mathcal{N} : L^I \to L^I$ for which $\mathcal{N}(0_{\mathcal{L}^I}) = 1_{\mathcal{L}^I}$ and $\mathcal{N}(1_{\mathcal{L}^I}) = 0_{\mathcal{L}^I}$. If $\mathcal{N}(\mathcal{N}(x)) = x$, for all $x \in L^I$, then $\mathcal{N}$ is called involutive.*

Let N be a negation on $([0,1], \leq)$. Then the mapping $\mathcal{N}_N : L^I \to L^I$ defined by, for all $x \in L^I$,

$$\mathcal{N}_N(x) = [N(x_2), N(x_1)], \tag{5}$$

is a negation on $\mathcal{L}^I$.

We will also need the following result and definition (see [9, 10, 11, 12, 13]).

Theorem 1.1 *Let $(T_\alpha)_{\alpha \in A}$ be a family of t-norms on $([0,1], \leq)$ and $(]a_\alpha, e_\alpha[)_{\alpha \in A}$ be a family of non-empty, pairwise disjoint open subintervals of $[0,1]$. Then the function $T : [0,1]^2 \to [0,1]$ defined by, for all x, y in $[0,1]$,*

$$T(x,y) = \begin{cases} a_\alpha + (e_\alpha - a_\alpha) \cdot T_\alpha\left(\dfrac{x - a_\alpha}{e_\alpha - a_\alpha}, \dfrac{y - a_\alpha}{e_\alpha - a_\alpha}\right), \\ \quad \text{if } (x,y) \in [a_\alpha, e_\alpha]^2, \\ \min(x,y), \quad \text{otherwise}, \end{cases} \tag{6}$$

is a t-norm on $([0,1], \leq)$.

Definition 1.6 *Let $(T_\alpha)_{\alpha \in A}$ be a family of t-norms on $([0,1], \leq)$ and $(]a_\alpha, e_\alpha[)_{\alpha \in A}$ be a family of non-empty, pairwise disjoint open subintervals of $[0,1]$. The t-norm T defined by (6) is called the ordinal sum of the summands $\langle a_\alpha, e_\alpha, T_\alpha \rangle$, $\alpha \in A$, and we will write*

$$T = (\langle a_\alpha, e_\alpha, T_\alpha \rangle)_{\alpha \in A}. \tag{7}$$

2 Uninorms on $\mathcal{L}^I$

The following definition of a uninorm on $\mathcal{L}^I$ is a straightforward generalization of the definition of a uninorm on the unit interval introduced by Yager and Rybalov [7, 14].

Definition 2.1 [8] *A uninorm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{U} : (L^I)^2 \to L^I$ for which there exists an $e \in L^I$ such that $\mathcal{U}(e,x) = x$, for all $x \in L^I$. The element e is called the neutral element of $\mathcal{U}$.*

For any uninorm U on the unit interval, there exist increasing bijections $\phi_e : [0,e] \to [0,1]$ and $\psi_e : [e,1] \to [0,1]$ with increasing inverse, a t-norm T_U and a t-conorm S_U on $([0,1], \leq)$ such that [14]

(i) $(\forall (x,y) \in [0,e]^2)(U(x,y) = \phi_e^{-1}(T_U(\phi_e(x), \phi_e(y))))$;

(ii) $(\forall (x,y) \in [e,1]^2)(U(x,y) = \psi_e^{-1}(S_U(\psi_e(x), \psi_e(y))))$.

Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in L^I$. We define $E = \{x \mid x \in L^I \text{ and } x \leq_{L^I} e\}$ and $E' = \{x \mid x \in L^I \text{ and } x \geq_{L^I} e\}$. In [8] it is shown that if $e \notin D$, then there does not exist increasing bijections $\Phi_e : E \to L^I$ and $\Psi_e : E' \to L^I$ such that Φ_e^{-1} and Ψ_e^{-1} are increasing. On the other hand, if $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$, then the mappings $\Phi_e : E \to L^I$ and $\Psi_e : E' \to L^I$ defined by, for all $x \in L^I$,

$$\Phi_e(x) = \left[\frac{x_1}{e_1}, \frac{x_2}{e_1}\right], \tag{8}$$

$$\Psi_e(x) = \left[\frac{x_1 - e_1}{1 - e_1}, \frac{x_2 - e_1}{1 - e_1}\right]. \tag{9}$$

are increasing bijections for which the inverse is also increasing. As a consequence, the above result can only be extended if $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$.

From now on, we denote for any t-norm T and t-conorm S on $([0,1], \leq)$, $T_{\phi_e} = \phi_e^{-1} \circ T \circ (\phi_e \times \phi_e)$ and $S_{\psi_e} = \psi_e^{-1} \circ S \circ (\psi_e \times \psi_e)$, where $\times$ denotes the product operation [15]. A similar notation will be used for t-(co)norms and bijections on $\mathcal{L}^I$.

Theorem 2.1 [8] *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then:*

(i) the mapping $\mathcal{T}_\mathcal{U} : (L^I)^2 \to L^I$ defined by, for all $x, y \in L^I$,

$$\mathcal{T}_\mathcal{U}(x,y) = \Phi_e(\mathcal{U}(\Phi_e^{-1}(x), \Phi_e^{-1}(y))) \tag{10}$$

is a t-norm on $\mathcal{L}^I$;

(ii) the mapping $\mathcal{S}_\mathcal{U} : (L^I)^2 \to L^I$ defined by, for all $x, y \in L^I$,

$$\mathcal{S}_\mathcal{U}(x,y) = \Psi_e(\mathcal{U}(\Psi_e^{-1}(x), \Psi_e^{-1}(y))) \tag{11}$$

is a t-conorm on $\mathcal{L}^I$.

Theorem 2.2 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then for all x, y in L^I,*

$$x \leq_{L^I} e \leq_{L^I} y \implies \inf(x,y) \leq_{L^I} \mathcal{U}(x,y) \leq_{L^I} \sup(x,y). \tag{12}$$

These properties show that uninorms are well suited to model human evaluations (e.g. customer satisfaction). Customers which evaluate the performance of all aspects of a certain product high, have a tendency to give the global satisfaction degree an even higher value; on the other hand customers which globally consider the performance of the various aspects as insufficient, will give a low global evaluation. So we observe "reinforcement": a collection of high (low) rates "reinforce" each other and yield a global evaluation rate that is even higher (resp. lower) than each individual rate. If, however, a customer gives high scores only to some aspects and low scores for other aspects, then the global score will in general be located between the lowest and the highest value. This is "compensation". From Theorem 2.1 it follows that $\mathcal{U}\big|_{E^2}$ behaves like a t-norm, in particular $\mathcal{U}(x,y) \leq_{L^I} \inf(x,y)$, for all x, y in E. On the other hand, $\mathcal{U}\big|_{E'^2}$ behaves like a t-conorm, so $\mathcal{U}(x,y) \geq_{L^I} \sup(x,y)$, for all x, y in E'. Finally, if $x \leq_{L^I} e$ and $y \geq_{L^I} e$ (or conversely), then $\mathcal{U}(x,y)$ is a number between

ISBN: 978-989-95079-6-8

$\inf(x,y)$ and $\sup(x,y)$. So, clearly, uninorms show a reinforcing behaviour on E^2 and E'^2, and a compensating behaviour on $E \times E'$ and $E' \times E$ (see [16, 17, 18, 19] for more details).

For uninorms on the unit interval, however, $U(0,1)$ can only have two values: 0 or 1 (see [14]). In the first case the uninorm is called "conjunctive" and in the second case "disjunctive". However, in both cases the compensatory behaviour of the uninorm is violated. For uninorms on $\mathcal{L}^I$ we have the following property.

Theorem 2.3 [8] *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then either $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = 0_{\mathcal{L}^I}$ or $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = 1_{\mathcal{L}^I}$ or $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$.*

Hence uninorms on $\mathcal{L}^I$ are not necessarily conjunctive or disjunctive. It is possible that a uninorm on $\mathcal{L}^I$ shows compensatory behaviour between $0_{\mathcal{L}^I}$ and $1_{\mathcal{L}^I}$. If one aspect of a product has a very negative evaluation ($0_{\mathcal{L}^I}$) and another aspect is very positively evaluated ($1_{\mathcal{L}^I}$), then in general it will be very difficult to give a global evaluation of the product, in fact the global evaluation will contain a lot of uncertainty. Therefore it makes more sense to use a uninorm $\mathcal{U}$ for which $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$.

3 Uninorms on $\mathcal{L}^I$ which are neither conjunctive nor disjunctive

In this section we try to obtain more information about the structure of uninorms which are neither conjunctive nor disjunctive by investigating the possible values of $\mathcal{U}(x,y)$ with x, y in L^I. First we give an example of a uninorm on $\mathcal{L}^I$ that is neither conjunctive nor disjunctive, in order to show that such uninorms do exists.

Example 3.1 *Let for all $e_1 \in]0,1[$, U_{e_1} be the uninorm on $([0,1], \leq)$ defined by, for all x_1, y_1 in $[0,1]$,*

$$U_{e_1}(x_1,y_1) = \begin{cases} \max(x_1,y_1), & \text{if } x_1 \geq e_1 \text{ and } y_1 \geq e_1; \\ \min(x_1,y_1), & \text{else.} \end{cases} \quad (13)$$

Let now, for all x, y in L^I,

$$\mathcal{U}(x,y) = [U_{e_1}(x_1,y_1), 1 - U_{1-e_1}(1-x_2, 1-y_2)]. \quad (14)$$

Then $\mathcal{U}$ is a uninorm on $\mathcal{L}^I$ with neutral element $e = [e_1, e_1]$. Since $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = [0,1]$, $\mathcal{U}$ is neither conjunctive nor disjunctive.

In general, if U_1 is an arbitrary conjunctive uninorm and U_2 an arbitrary disjunctive uninorm on $([0,1], \leq)$, then the mapping $\mathcal{U} : (L^I)^2 \to L^I : (x,y) \mapsto [U_1(x_1,y_1), U_2(x_2,y_2)]$, for all x, y in L^I, is a uninorm on $\mathcal{L}^I$ for which $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = [0,1]$.

Lemma 3.1 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then, for all $x \in L^I$,*

(i) either $\mathcal{U}(0_{\mathcal{L}^I}, x) = 0_{\mathcal{L}^I}$ or $\mathcal{U}(0_{\mathcal{L}^I}, x) \notin E$,

(ii) either $\mathcal{U}(1_{\mathcal{L}^I}, x) = 1_{\mathcal{L}^I}$ or $\mathcal{U}(1_{\mathcal{L}^I}, x) \notin E'$.

Theorem 3.2 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. If $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$, then, for all $x \in L^I$,*

(i) $\mathcal{U}(0_{\mathcal{L}^I}, x) \parallel_{L^I} e$ or $\mathcal{U}(0_{\mathcal{L}^I}, x) = 0_{\mathcal{L}^I}$,

(ii) $\mathcal{U}(1_{\mathcal{L}^I}, x) \parallel_{L^I} e$ or $\mathcal{U}(1_{\mathcal{L}^I}, x) = 1_{\mathcal{L}^I}$.

If one aspect of a product has a negative evaluation $x \in L^I$ with $x \leq_{L^I} e$ and another aspect has a positive evaluation $y \in L^I$ with $y \geq_{L^I} e$, then the global evaluation will be rather neutral and contain some uncertainty. Therefore it is natural to expect that $\mathcal{U}(x,y) \parallel_{L^I} e$. We investigate for which x and y in L^I this is the case.

Lemma 3.3 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Assume that $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$.*

(i) Let arbitrarily $x \in E$. If $\mathcal{U}(1_{\mathcal{L}^I}, x) = 1_{\mathcal{L}^I}$, then $\mathcal{U}(1_{\mathcal{L}^I}, [x_1, y_2]) = 1_{\mathcal{L}^I}$, for all $y_2 \in [x_1, e_1]$.

(ii) Let arbitrarily $x \in E'$. If $\mathcal{U}(0_{\mathcal{L}^I}, x) = 0_{\mathcal{L}^I}$, then $\mathcal{U}(0_{\mathcal{L}^I}, [y_1, x_2]) = 0_{\mathcal{L}^I}$, for all $y_1 \in [e_1, x_2]$.

Theorem 3.4 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. If $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$, then*

(i) there exists an $\alpha \in D \cap E$ such that (see Figure 2)

- *$\mathcal{U}(1_{\mathcal{L}^I}, x) \parallel_{L^I} e$ for all $x \in L^I$ satisfying $x_1 < \alpha_1$ and $x_2 \leq e_1$, and*
- *$\mathcal{U}(1_{\mathcal{L}^I}, x) = 1_{\mathcal{L}^I}$, for all $x \in L^I$ satisfying $x_1 > \alpha_1$,*

(ii) there exists a $\beta \in D \cap E'$ such that

- *$\mathcal{U}(0_{\mathcal{L}^I}, x) \parallel_{L^I} e$ for all $x \in L^I$ satisfying $x_1 \geq e_1$ and $x_2 > \beta_1$, and*
- *$\mathcal{U}(0_{\mathcal{L}^I}, x) = 0_{\mathcal{L}^I}$, for all $x \in L^I$ satisfying $x_2 < \beta_1$.*

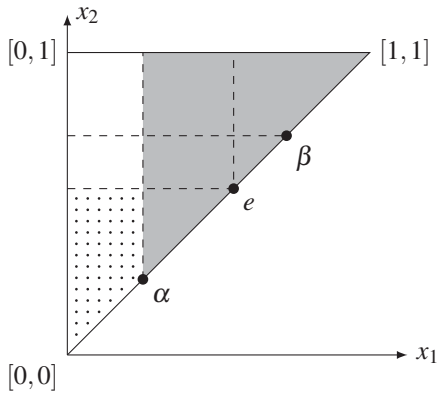


Figure 2: The grey area is the set of elements x for which $\mathcal{U}(x, 1_{\mathcal{L}^I}) = 1_{\mathcal{L}^I}$, the dotted area is the set of elements x for which $\mathcal{U}(x, 1_{\mathcal{L}^I}) \parallel_{L^I} e$.

Example 3.2 We give an example of a uninorm which satisfies the results in Theorem 3.4 for a non-trivial α and β. Let arbitrarily $\alpha \in D \cap E \setminus \{0_{\mathcal{L}^I}, e\}$ and $\beta \in D' \cap E' \setminus \{e, 1_{\mathcal{L}^I}\}$. Let T_{1a} and T_{1b} be arbitrary t-norms, S_{1a} and S_{1b} arbitrary t-conorms on $([0,1], \leq)$, and define

$$T_1 = (\langle 0, \phi_e(\alpha_1), T_{1a} \rangle, \langle \phi_e(\alpha_1), 1, T_{1b} \rangle), \quad (15)$$

$$S_2 = (\langle 0, \psi_e(\beta_1), S_{2a} \rangle, \langle \psi_e(\beta_1), 1, S_{2b} \rangle). \quad (16)$$

ISBN: 978-989-95079-6-8

Define the mappings $U_1 : [0,1]^2 \to [0,1]$ and $U_2 : [0,1]^2 \to [0,1]$ by, for all x_1, y_1, x_2, y_2 in $[0,1]$,

$$U_1(x_1,y_1) = \begin{cases} (T_1)_{\phi_e}(x_1,y_1), \text{ if } \max(x_1,y_1) \le e_1, \\ (S_1)_{\psi_e}(x_1,y_1), \text{ if } \min(x_1,y_1) \ge e_1, \\ 1, \text{ if } (x_1 > \alpha_1 \text{ and } y_1 = 1) \\ \quad \text{or } (y_1 > \alpha_1 \text{ and } x_1 = 1), \\ \min(x_1,y_1), \text{ else}, \end{cases} \tag{17}$$

$$U_2(x_2,y_2) = \begin{cases} (T_2)_{\phi_e}(x_2,y_2), \text{ if } \max(x_2,y_2) \le e_1, \\ (S_2)_{\psi_e}(x_2,y_2), \text{ if } \min(x_2,y_2) \ge e_1, \\ 0, \text{ if } (x_2 < \beta_1 \text{ and } y_2 = 0) \\ \quad \text{or } (y_2 < \beta_1 \text{ and } x_2 = 0), \\ \max(x_2,y_2), \text{ else}, \end{cases} \tag{18}$$

Then U_1 is a conjunctive uninorm and U_2 is a disjunctive uninorm on $([0,1],\le)$. The mapping $\mathcal{U} : (L^I)^2 \to L^I$ defined by, for all x, y in L^I,

$$\mathcal{U}(x,y) = [U_1(x_1,y_1), U_2(x_2,y_2)], \tag{19}$$

is a uninorm on $\mathcal{L}^I$ for which $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = [0,1]$ and for which the results in Theorem 3.4 hold for the given α and β.

From now on α and β will be the elements of $\mathcal{L}^I$ introduced in Theorem 3.4.

Lemma 3.5 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. If $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \|_{L^I} e$, then for all $x \in E$ and $y \in E'$ satisfying $x_1 < \alpha_1$ and $y_2 > \beta_1$ it holds that $\mathcal{U}(x,y) \|_{L^I} e$.*

Theorem 3.6 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. If $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \|_{L^I} e$, then for all $x \in E$ and $y \in E'$ satisfying $x_1 < \alpha_1$ and $y_2 > \beta_1$ it holds that $(\mathcal{U}(x,y))_1 \le \alpha_1$ and $(\mathcal{U}(x,y))_2 \ge \beta_1$.*

Corollary 3.7 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Assume that $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \|_{L^I} e$.*

(i) Let arbitrarily $a = [\alpha_1, a_2] \in E$ and $y \in E'$ such that $y_2 > \beta_1$. Then

$$\lim_{\substack{x \to a \\ x_1 < \alpha_1}} (\mathcal{U}(x,y))_1 = \alpha_1. \tag{20}$$

(ii) Let arbitrarily $b = [b_1, \beta_2] \in E'$ and $x \in E$ such that $x_1 < \alpha_1$. Then

$$\lim_{\substack{y \to b \\ y_2 > \beta_2}} (\mathcal{U}(x,y))_2 = \beta_2. \tag{21}$$

In the above, the limits are calculated using on L^I the Euclidean metric function $d^E(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$, for all x, y in L^I.

Theorem 3.8 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Assume that $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \|_{L^I} e$.*

(i) For all $x \in E$ and $y \in E'$ satisfying $x_1 > \alpha_1$ and $y_2 > \beta_1$ it holds that $\mathcal{U}(x,y) \ge_{L^I} [\alpha_1, \beta_1]$.

(ii) For all $x \in E$ and $y \in E'$ satisfying $x_1 < \alpha_1$ and $y_2 < \beta_1$ it holds that $\mathcal{U}(x,y) \le_{L^I} [\alpha_1, \beta_1]$.

4 The value of $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I})$

In this section we check which are the possible values for $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I})$ in the case that $\mathcal{U}$ is neither conjunctive nor disjunctive.

Lemma 4.1 *For any $\alpha \in L^I$ and $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ such that $\alpha \|_{L^I} e$, $\alpha_1 > 0$ and $\alpha_2 < 1$, there exists an involutive negation N on $([0,1],\le)$ such that $N(\alpha_1) = \alpha_2$ and $N(e_1) = e_1$.*

Theorem 4.2 *Let $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$, $\alpha \in L^I$, T_1 and T_2 be t-norms, S_1 and S_2 be t-conorms on $([0,1],\le)$ such that*

(i) $\alpha \|_{L^I} e$,

(ii) there exist t-norms T_{1a} and T_{1b} on $([0,1],\le)$ such that $T_1 = (\langle 0, \phi_e(\alpha_1), T_{1a}\rangle, \langle \phi_e(\alpha_1), 1, T_{1b}\rangle)$,

(iii) there exist t-conorms S_{2a} and S_{2b} on $([0,1],\le)$ such that $S_2 = (\langle 0, \psi_e(\alpha_2), S_{2a}\rangle, \langle \psi_e(\alpha_2), 1, S_{2b}\rangle)$,

(iv) $T_1(x_1,y_1) \le T_2(x_1,y_1)$ and $S_1(x_1,y_1) \le S_2(x_1,y_1)$, for all x_1, y_1 in $[0,1]$.

Define the mapping $\mathcal{U} : (L^I)^2 \to L^I$ by, for all x, y in L^I,

$$(\mathcal{U}(x,y))_1 = \begin{cases} \alpha_1, \text{ if } (x_1 < \alpha_1 \text{ and } y_1 \ge \alpha_1 \text{ and } y_2 > e_1) \\ \quad \text{or } (y_1 < \alpha_1 \text{ and } x_1 \ge \alpha_1 \text{ and } x_2 > e_1), \\ U_1(x_1,y_1), \text{ else}, \end{cases} \tag{22}$$

$$(\mathcal{U}(x,y))_2 = \begin{cases} \alpha_2, \text{ if } (x_2 > \alpha_2 \text{ and } y_2 \le \alpha_2 \text{ and } y_1 < e_1) \\ \quad \text{or } (y_2 > \alpha_2 \text{ and } x_2 \le \alpha_2 \text{ and } x_1 < e_1), \\ U_2(x_2,y_2), \text{ else}. \end{cases} \tag{23}$$

where, for all x_1, y_1, x_2, y_2 in $[0,1]$,

$$U_1(x_1,y_1) = \begin{cases} (T_1)_{\phi_e}(x_1,y_1), & \text{if } \max(x_1,y_1) \le e_1, \\ (S_1)_{\psi_e}(x_1,y_1), & \text{if } \min(x_1,y_1) \ge e_1, \\ \min(x_1,y_1), & \text{else}, \end{cases} \tag{24}$$

$$U_2(x_2,y_2) = \begin{cases} (T_2)_{\phi_e}(x_2,y_2), & \text{if } \max(x_2,y_2) \le e_1, \\ (S_2)_{\psi_e}(x_2,y_2), & \text{if } \min(x_2,y_2) \ge e_1, \\ \max(x_2,y_2), & \text{else}. \end{cases} \tag{25}$$

Then $\mathcal{U}$ is a uninorm on $\mathcal{L}^I$ with neutral element e for which $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = \alpha$.

Theorem 4.2 shows that for any $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ and any $\alpha \in L^I$ such that $\alpha \|_{L^I} e$, there exists a uninorm $\mathcal{U}$ on $\mathcal{L}^I$ with neutral element e such that $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = \alpha$.

In the following theorem we show that for most values of $\alpha \in L^I$ such that $\alpha \|_{L^I} e$, it is even possible to find uninorms satisfying $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = \alpha$, which are self-dual.

Theorem 4.3 *Let $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$, $\alpha \in L^I$, T be a t-norm, S a t-conorm and N a negation on $([0,1],\le)$ such that*

(i) $\alpha \|_{L^I} e$ and either $\alpha \|_{L^I} [0,1]$ or $\alpha = [0,1]$,

(ii) N is involutive, $N(\alpha_1) = \alpha_2$ and $N(e_1) = e_1$,

(iii) there exist t-norms T_a and T_b on $([0,1],\le)$ such that $T = (\langle 0, \phi_e(\alpha_1), T_a\rangle, \langle \phi_e(\alpha_1), 1, T_b\rangle)$,

ISBN: 978-989-95079-6-8

(iv) $T_{\phi_e}(x_1,y_1) \leq N(S_{\psi_e}(N(x_1),N(y_1)))$, *for all* x_1 *and* y_1 *in* $[0,e_1]$.

Define the mapping $\mathcal{U} : (L^I)^2 \to L^I$ *by, for all* x,y *in* L^I,

$$(\mathcal{U}(x,y))_1 = \begin{cases} \alpha_1, \text{ if } (x_1 < \alpha_1 \text{ and } y_1 \geq \alpha_1 \text{ and } y_2 > e_1) \\ \quad \text{or } (y_1 < \alpha_1 \text{ and } x_1 \geq \alpha_1 \text{ and } x_2 > e_1), \\ U(x_1,y_1), \text{ else}, \end{cases} \tag{26}$$

$$(\mathcal{U}(x,y))_2 = N((\mathcal{U}(\mathcal{N}_N(x),\mathcal{N}_N(y)))_1). \tag{27}$$

where, for all x_1,y_1 *in* $[0,1]$,

$$U(x_1,y_1) = \begin{cases} T_{\phi_e}(x_1,y_1), & \text{if } \max(x_1,y_1) \leq e_1, \\ S_{\psi_e}(x_1,y_1), & \text{if } \min(x_1,y_1) \geq e_1, \\ \min(x_1,y_1), & \text{else}. \end{cases} \tag{28}$$

Then $\mathcal{U}$ *is a uninorm on* $\mathcal{L}^I$ *with neutral element* e *for which* $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I}) = \alpha$ *and, for all* x,y *in* L^I,

$$\mathcal{U}(x,y) = \mathcal{N}_N(\mathcal{U}(\mathcal{N}_N(x),\mathcal{N}_N(y))). \tag{29}$$

Example 4.1 *Let arbitrarily* $e \in D$ *and* $\alpha \in L^I$ *with* $\alpha \|_{L^I} e$ *and* $\alpha \|_{L^I} [0,1]$. *Define for all* $x_1 \in [0,1]$,

$$N(x_1) = \begin{cases} 1 - \frac{1-\alpha_2}{\alpha_1}x_1, & \text{if } x_1 \in [0,\alpha_1], \\ e_1 + \frac{\alpha_2 - e_1}{\alpha_1 - e_1}(x_1 - e_1), & \text{if } x_1 \in [\alpha_1,e_1], \\ e_1 + \frac{\alpha_1 - e_1}{\alpha_2 - e_1}(x_1 - e_1), & \text{if } x_1 \in [e_1,\alpha_2], \\ -\frac{\alpha_1}{1-\alpha_2}(x_1 - e_1), & \text{if } x_1 \in [\alpha_2,1]. \end{cases} \tag{30}$$

Then N *is an involutive negation with* $N(\alpha_1) = \alpha_2$ *and* $N(e_1) = e_1$. *Define* $T = (\langle 0, \phi_e(\alpha_1), P\rangle, \langle \phi_e(\alpha_1), 1, \min\rangle)$, *where* P *is the product t-norm on the unit interval. Then for all* $(x_1,y_1) \in [0,e_1]^2$,

$$T_{\phi_e}(x_1,y_1) = \begin{cases} \frac{1}{\alpha_1}x_1y_1, & \text{if } (x_1,y_1) \in [0,\alpha_1]^2, \\ \min(x_1,y_1), & \text{else}. \end{cases} \tag{31}$$

Let now for all $(x_1,y_1) \in [e_1,1]^2$, $S_{\psi_e} = N \circ T_{\phi_e} \circ (N \times N)$. *Define* U, $(\mathcal{U}(x,y))_1$ *and* $(\mathcal{U}(x,y))_2$ *in a similar way as in Theorem 4.3. Then* $\mathcal{U}$ *is a uninorm on* $\mathcal{L}^I$ *with neutral element* e *for which* $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I}) = \alpha$ *and which is self-dual w.r.t.* $\mathcal{N}_N$.

The question remains whether for any $e \in D \setminus \{0_{\mathcal{L}^I},1_{\mathcal{L}^I}\}$, any $\alpha \in L^I$ such that $\alpha \|_{L^I} e$, and also any t-norm $\mathcal{T}$ and any t-conorm $\mathcal{S}$ on $\mathcal{L}^I$, there exists a uninorm $\mathcal{U}$ on $\mathcal{L}^I$ with neutral element e such that $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I}) = \alpha$, $\mathcal{T}_{\mathcal{U}} = \mathcal{T}$ and $\mathcal{S}_{\mathcal{U}} = \mathcal{S}$.

5 Conclusion

In this paper we have studied uninorms on the lattice $\mathcal{L}^I$, which is the underlying lattice of both Atanassov's intuitionistic fuzzy set theory and interval-valued fuzzy set theory. Such uninorms $\mathcal{U}$ can be neither conjunctive nor disjunctive, in which case $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I})$ is an element of L^I which is incomparable to the neutral element of $\mathcal{U}$. We have investigated the value $\mathcal{U}(x,y)$ in the case that x and y are located in certain areas of L^I and we have found several restrictions. For any value of $\alpha \in L^I$ which is incomparable to an arbitrary element e, we have constructed a uninorm $\mathcal{U}$ with neutral element e and for which $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I}) = \alpha$.

References

[1] M. B. Gorzałczany. A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 21(1):1–17, Jan 1987.

[2] R. Sambuc. *Fonctions Φ-floues. Application à l'aide au diagnostic en pathologie thyroidienne*. PhD thesis, Université de Marseille, France, 1975.

[3] D. Dubois. On ignorance and contradiction considered as truth-values. *Logic Journal of the IGPL*, 16(2):195–216, 2008.

[4] K. T. Atanassov. *Intuitionistic fuzzy sets*. Physica-Verlag, Heidelberg, New York, 1999.

[5] G. Deschrijver and E. E. Kerre. On the relationship between some extensions of fuzzy set theory. *Fuzzy Sets and Systems*, 133(2):227–235, 2003.

[6] J. A. Goguen. L-fuzzy sets. *Journal of Mathematical Analysis and Applications*, 18(1):145–174, Apr 1967.

[7] R. R. Yager and A. Rybalov. Uninorm aggregation operators. *Fuzzy Sets and Systems*, 80(1):111–120, May 1996.

[8] G. Deschrijver and E. E. Kerre. Uninorms in L^*-fuzzy set theory. *Fuzzy Sets and Systems*, 148(2):243–262, 2004.

[9] A. H. Clifford. Naturally totally ordered commutative semigroups. *Amer. J. Math.*, 76:631–646, 1954.

[10] S. Jenei. A note on the ordinal sum theorem and its consequence for the construction of triangular norms. *Fuzzy Sets and Systems*, 126(2):199–205, Mar 2002.

[11] C.-H. Ling. Representation of associative functions. *Publ. Math. Debrecen*, 12:189–212, 1965.

[12] S. Saminger. On ordinal sums of triangular norms on bounded lattices. *Fuzzy Sets and Systems*, 157(10):1403–1416, May 2006.

[13] B. Schweizer and A. Sklar. Associative functions and abstract semigroups. *Publ. Math. Debrecen*, 10:69–81, 1963.

[14] J. C. Fodor, R. R. Yager, and A. Rybalov. Structure of uninorms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5(4):411–427, Aug 1997.

[15] M. Eisenberg. *Axiomatic theory of sets and classes*. Holt, Rinehart and Winston, Inc., New York, 1971.

[16] B. Depaire, K. Vanhoof, and G. Wets. The application of uninorms in importance-performance analysis. In *FS'06: Proceedings of the 7th WSEAS International Conference on Fuzzy Systems*, pages 1–7, Stevens Point, Wisconsin, USA, 2006. World Scientific and Engineering Academy and Society (WSEAS).

[17] K. Vanhoof, P. Pauwels, J. Dombi, T. Brijs, and G. Wets. Penalty-reward analysis with uninorms: A study of customer (dis)satisfaction. In D. Ruan, C. Chen, E. E. Kerre, and G. Wets, editors, *Intelligent Data Mining: Techniques and Applications*, volume 5, pages 237–252. 2005.

[18] B. Depaire, K. Vanhoof, and G. Wets. Managerial opportunities of uninorm-based importance-performance analysis. *WSEAS Transactions on Business and Economics*, 3(3):101–108, 2007.

[19] B. Depaire, K. Vanhoof, and G. Wets. Expectation-performance compatibility in a customer satisfaction context modelled by means of aggregation operators. In B. De Baets, K. Maes, and R. Mesiar, editors, *Proceedings of the 4th International Summer School on Aggregation Operators*, pages 45–50. Academia Press, 2007.

ISBN: 978-989-95079-6-8

Fuzzy Descriptions Logics with Fuzzy Truth Values

Fernando Bobillo[1] Umberto Straccia[2]

1. Dpt. of Computer Science & Systems Engineering University of Zaragoza, Spain
2. ISTI - CNR, Pisa, Italy
Email: fbobillo@unizar.es, straccia@isti.cnr.it

***Abstract*— *Fuzzy Description Logics are a family of logics which allow to deal with structured knowledge affected by vagueness. Although a relatively important amount of work has been carried out in the last years, current fuzzy DLs are open to be extend with several features worked out in the fuzzy logic literature. In this work, we extend fuzzy DLs with fuzzy truth values, allowing to state sentences such as "Tina is young is* almost *true".**

***Keywords*—** Fuzzy Description Logics, Fuzzy Truth Values

1 Introduction

Description Logics (DLs) [1], have gained popularity due to their application in the context of the *Semantic Web* [2]. *Ontologies* play a key role in the Semantic Web. An ontology consists of a hierarchical description of important concepts in a particular domain, along with the description of the properties (of the instances) of each concept. DLs play a particular role in this context as they are essentially the theoretical counterpart of the *Web Ontology Language OWL DL*, a state of the art language to specify ontologies.

It is well-known that "classical" ontology languages are not appropriate to deal with *fuzzy/vague/imprecise knowledge*, which is inherent to several real world domains [3, 4]. Since fuzzy set theory and fuzzy logic are suitable formalisms to handle these types of knowledge, fuzzy ontologies emerge as useful in several applications, such as (multimedia) information retrieval, image interpretation, ontology mapping, matchmaking and the Semantic Web [5]. So far, several fuzzy extensions of DLs can be found in the literature (see the survey in [5]) and some fuzzy DL reasoners exist, such as FUZZYDL [6], DELOREAN [7], FIRE [8] or DLMEDIA [9].

In this paper we allow fuzzy DL sentences to be qualified with *fuzzy truth values* [10], and, thus, allow expressions such as "Tina is young is *very true*" and "Tina is young is *almost true*". We show the syntax, semantics and reasoning algorithms for the extension provided in this paper.

We proceed as follows. To make the paper self-contained, the next section recalls salient notions of mathematical fuzzy logics [11]. Section 3 introduces fuzzy $\mathcal{ALC}$ [3], which is the fuzzy variant of the Description Logic $\mathcal{ALC}$. $\mathcal{ALC}$ is usually considered as a reference language, whenever new features are introduced into DLs. Section 4 extends fuzzy DLs with fuzzy truth values and Section 5 provides reasoning algorithms. Section 6 concludes and describes future work.

2 Preliminaries: Mathematical Fuzzy Logic

In the setting of fuzzy logics, the convention prescribing that a statement is either true or false is changed and is a matter of degree measured on an ordered scale $\mathcal{S}$ that is no longer $\{0, 1\}$, but, e.g., the unit interval $[0, 1]$. This degree of fit is called *degree of truth* of the statement ϕ in the interpretation $\mathcal{I}$. Fuzzy logics logics provide compositional calculi of degrees of truth, including degrees between "true" and "false". A statement is now not true or false only, but may have a truth degree taken from a *truth space* $\mathcal{S}$, usually $[0, 1]$ (in that case we speak about *Mathematical Fuzzy Logic* [11]).

In the illustrative fuzzy logic that we consider in this section, *fuzzy statements* have the form $\phi \geqslant l$ or $\phi \leqslant u$, where $l, u \in [0, 1]$ [13, 11] and ϕ is a statement, which encode that the degree of truth of ϕ is *at least* l resp. *at most* u. For example, $ripe_tomato \geqslant 0.9$ says that we have a rather ripe tomato (the degree of truth of $ripe_tomato$ is at least 0.9).

A *fuzzy interpretation* $\mathcal{I}$ maps each basic statement p_i into $[0, 1]$ and is then extended inductively to all statements:

$$\begin{array}{lcl} \mathcal{I}(\phi \wedge \psi) & = & \mathcal{I}(\phi) \otimes \mathcal{I}(\psi)\,; \\ \mathcal{I}(\phi \vee \psi) & = & \mathcal{I}(\phi) \oplus \mathcal{I}(\psi)\,; \\ \mathcal{I}(\phi \rightarrow \psi) & = & \mathcal{I}(\phi) \Rightarrow \mathcal{I}(\psi)\,; \\ \mathcal{I}(\neg\phi) & = & \ominus \mathcal{I}(\phi)\,, \end{array}$$

where $\otimes$, $\oplus$, $\Rightarrow$, and $\ominus$ are so-called *truth combination functions*, namely, *triangular norms* (or *t-norms*), *triangular conorms* (*t-conorms*), *implication functions*, and *negation functions*, respectively, which extend the classical Boolean conjunction, disjunction, implication, and negation, respectively, to the fuzzy case (see [11] for a formal definition of these functions and their properties). Several t-norms, t-conorms, implication functions, and negation functions have been given in the literature, giving raise to different fuzzy logics with different logical properties. In fuzzy logic, one usually distinguishes three different logics (see Fig. 1), namely Łukasiewicz, Gödel, and Product logic [11]. Zadeh "logic" (fuzzy operators originally considered by Zadeh [12]) is entailed by Łukasiewicz logic.

A *fuzzy set* R over a countable crisp set X is a function $R\colon X \rightarrow [0, 1]$. A (binary) *fuzzy relation* R over two countable crisp sets X and Y is a function $R\colon X \times Y \rightarrow [0, 1]$. The *degree of subsumption* between two fuzzy sets A and B is defined as $\inf_{x \in X} A(x) \Rightarrow B(x)$.

The notions of satisfiability and logical consequence are defined in the standard way. A fuzzy interpretation $\mathcal{I}$ *satisfies* a fuzzy statement $\phi \geqslant l$ (resp., $\phi \leqslant u$) or $\mathcal{I}$ is a *model* of $\phi \geqslant l$ (resp., $\phi \leqslant u$), denoted $\mathcal{I} \models \phi \geqslant l$ (resp., $\mathcal{I} \models \phi \leqslant u$), iff $\mathcal{I}(\phi) \geqslant l$ (resp., $\mathcal{I}(\phi) \leqslant u$). $\phi \geqslant l$ is a *tight logical consequence* of a set of fuzzy statements $\mathcal{K}$ iff l is the infimum of $\mathcal{I}(\phi)$ subject to all models $\mathcal{I}$ of $\mathcal{K}$. The latter value is equivalent to $l = \sup\{r \mid \mathcal{K} \models \phi \geqslant r\}$, it is called *Best Entailment Degree* (BED), and is denoted $bed(\mathcal{K}, \phi)$, while the *Best Satisfiability Degree*, denoted as $bsd(\mathcal{K}, \phi)$ is defined as $\sup_{\mathcal{I} \models \mathcal{K}} \mathcal{I}(\phi)$. We refer the reader to [11, 13, 14] for reasoning algorithms for fuzzy propositional and First-Order Logics (FOLs).

ISBN: 978-989-95079-6-8

	Łukasiewicz Logic	Gödel Logic	Product Logic	Zadeh Logic
$a \otimes b$	$\max(a+b-1,0)$	$\min(a,b)$	$a \cdot b$	$\min(a,b)$
$a \oplus b$	$\min(a+b,1)$	$\max(a,b)$	$a+b-a \cdot b$	$\max(a,b)$
$a \Rightarrow b$	$\min(1-a+b,1)$	$\begin{cases} 1 & \text{if } a \leq b \\ b & \text{otherwise} \end{cases}$	$\min(1,b/a)$	$\max(1-a,b)$
$\ominus a$	$1-a$	$\begin{cases} 1 & \text{if } a=0 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 & \text{if } a=0 \\ 0 & \text{otherwise} \end{cases}$	$1-a$

Figure 1: Combination functions of various fuzzy logics.

For illustrative purposes, we provide a simple and effective way to solve the entailment problem, in the case of Zadeh logic and Lukasiewicz logic, in terms of Mixed Integer Linear Programming (MILP) –see also [14]. The calculus depends on the t-norm, t-conorm and negation functions considered. Suppose we are looking for $bed(\mathcal{K},\phi)$, then,

$$bed(\mathcal{K},\phi) = \min\{x \mid \mathcal{K} \cup \{\phi \leq x\} \text{ is satisfiable.}\} \quad (1)$$

Indeed, suppose the minimal value is $\bar{n}$. We will know then that for any interpretation $\mathcal{I}$ satisfying $\mathcal{K}$, it cannot be $\mathcal{I}(\phi) < \bar{n}$, and, thus, $\mathcal{I}(\phi) \geq \bar{n}$ has to hold.

The above problem can be solved by means of MILP (we consider Zadeh logic only, for Lukasiewicz logic the procedure is similar) [1]. For a formula ϕ consider a variable x_ϕ. The intuition is that the degree of truth of ϕ is greater or equal to x_ϕ. The MILP problem determining $bed(\mathcal{K},\phi)$ is as follows:

$$\begin{array}{l} \min x. \text{ such that } x \in [0,1] \\ \quad x_{\neg\phi} \geq 1-x, \sigma(\neg\phi), \\ \quad \text{for all } (\phi' \geq n) \in \mathcal{K}, x_{\phi'} \geq n, \sigma(\phi'), \\ \quad \text{for all } (\phi' \leq n) \in \mathcal{K}, x_{\neg\phi'} \geq 1-n, \sigma(\neg\phi') \end{array} \quad (2)$$

where the function σ, transforming a many-valued proposition into a set of inequations, is inductively defined as follows:

$$\sigma(\phi) = \begin{cases} x_p \in [0,1] & \text{if } \phi = p \\ x_{\phi'} = 1 - x_\phi, x_\phi \in [0,1] & \text{if } \phi = \neg\phi' \\ x_{\phi_1} \geq x_\phi, x_{\phi_2} \geq x_\phi, \\ \sigma(\phi_1), \sigma(\phi_2), x_\phi \in [0,1] & \text{if } \phi = \phi_1 \wedge \phi_2 \\ x_{\phi_1} + x_{\phi_2} = x_\phi, x_{\phi_1} \leq y, \\ x_{\phi_2} \leq 1-y, \sigma(\phi_1), \sigma(\phi_2), \\ y \in \{0,1\}, x_\phi \in [0,1], \\ \text{where } y \text{ is a new binary variable} & \text{if } \phi = \phi_1 \vee \phi_2 \end{cases}.$$

In a similar way, we may determine $bsd(\mathcal{K},\phi)$ as

$$\begin{array}{l} \min -x. \text{ such that } x \in [0,1] \\ \quad x_\phi \geq x, \sigma(\phi), \\ \quad \text{for all } (\phi' \geq n) \in \mathcal{K}, x_{\phi'} \geq n, \sigma(\phi'), \\ \quad \text{for all } (\phi' \leq n) \in \mathcal{K}, x_{\neg\phi'} \geq 1-n, \sigma(\neg\phi') \end{array} \quad (3)$$

Note that we may verify whether $\mathcal{K}$ is satisfiable by checking if $bed(\mathcal{K},p) = 1$, where p is a new propositional letter not occurring in $\mathcal{K}$, and that under Łukasiewicz logic and Zadeh semantics we end up with a *bounded Mixed Integer Linear Program* (bMILP) optimization problem [15].

3 Fuzzy $\mathcal{ALC}$

Syntax. Consider an alphabet of *concepts names* (denoted A), *abstract roles names* (denoted R), *abstract individual names* (denoted a). From a First-Order Logic point of view, concepts may be seen as a formulae with one free variable (and, thus, may be seen as class descriptors), while roles as binary predicates (and, thus, may be used to describe properties of a class). *Concepts* (denoted C or D) of the language can be built inductively from atomic concepts (A), top concept $\top$, bottom concept $\bot$, abstract roles (R) as:

$C, D \rightarrow$	A	$\mid$	(atomic concept)
	$\top$	$\mid$	(top concept)
	$\bot$	$\mid$	(bottom concept)
	$C \sqcap D$	$\mid$	(concept conjunction)
	$C \sqcup D$	$\mid$	(concept disjunction)
	$\neg C$	$\mid$	(concept negation)
	$\forall R.C$	$\mid$	(universal quantification)
	$\exists R.C$		(existential quantification)

As illustrative purpose, Fig. 2 provides a First-Order Logic translation of $\mathcal{ALC}$ concepts and examples.

Syntax			*FOL*	*Example*
C, D	$\rightarrow$	$\top$	$\top(x)$	
		$\bot$	$\bot(x)$	
		A	$A(x)$	$Human$
		$C \sqcap D$	$C(x) \wedge D(x)$	$Human \sqcap Male$
		$C \sqcup D$	$C(x) \vee D(x)$	$Nice \sqcup Rich$
		$\neg C$	$\neg C(x)$	$\neg Meat$
		$\exists R.C$	$\exists y.R(x,y) \wedge C(y)$	$\exists has_child.Blond$
		$\forall R.C$	$\forall y.R(x,y) \rightarrow C(y)$	$\forall has_child.Human$

Figure 2: $\mathcal{ALC}$ concepts and First-Order Logic reading.

A *Fuzzy Knowledge Base* $\mathcal{K}$ comprises a fuzzy ABox $\mathcal{A}$ and a fuzzy TBox $\mathcal{T}$. A *fuzzy ABox* consists of a finite set of *fuzzy assertions* of one of the following types: a fuzzy *concept assertion* of the form $\langle a{:}C, n\rangle$ (with informal meaning, individual a is an instance of concept C with degree at least n) or a fuzzy *role assertion* of the form $\langle (a,b){:}R, n\rangle$ (the pair of individuals (a,b) is an instance of role R with degree at least n). In FOL, $\langle a{:}C, n\rangle$ may be seen as a fuzzy statement of the form $C(a) \geqslant n$, while $\langle (a,b){:}R, n\rangle$ may be seen as a fuzzy statement of the form $R(a,b) \geqslant n$.

In general, a *fuzzy TBox* $\mathcal{T}$ is a finite set of *fuzzy concept inclusion axioms* $\langle C \sqsubseteq D, n\rangle$, where C, D are concepts and $n \in (0,1]$. Informally, $\langle C \sqsubseteq D, n\rangle$ states that all instances of concept C are instances of concept D to degree n.$C = D$ is a shorthand for the two axioms $\langle C \sqsubseteq D, 1\rangle$ and $\langle D \sqsubseteq C, 1\rangle$.

However, for computational reasons, we will restrict TBoxes to be *acyclic*. That is, $\mathcal{T}$ is a finite set of *fuzzy concept inclusion axioms* $\langle A \sqsubseteq C, n\rangle$, and *concept definitions* $A = C$, where A is an atomic concept. Furthermore, we assume that $\mathcal{T}$ verifies two additional constraints: *(i)* there is no concept A such that it appears more than once on the left hand side of some axiom in $\mathcal{T}$. *(ii)* no cyclic definitions are present in $\mathcal{T}$. We will say that A *directly uses* primitive concept B in $\mathcal{T}$, if there is some axiom $\tau \in \mathcal{T}$ such that A is on the left hand side of τ and B occurs in the right hand side of τ. Let *uses* be the transitive closure of the relation *directly uses* in $\mathcal{T}$. $\mathcal{T}$ is cyclic iff there is A such that A uses A in $\mathcal{T}$.

[1] Note that this is an optimized version w.r.t. [13, 14].

ISBN: 978-989-95079-6-8

It is well known that such TBoxes can be eliminated through a finite (although it can create an exponential growth of the KB), *expansion process*, both in the crisp and in the fuzzy case [3]. Instead, we will use a extension to the fuzzy case of the *lazy expansion* technique [16], which has proved to be more useful in practice.

Semantics. From a semantics point of view, a fuzzy interpretation $\mathcal{I}$ is a pair $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consisting of a non empty set $\Delta^{\mathcal{I}}$ (the interpretation domain) and a fuzzy interpretation function $\cdot^{\mathcal{I}}$ mapping: (i) an *abstract individual* a onto an element $a^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ such that if $a \neq b$ then $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ (Unique Name Assumption); (ii) a *concept* C onto a function $C^{\mathcal{I}} : \Delta^{\mathcal{I}} \to [0,1]$; (iii) an *abstract role* R onto a function $R^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \to [0,1]$.

Given arbitraries t-norm $\otimes$, t-conorm $\oplus$, negation function $\ominus$ and implication function $\Rightarrow$, the fuzzy interpretation function is extended to *complex concepts and roles* as shown in Fig. 3, where $C^{\mathcal{I}}$ denotes the membership function of the fuzzy concept C with respect to the fuzzy interpretation $\mathcal{I}$. $C^{\mathcal{I}}(x)$ gives us the degree of being the individual x an element of the fuzzy concept C under $\mathcal{I}$. Similarly, $R^{\mathcal{I}}$ denotes the membership function of the fuzzy role R with respect to $\mathcal{I}$. $R^{\mathcal{I}}(x,y)$ gives us the degree of being (x,y) an element of the fuzzy role R under $\mathcal{I}$. The fuzzy interpretation function is extended to *fuzzy axioms* in Fig. 3.

Concept		Semantics
$(\top)^{\mathcal{I}}(x)$	=	1
$(\bot)^{\mathcal{I}}(x)$	=	0
$(A)^{\mathcal{I}}(x)$	=	$A^{\mathcal{I}}(x)$
$(C \sqcap D)^{\mathcal{I}}(x)$	=	$C^{\mathcal{I}}(x) \otimes D^{\mathcal{I}}(x)$
$(C \sqcup D)^{\mathcal{I}}(x)$	=	$C^{\mathcal{I}}(x) \oplus D^{\mathcal{I}}(x)$
$(\neg C)^{\mathcal{I}}(x)$	=	$\ominus C^{\mathcal{I}}(x)$
$(\forall R.C)^{\mathcal{I}}(x)$	=	$\inf_{y \in \Delta^{\mathcal{I}}} \{R^{\mathcal{I}}(x,y) \Rightarrow C^{\mathcal{I}}(y)\}$
$(\exists R.C)^{\mathcal{I}}(x)$	=	$\sup_{y \in \Delta^{\mathcal{I}}} \{R^{\mathcal{I}}(x,y) \otimes C^{\mathcal{I}}(y)\}$
Axiom		**Semantics**
$(a:C)^{\mathcal{I}}$	=	$C^{\mathcal{I}}(a^{\mathcal{I}})$
$((a,b):R)^{\mathcal{I}}$	=	$R^{\mathcal{I}}(a^{\mathcal{I}}, b^{\mathcal{I}})$
$(C \sqsubseteq D)^{\mathcal{I}}$	=	$\inf_{x \in \Delta^{\mathcal{I}}} \{C^{\mathcal{I}}(x) \Rightarrow D^{\mathcal{I}}(x)\}$

Figure 3: Semantics of fuzzy concepts and axioms.

A fuzzy interpretation $\mathcal{I}$ *satisfies* (is a *model* of) a fuzzy statement $\langle \alpha, n \rangle$ iff $\alpha^{\mathcal{I}} \geq n$. The notions of logical consequence, best entailment degree and best satisfiability degree of α are as for Section 2. We additionally define the *Best Satisfiability Degree* [6] of a concept C w.r.t. a fuzzy KB $\mathcal{K}$ as $bsd(\mathcal{K}, C) = \sup_{\mathcal{I} \models \mathcal{K}} \sup_{x \in \Delta^{\mathcal{I}}} C^{\mathcal{I}}(x)$.

4 Fuzzy DLs with fuzzy truth values

So far, we have seen that fuzzy statements are of the form $\langle \alpha, n \rangle$, where $n \in [0,1]$. As next, we extend such statements to allow fuzzy truth values, such as "very true, almost true, almost false", in place of a "crisp" value $n \in [0,1]$. Essentially, we have fuzzy truth-qualified statements in which the truth is now a fuzzy set and, thus, we allow statements such as "Tina is young is very true".

As pointed out in [17] (see also [4, 6, 18]), there are many functions to specify fuzzy set membership degrees in fuzzy set theory and practice. The most frequently used are the trapezoidal (Fig. 4 (a)), the triangular (Fig. 4 (b)), the L-function (left-shoulder function, Fig. 4 (c)), the R-function (right-shoulder function, Fig. 4 (d)) and linear hedges (Fig. 4 (e), where $a = c/(c+1), b = 1/(c+1)$). We will call these functions (defined over $[0,1]$) *truth qualifiers* and allow them to be used to modify the degree of truth of a sentence. We will use

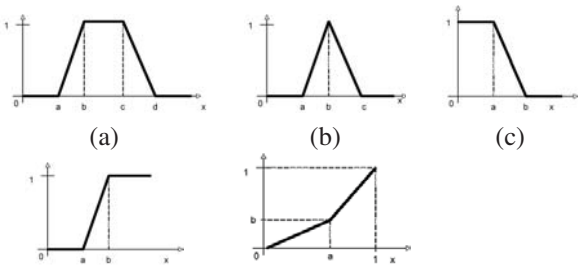


Figure 4: (a) Trapezoidal function; (b) Triangular function; (c) L-function; (d) R-function; (e) Linear function.

the abbreviations $\mathsf{trpz(a,b,c,d)}, \mathsf{tri(a,b,c)}, \mathsf{ls(a,b)}, \mathsf{rs(a,b)}$ and $\mathsf{ln(c)}$ to denote them, respectively.

Syntax. Let α be a concept assertion $a{:}C$, a role assertion $(a,b){:}R$ or a GCI $C \sqsubseteq D$, and $n \in [0,1]$. Let q be a truth qualifier, as in Fig. 4. Then we extend $\mathcal{ALC}$ axioms $\langle \alpha, n \rangle$ also to be of the form $\langle \alpha, q \rangle$, where a truth qualifier q may occur in place of a value $n \in [0,1]$. For instance, "Tina is young is very true" may be represented with $\langle \mathsf{tina{:}Young}, \mathsf{ln(4)} \rangle$. Similarly, we may represent an imprecision about the actual degree of truth, such as, "Tina is young is true to degree around 0.7", as an axioms of the form $\langle \mathsf{tina{:}Young}, \mathsf{tri(0.6, 0.7, 0.8)} \rangle$.

In the following, a knowledge base will be split into two parts, $\mathcal{K} = \mathcal{K}_U \cup \mathcal{K}_Q$ in which $\mathcal{K}_U$ contains only unqualified fuzzy axioms of type $\langle \alpha, n \rangle$, where $n \in [0,1]$, while $\mathcal{K}_Q$ contains only qualified fuzzy axioms of type $\langle \alpha, q \rangle$, where q is a truth qualifier. In FOL, we may see $\mathcal{K}$ as the formula

$$\Gamma_{\mathcal{K}} = \bigwedge_{\tau_i \in \mathcal{K}_U} \tau_i \wedge \bigwedge_{\tau_j \in \mathcal{K}_Q} \tau_j \,. \tag{4}$$

We will also use Γ_U for the left conjunct, while use Γ_Q for the right conjunct and, thus, $\Gamma_{\mathcal{K}} = \Gamma_U \wedge \Gamma_Q$.

Semantics. So far, given an interpretation $\mathcal{I}$, an axiom of the form $\langle \alpha, n \rangle$ is crisp under $\mathcal{I}$, in the sense that either $\mathcal{I}$ satisfies $\langle \alpha, n \rangle$ (i.e., $\alpha^{\mathcal{I}} \geq n$) or $\mathcal{I}$ does not satisfy $\langle \alpha, n \rangle$ (i.e., $\alpha^{\mathcal{I}} < n$). Once we move to truth qualified axioms, axioms are no longer true or false, but have a degree of truth depending on the qualifier. More precisely, the *degree of truth* of a fuzzy axiom τ of the form $\langle \alpha, q \rangle$ under $\mathcal{I}$, denoted $\mathcal{I}(\tau)$ is defined as the value $q(\alpha^{\mathcal{I}})$ (the application of the qualifier q to the truth value $\alpha^{\mathcal{I}}$). So, if "Tina is Young" is true to degree 0.9 under $\mathcal{I}$, then "Tina is young is very true" to degree 0.6 under $\mathcal{I}$. We may also extend this notion the axioms of the form $\langle \alpha, n \rangle$ by defining the *degree of truth* of a fuzzy axiom τ of the form $\langle \alpha, n \rangle$ to be $\mathcal{I}(\tau) = 1$ if $\alpha^{\mathcal{I}} \geq n$, otherwise $\mathcal{I}(\tau) = 0$, which is compatible with what we have defined so far.

Now, consider a knowledge base $\mathcal{K} = \mathcal{K}_U \cup \mathcal{K}_Q$, a fuzzy axiom τ and an interpretation $\mathcal{I}$. The *degree of truth* of $\mathcal{K}_x$ ($x \in \{U, Q\}$) under $\mathcal{I}$, denoted $\mathcal{I}(\mathcal{K}_x)$ is defined as

$$\mathcal{I}(\mathcal{K}_x) = \bigotimes_{\tau_i \in \mathcal{K}_x} \mathcal{I}(\tau_i) \,.$$

In FOL, this is the same as $\mathcal{I}(\Gamma_x)$. Note that if $\mathcal{I}$ is a model of $\mathcal{K}_U$ then $\mathcal{I}(\mathcal{K}_U) = 1$, else $\mathcal{I}(\mathcal{K}_U) = 0$. Furthermore, the *degree of entailment* of a τ w.r.t. $\mathcal{K}$ under $\mathcal{I}$, denoted $\mathcal{I}(\mathcal{K}, \tau)$, is defined as

$$\mathcal{I}(\mathcal{K}, \tau) = (\mathcal{I}(\mathcal{K}_U) \otimes \mathcal{I}(\mathcal{K}_Q)) \Rightarrow \mathcal{I}(\tau) \,. \tag{5}$$

ISBN: 978-989-95079-6-8

Essentially, the degree of entailment under $\mathcal{I}$ is the evaluation under $\mathcal{I}$ of the FOL implication $(\Gamma_U \wedge \Gamma_Q) \rightarrow \tau$, which is quite natural. Please note that if $\mathcal{I} \not\models \mathcal{K}_U$ then $\mathcal{I}(\mathcal{K}_U) = 0$ and, thus, $\mathcal{I}(\Gamma_\mathcal{K} \rightarrow \tau) = 1$, while if $\mathcal{I} \models \mathcal{K}_U$ then $\mathcal{I}(\Gamma_\mathcal{K} \rightarrow \tau) = \mathcal{I}(\Gamma_Q \rightarrow \tau)$. The *Best Entailment Degree* of a τ w.r.t. $\mathcal{K}$, denoted $bed(\mathcal{K}, \tau)$ is defined as $bed(\mathcal{K}, \tau) = \inf_\mathcal{I} \mathcal{I}(\mathcal{K}, \tau)$. By the previous observations, it easily follows that

$$bed(\mathcal{K}, \tau) = \inf_{\mathcal{I} \models \mathcal{K}_U} \mathcal{I}(\Gamma_Q \rightarrow \tau) \,, \tag{6}$$

where $\inf \emptyset = 1$. Finally, for an axiom α of the form $a{:}C$, $(a,b){:}R$ or a $C \sqsubseteq D$, the *Best Entailment Degree* of α w.r.t. $\mathcal{K}$, denoted $bed(\mathcal{K}, \alpha)$ is defined as

$$bed(\mathcal{K}, \alpha) = \sup\{n \mid 1 = bed(\mathcal{K}, \langle \alpha, n \rangle)\} \,, \tag{7}$$

and the *Best Satisfiability Degree* of a C w.r.t. $\mathcal{K}$, is

$$bsd(\mathcal{K}, C) = \sup_{\mathcal{I} \models \mathcal{K}_U} \sup_{x \in \Delta^\mathcal{I}} \mathcal{I}(\Gamma_Q) \Rightarrow C^\mathcal{I}(x) \,.$$

5 Reasoning

We next provide a reasoning algorithm for fuzzy $\mathcal{ALC}$ with truth qualifiers. To start with, we require a calculus for fuzzy $\mathcal{ALC}$ without truth qualifiers. Our algorithm is inspired by the one implemented within the FUZZYDL system [6] (which follows from the one presented in [17]). However, the presence of truth qualifiers requires some modifications. Furthermore, the version we present here for $\mathcal{ALC}$, introduces some optimizations that will require less "variables" and, thus, is expected to be more efficient. For the sake of this paper, *we provide a calculus under Łukasiewicz logic*, as in [17, 19]. A similar algorithm under product logic is expected to be more involved (see, e.g., [20, 21]).

W.l.o.g., we may assume that concepts are in *Negation Normal Form* (NNF), which is obtained by pushing in the usual manner negation on front of concept names only, by applying recursively the equivalences below.

$\neg\top \equiv \bot$	$\neg(C_1 \sqcup C_2) \equiv \neg C_1 \sqcap \neg C_2$	$\neg\exists R.C \equiv \forall R.\neg C$
$C \sqcap \bot \equiv \bot$	$C \sqcup \bot \equiv C$	$\neg\neg C \equiv C$
$\neg\bot \equiv \top$	$\neg(C_1 \sqcap C_2) \equiv \neg C_1 \sqcup \neg C_2$	$\neg\forall R.C \equiv \exists R.\neg C$
$C \sqcap \top \equiv C$	$C \sqcup \top \equiv \top$	

Now, our goal is to provide a terminating procedure determining $bed(\mathcal{K}, \alpha)$, where α is of the form $a{:}C$, $(a,b){:}R$ or a $C \sqsubseteq D$, and $bed(\mathcal{K}, \tau,)$, for truth qualified fuzzy axioms τ of the form $\langle \alpha, q \rangle$ (q is a truth qualifier).

5.1 Reasoning in $\mathcal{ALC}$ without truth qualifiers

Let's focus first on $bed(\mathcal{K}, \alpha)$, where $\mathcal{K}$ *does not* contain any truth qualifier, i.e. $\mathcal{K}_Q = \emptyset$. If α is a fuzzy role assertion of the form $(a,b){:}R$ then, in order to determine $bed(\mathcal{K}, (a,b){:}R)$, we may reduce it to the BED for concept assertions, as

$$bed(\mathcal{K}, (a,b){:}R) = bed(\mathcal{K} \cup \{\langle b{:}B, 1\rangle\}, a{:}\exists R.B) \,,$$

where B is a new concept (i.e., it does not occur in $\mathcal{K}$).

Consider $\mathcal{K} = \langle \mathcal{A}, \mathcal{T} \rangle$, where $\mathcal{T}$ is acyclic. The basic idea behind our reasoning algorithm is based on the observations and algorithm of Section 2. So, not surprisingly, in order to determine $bed(\mathcal{K}, \alpha)$, we will compute it as

$$bed(\mathcal{K}, \alpha) = \min x. \text{ such that } \mathcal{K} \cup \{(\alpha \leq x)\} \text{ satisfiable.} \tag{8}$$

Then by applying satisfiability preserving rules, we generate new *inequations* over $[0,1]$-valued variables. These inequations have to hold in order to respect the semantics of the DL constructors. Finally, in order to determine the BED, we *minimize* the original variable x such that all constraints are satisfied. More specifically,

$$\begin{array}{lcl} bed(\mathcal{K}, a{:}C) & = & \min x \text{ such that } \mathcal{K} \cup \{\langle a{:}\neg C, 1-x\rangle\} \text{ satisfiable} \\ bed(\mathcal{K}, C \sqsubseteq D) & = & \min x \text{ such that } \mathcal{K} \cup \{\langle b{:}C \sqcap \neg D, 1-x\rangle\} \text{ satisfiable} \\ bsd(\mathcal{K}, C) & = & \min -x \text{ such that } \mathcal{K} \cup \{\langle b{:}C, x\rangle\} \text{ satisfiable} \,, \end{array}$$

where b is a new individual. This means that determining the minimum degree of satisfiability [2] of a KB is the main reasoning issue to be addressed. Note also that we may determine if $\mathcal{K}$ has a model by e.g. checking whether $bed(\mathcal{K}, b{:}A) = 1$, where b and A do not occur in $\mathcal{K}$.

Like most of the tableaux algorithms, our algorithm works on *completion-forests* since an ABox might contain several individuals with arbitrary roles connecting them. A completion-forest $\mathcal{F}$ is a collection of trees whose distinguished roots are arbitrarily connected by edges. The forest has associated a set $\mathcal{C}_\mathcal{F}$ of constraints of the form $l \leq l'$, $l = l'$, $x_i \in [0,1], y_i \in \{0,1\}$, where l, l' are arithmetic expressions, on the variables occurring the node labels and edge labels.

Each node v is labelled with a set $\mathcal{L}(v)$ of concepts C. If $C \in \mathcal{L}(v)$ then we consider a variable $x_{v:C}$. The intuition here is that v is an instance of C to degree equal or greater than of the value of the variable $x_{v:C}$ in a minimal solution. Essentially, $x_{v:C}$ will hold the degree of truth of $v{:}C$. Each edge $\langle v, w\rangle$ is labelled with a set $\mathcal{L}(\langle v, w\rangle)$ of roles R. If $R \in \mathcal{L}(\langle v, w\rangle)$ then we consider a variable $x_{(v,w):R}$ (the intuition here is that $\langle v, w\rangle$ is an instance of R to degree equal or greater than the value of the variable $x_{(v,w):R}$ in a minimal solution (as for concept assertions, $x_{(v,w):R}$ will hold the degree of truth of $(v,w){:}R$). We will assume that there is a bijection between assertions α and variables x_α.

We are ready now to present our calculus. We first start with an initialization step, which builds the starting completion-forest. Then we apply to it a set of completion-forest transforming rules until no more rule can be applied. Finally, we solve the MILP problem associated to the set of constraints.

The algorithm initializes a forest $\mathcal{F}$ as follows. Consider $\mathcal{K}$: (i) $\mathcal{F}$ contains a root node a_i for each individual a_i occurring in $\mathcal{A}$; (ii) $\mathcal{F}$ contains an edge $\langle a, b\rangle$ for each fuzzy assertion $\langle (a,b){:}R, n\rangle \in \mathcal{A}$; (iii) for each fuzzy assertion $\langle a{:}C, n\rangle \in \mathcal{A}$, add both C to $\mathcal{L}(a)$ and $x_{a:C} \geq n$ to the set of constraints $\mathcal{C}_\mathcal{F}$; (iv) for each fuzzy assertion $\langle (a,b){:}R, n\rangle \in \mathcal{A}$, add both R to $\mathcal{L}(\langle a, b\rangle)$ and $x_{(a,b):R} \geq n$ to the set of constraints $\mathcal{C}_\mathcal{F}$; (v) for all introduced variables x_α, add $x_\alpha \in [0,1]$ to $\mathcal{C}_\mathcal{F}$.

$\mathcal{F}$ is then expanded by repeatedly applying the completion rules described below. The completion-forest is complete when none of the completion rules are applicable. Then, the bMILP problem on the set of constraints $\mathcal{C}_\mathcal{F}$ is solved.

As anticipated, we will use a extension to the fuzzy case of the *lazy expansion* technique in order to remove the axioms in $\mathcal{T}$. The basic idea is as follows (recall that there are only two types of fuzzy concept inclusions): given $\langle A \sqsubseteq C, n\rangle$, add C only to nodes with a label containing A, and given $\langle C \sqsubseteq A, 1\rangle$, add $\neg C$ only to nodes with a label containing $\neg A$.

[2] In the last equation, we use the fact that in our setting $\max x \; s.t. \; x \in S = \min -x \; s.t. \; x \in S$.

ISBN: 978-989-95079-6-8

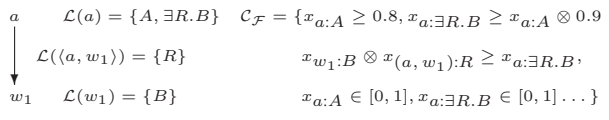


Figure 5: Complete forest for Example 1.

We assume a fixed rule application strategy as e.g., the order of rules below, such that the rule for $(\exists)$ is applied as last. Also, all expressions in node labels are processed according to the order they are introduced into $\mathcal{F}$. Note that we do not need a notion of *blocking* as $\mathcal{T}$ is acyclic.

Now we are ready to present the inference rules:

$(\bar{A})$. If $\neg A \in \mathcal{L}(v)$ then $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:A} = 1 - x_{v:\neg A}\} \cup \{x_{v:A} \in [0,1]\}$.

$(\bot)$. If $\bot \in \mathcal{L}(v)$ then $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:\bot} = 0\}$.

$(\top)$. If $\top \in \mathcal{L}(v)$ then $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:\top} = 1\}$.

$(\sqcap)$. If *(i)* $C \sqcap D \in \mathcal{L}(v)$, and *(ii)* not both $C \in \mathcal{L}(v)$ and $D \in \mathcal{L}(v)$ then add C and D to $\mathcal{L}(v)$, and $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:C} \otimes x_{v:D} = x_{a:C \sqcap D}\} \cup \{x_{v:C}, x_{v:D} \in [0,1]\}$.

$(\sqcup)$. If *(i)* $C \sqcup D \in \mathcal{L}(v)$, and *(ii)* not both $C \in \mathcal{L}(v)$ and $D \in \mathcal{L}(v)$ then add C and D to $\mathcal{L}(v)$, and $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:C} \oplus x_{v:D} = x_{a:C \sqcup D}\} \cup \{x_{v:C}, x_{v:D} \in [0,1]\}$.

$(\forall)$. If *(i)* $\forall R.C \in \mathcal{L}(v)$, $R \in \mathcal{L}(\langle v, w\rangle)$, and *(ii)* the rule has not been already applied to this pair then add C to $\mathcal{L}(w)$, and $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{w:C} \geq x_{a:\forall R.C} \otimes x_{(v,w):R}\} \cup \{x_{w:C} \in [0,1]\}$.

$(\sqsubseteq)$. If *(i)* $\langle A \sqsubseteq C, n\rangle \in \mathcal{T}$, *(ii)* $A \in \mathcal{L}(v)$, and *(iii)* v is a node to which this rule has not yet been applied then *(i)* append C to $\mathcal{L}(v)$, and *(ii)* $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:C} \geq x_{v:A} \otimes n\} \cup \{x_{v:C} \in [0,1]\}$.

$(\bar{\sqsubseteq})$. If *(i)* $\langle C \sqsubseteq A, 1\rangle \in \mathcal{T}$, *(ii)* $\neg A$, and *(iii)* v is a node to which this rule has not yet been applied then *(i)* append $\neg C$ to $\mathcal{L}(v)$, and *(ii)* $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:\neg A} \geq x_{v:\neg C}\} \cup \{x_{v:\neg C} \in [0,1]\}$.

$(\exists)$. If $\exists R.C \in \mathcal{L}(v)$ then create a new node w, add R to $\mathcal{L}(\langle v, w\rangle)$, add C to $\mathcal{L}(w)$, and $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{w:C} \otimes x_{(v,w):R} = x_{v:\exists R.C}\} \cup \{x_{(v,w):R}, x_{w:C} \in [0,1]\}$.

Note that in order to write the fuzzy operators, we may need to create some new control variables. For example, under Łukasiewicz t-norm, $x_1 \otimes x_2 \geq l$ can be written as $\{l \leq y, x_1 + x_2 \leq 1 + l, x_1 + x_2 - l \geq y, y_i \in \{0,1\}\}$. If $y = 0$, then $l = 0$ (it simulates the case where $x_1 + x_2 \leq 1$, and hence $x_1 \otimes x_2 = 0$), and if $y = 1$, then $l = x_1 + x_2 - 1$ [17, 19].

Example 1 illustrates the behaviour of the algorithm.

Example 1 *Consider* $\mathcal{K} = \{\langle A \sqsubseteq \exists R.B, 0.9\rangle, \langle a{:}A, 0.8\rangle\}$. *Let's show that* $\mathcal{K}$ *is satisfiable. The forest* $\mathcal{F}$ *is initialized with a root node* a*, labelled with* $\mathcal{L}(a) = \{A\}$*, and set of constraints* $\mathcal{C}_\mathcal{F} = \{x_{a:A} \geq 0.8\}$*. The we apply the* $(\sqsubseteq)$ *rule and add* A *and* $\exists R.B$ *to* $\mathcal{L}(a)$*, and add to the constraint set* $\{x_{a:\exists R.B} \geq x_{a:A} \otimes 0.9\} \cup \{x_{a:A}, x_{a:\exists R.B} \in [0,1]\}$*. As next, we apply the* $(\exists)$ *rule to node* a *and, thus, we create a new node* w_1*, labelled with* $\mathcal{L}(w_1) = \{B\}$*, and an edge* $\langle a, w_1\rangle$ *labelled with* $\mathcal{L}(\langle a, w_1\rangle) = \{R\}$*, and we update the constraint set with* $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{w_1:B} \otimes x_{(a,w_1):R} = x_{a:\exists R.B}\} \cup \{x_{(a,w_1):R}, x_{w_1:B} \in [0,1]\}$*. Rule* $(\sqsubseteq)$ *is not applicable to node* w_1 *because* $A \notin \mathcal{L}(b)$*.*

The complete forest $\mathcal{F}$ *in Figure 5 shows the computation so far. It only remains to find a solution to the inequalities.* □

Note that there is a significative difference with other similar algorithms for fuzzy DLs combining tableau algorithms with optimization problems [17, 20, 9]. In those algorithms, every time a concept C appears in the list of expressions of a node v, a new variable x is created. Instead, we introduce a variable $x_{v:C}$ once, and reuse it the following times. This reduction in the number of generated variables is important because it makes the bMILP problem easier to solve. We have:

Proposition 1 *For each KB* $\mathcal{K}$*, the tableau algorithm terminates and computes* $bed(\mathcal{K}, \alpha)$ *and* $bsd(\mathcal{K}, C)$*.*

5.2 Reasoning in $\mathcal{ALC}$ with truth qualifiers

Let us now consider the case $\mathcal{K}_Q \neq \emptyset$. We next address the problem of determining $bed(\mathcal{K}, \tau)$ and $bed(\mathcal{K}, \alpha)$. From Eq. (6), we now that if $\mathcal{K}_U$ has no model then immediately $bed(\mathcal{K}, \tau) = 1$ and, thus, this case is not of particular interest. So, let us assume that $\mathcal{K}_U$ has a model. The case $\tau = \langle \alpha, n\rangle$, with $n \in [0,1]$, is also not of particular interest as it suffices to compute $m = bed(\mathcal{K}, \alpha)$ and check whether $m \geq n$. So, it remains to determine both $bed(\mathcal{K}, \tau)$, for a fuzzy truth qualified statement τ, and $bed(\mathcal{K}, \alpha)$, where α is of the form $a{:}C$, $(a,b){:}R$ or a $C \sqsubseteq D$. From what we have seen so far, it is immediate that (where a is new individual [3])

$$bed(\mathcal{K}, \tau) = \min x. \text{ such that } \mathcal{K}_U \cup \{(\Gamma_Q \rightarrow \tau) \leq x\} \text{ satisfiable.} \quad (9)$$

$$bsd(\mathcal{K}, C) = \min -x. \text{ such that } \mathcal{K}_U \cup \{(\Gamma_Q \rightarrow a{:}C) \geq x\} \text{ satisfiable.} \quad (10)$$

$$bed(\mathcal{K}, \alpha) = \min x. \text{ such that } \mathcal{K}_U \cup \{(\Gamma_Q \rightarrow (\alpha \leq x)) = 1\} \text{ satisfiable.} \quad (11)$$

Now, let us address first the problem (9). The calculus is of the same style as in the previous section. We have that

$$bed(\mathcal{K}, \tau) = \min x. \text{ such that } \mathcal{K}_U \cup \{(\Gamma_Q \wedge \neg\tau) \geq 1 - x\} \text{ satisfiable .}$$

Therefore, it suffices to provide rules for encoding $(\Gamma_Q \wedge \neg\tau) \geq 1 - x$ as a set of MILP constraints. This is obtained as follows. Consider $\Gamma_Q = \bigwedge_{\tau_j \in \mathcal{K}_Q} \tau_j$ and assume $\tau = \langle \alpha, q\rangle$. Consider variables $x_{QC}, x_{\Gamma_Q}, x_{\neg\tau}$, and x_{τ_j}, where x_{Γ_Q} will hold the degree of truth of Γ_Q, x_{τ_j} ($x_{\neg\tau}$) will hold the degree of truth of τ_j ($\neg\tau$) and x_{QC} will hold the degree of truth of $(\Gamma_Q \wedge \neg\tau)$. Then, to encode $(\Gamma_Q \wedge \neg\tau)) \geq 1 - x$ we need

$$\begin{array}{lcl} x_{QC} & \geq & 1 - x \\ x_{\Gamma_Q} \otimes x_{\neg\tau} & = & x_{QC} \\ x_{\Gamma_Q} & = & \bigotimes_{\tau_j \in \mathcal{K}_Q} x_{\tau_j} \\ x_{QC}, x_{\Gamma_Q}, x_{\tau_j}, x_{\neg\tau} & \in & [0,1]\,. \end{array} \quad (12)$$

and we add $\langle \alpha, x_\alpha\rangle$ to $\mathcal{K}_U$. We now have to connect correctly the value of x_{τ_j} ($x_{\neg\tau}$) to the degree of truth of $\tau_j = \langle \alpha_j, q_j\rangle$ ($\neg\tau$). To this end, we need the constraints

$$x_{\tau_j} = q_j(x_{\alpha_j}) \quad (13)$$

$$x_{\neg\tau} = 1 - q(x_\alpha) \quad (14)$$

and we add $\langle \alpha, x_\alpha\rangle$ to $\mathcal{K}_U$. Note that, e.g. in Eq. 13, the equation $x_{\tau_j} = q(x_\alpha)$ is MILP expressible as q is MILP expressible function (see, e.g. [17]). For instance, if q is $\mathsf{ls}(\mathsf{a}, \mathsf{b})$ then $x_{\tau_j} = q(x_\alpha)$ may be expressed as

$$\begin{array}{l} x_\alpha + (1-a)y_1 \leq 1, x_\alpha - ay_2 \geq 0, x_\alpha + (1-b)y_2 \leq 1, \\ (b-a)x_{\tau_j} \geq x_\alpha - a - (1-a)y_2, \\ (b-a)x_{\tau_j} \leq x_\alpha - a - (1+a)y_1 + (b-a)y_2, \\ x_\alpha - by_3 \geq 0, x_{\tau_j} + y_3 \leq 1, y_1 + y_2 + y_3 = 1, \\ x_\alpha \in [0,1], y_i \in \{0,1\}\}, \end{array}$$

[3] Concerning Equation (10), suppose the minimal value is $\bar{n}$. We will know then that for any interpretation $\mathcal{I}$ satisfying $\mathcal{K}_U$, it cannot be $\mathcal{I}(\Gamma_Q \rightarrow (\alpha < \bar{n})) = 1$, and, thus, $\mathcal{I}(\Gamma_Q \rightarrow (\alpha \geq \bar{n})) = 1)$ has to hold and, thus, $bed(\mathcal{K}, \langle \alpha, \bar{n}\rangle) = 1$ and $\bar{n}$ is tight.

ISBN: 978-989-95079-6-8

where y_i are new variables (Eq. 14 is managed similarly). Now, we may proceed as for Section 5.1, where we consider these additional constraints and in which expressions of the form $\langle \alpha, x_\alpha \rangle$ are handled as for the $\langle \alpha, n \rangle$ case, except that the value n is replaced with the variable x_α instead [4]. The following can be shown:

Proposition 2 *For each KB $\mathcal{K}$, the tableau algorithm terminates and computes $bed(\mathcal{K}, \tau)$.*

As next, let us address the problem of determining $bsd(\mathcal{K}, C)$. The way we proceed is pretty similar as for $bed(\mathcal{K}, \tau)$. Now, x_{QC} will hold the degree of truth of $(\Gamma_Q \rightarrow a{:}C)$. To encode $(\Gamma_Q \rightarrow a{:}C) \geq x$, we replace the constraints (12) with

$$\begin{array}{lcl} x_{QC} & \geq & x \\ x_{\Gamma_Q} \Rightarrow x_{a:C} & = & x_{QC} \\ x_{\Gamma_Q} & = & \bigotimes_{\tau_j \in \mathcal{K}_Q} x_{\tau_j} \\ x_{QC}, x_{\Gamma_Q}, x_{\tau_j} & \in & [0,1]\,, \end{array} \tag{15}$$

we add $\langle a{:}C, x_{a:C} \rangle$ to $\mathcal{K}_U$, and then we proceed as for determining $bed(\mathcal{K}, \tau)$. The following can be shown.

Proposition 3 *For each KB $\mathcal{K}$, the tableau algorithm terminates and computes $bsd(\mathcal{K}, C)$.*

Finally, let us address the problem of determining $bed(\mathcal{K}, \alpha)$. Since $bed(\mathcal{K}, a{:}C) = \min x.$ such that $\mathcal{K}_U \cup \{(\Gamma_Q \rightarrow (a{:}\neg C \geq 1 - x)) = 1\}$ satisfiable, $bed(\mathcal{K}, (a,b){:}R) = bed(\mathcal{K} \cup \{\langle b{:}B, 1\rangle\}, a{:}\exists R.B)$, and $bed(\mathcal{K}, C \sqsubseteq D) = \min x.$ such that $\mathcal{K}_U \cup \{(\Gamma_Q \rightarrow (a{:}C \sqcap \neg D \geq 1 - x)) = 1\}$ satisfiable, we may restrict our attention to show how to encode $(\Gamma_Q \rightarrow (a{:}E \geq 1 - x)) = 1$ in MILP. To this end, let $\tau = \langle a{:}E, 1 - x \rangle$. Then we replace the constraints (12) with (recall that τ is either true or false)

$$\begin{array}{lcl} x_{\Gamma_Q} & \leq & 1 - y \\ x_{a:E} & \geq & 1 - x - y \\ x_{\Gamma_Q} & = & \bigotimes_{\tau_j \in \mathcal{K}_Q} x_{\tau_j} \\ x_{\Gamma_Q}, x_{\tau_j}, x_{a:E} & \in & [0,1] \\ y & \in & \{0,1\}\,, \end{array} \tag{16}$$

we add $\langle a{:}E, x_{a:E} \rangle$ to $\mathcal{K}_U$, and then we proceed as for determining $bed(\mathcal{K}, \tau)$. The following can be shown.

Proposition 4 *For each KB $\mathcal{K}$, the tableau algorithm terminates and computes $bed(\mathcal{K}, \alpha)$.*

Note that we may safe some obvious variables and equations in the constraints derived so far, such as in Eq 12–16.

6 Conclusions & Outlook

We have addressed the problem of allowing to deal with fuzzy truth qualified statements with fuzzy DLs and, thus, allowing statements such as "Tina is young is *almost true*". We have provided syntax, semantics of a fuzzy DL with truth qualified axioms, and a calculus addressing the various reasoning problems presented under Łukasiewicz logic.

We have also provided a novel simplified calculus for fuzzy DLs without truth qualified axioms, which is closer to the usual tableau algorithms for DLs, and which reduces the number of generated variables, and hence the size of the optimization problem to be solved.

It would be possibly to unify both approaches by using truth qualifiers of the form $\mathtt{trpz}(n, n, 1, 1)$ for fuzzy DLs without truth qualified axioms. However, it is more efficient to deal with these two cases using different strategies.

Along the line of qualified statements, we think it is useful to further extend the language by allowing so-called *probability qualified* statements (cf. [10], page 222) and, thus, allowing statements such as "The probability of the temperature t at given place and time is around 35°C is likely", where both "around" and "likely" are fuzzy sets (with, e.g. a triangular membership function to model "around", and a right shoulder function to model "likely").

[4] Note that in Step 3 and 4 of the forest initialization, in case we consider a fuzzy assertion of the form $\langle \alpha, x_\alpha \rangle$, we may omit to add $x_\alpha \geq x_\alpha$ to the constraint set $\mathcal{C}_\mathcal{F}$.

References

[1] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[2] I. Horrocks, P. F. Patel-Schneider and F. van Harmelen. From $\mathcal{SHIQ}$ and RDF to OWL: The making of a web ontology language. *Journal of Web Semantics*, 1(1):7–26, 2003.

[3] U. Straccia. Reasoning within fuzzy Description Logics. *Journal of Artificial Intelligence Research*, 14:137–166, 2001.

[4] U. Straccia. A fuzzy Description Logic for the Semantic Web. In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 4, pages 73–90. Elsevier, 2006.

[5] U. Straccia. Managing uncertainty and vagueness in Description Logics, logic programs and Description Logic programs. In *Reasoning Web, 4th International Summer School, Tutorial Lectures*, Lecture Notes in Computer Science 5224, pages 54–103. Springer Verlag, 2008.

[6] F. Bobillo and U. Straccia. fuzzyDL: An expressive fuzzy Description Logic reasoner. In *Proceedings of the 2008 International Conference on Fuzzy Systems (FUZZ-IEEE 2008)*, pages 923–930. IEEE Computer Society, 2008.

[7] F. Bobillo, M. Delgado and J. Gómez-Romero. DeLorean: A reasoner for fuzzy OWL 1.1. In *Proc. of the 4th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2008)*, CEUR Workshop Proceedings 423, 2008.

[8] G. Stoilos, N. Simou, G. Stamou and S. Kollias. Uncertainty and the Semantic Web. *IEEE Intelligent Systems*, 21(5):84–87, 2006.

[9] U. Straccia and G. Visco. DL-Media: an ontology mediated multimedia information retrieval system. In *Proc. of the 4th Int. Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2008)*, CEUR Workshop Proceedings 423, 2008.

[10] G. J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice-Hall, Inc., 1995.

[11] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer, 1998.

[12] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[13] R. Hähnle. Advanced many-valued logics. In Dov M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd Edition*, volume 2. Kluwer, 2001.

[14] R. Hähnle. Many-valued logics and mixed integer programming. *Annals of Mathematics and Artificial Intelligence*, 3,4(12):231–264, 1994.

[15] H. Salkin and M. Kamlesh. *Foundations of Integer Programming*. North-Holland, 1988.

[16] F. Baader, E. Franconi, B. Hollunder, B. Nebel, and H. J. Profitlich. An empirical analysis of optimization techniques for terminological representation systems, or: Making KRIS get a move on. *Applied Artificial Intelligence* 4:109–132, 1994.

[17] U. Straccia. Description logics with fuzzy concrete domains. In Fahiem Bachus and Tommi Jaakkola, editors, *21st Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pages 559–567, 2005. AUAI Press.

[18] T. Lukasiewicz and U. Straccia. Managing uncertainty and vagueness in Description Logics for the Semantic Web. *Journal of Web Semantics*, 6:291–308, 2008.

[19] U. Straccia and F. Bobillo. Mixed integer programming, general concept inclusions and fuzzy Description Logics. *Mathware & Soft Computing*, 14(3):247–259, 2007.

[20] F. Bobillo and U. Straccia. A fuzzy Description Logic with product t-norm. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE-07)*, pages 652–657. IEEE Computer Society, 2007.

[21] F. Bobillo and U. Straccia. Fuzzy Description Logics with general t-norms and datatypes. In *Fuzzy Sets and Systems*. To appear.

ISBN: 978-989-95079-6-8

A new approach to fuzzy location of cephalometric landmarks in craniofacial superimposition

Oscar Ibáñez[1] Oscar Cordón[1] Sergio Damas[1] Sergio Guadarrama[1] José Santamaría[2]

[1]European Centre for Soft Computing. Mieres, Spain
[2]Computer Science Department, University of Jaén. Jaén, Spain
Email: {oscar.ibanez,oscar.cordon,sergio.damas,sergio.guadarrama}@softcomputing.es, jslopez@ujaen.es

Abstract— *Craniofacial superimposition is the second stage of a complex forensic technique that aims to identify a missing person from a photograph (or video shot) and the skull found. This specific task is devoted to find the most appropriate pose of the skull model to be projected onto the photograph. The process is guided by a number of landmarks identified both in the skull (craniometric landmarks) and in the face (cephalometric landmarks). In this contribution we extend our previous genetic algorithm-based approach to the problem by considering the uncertainty involved in the location of the cephalometric landmarks and its influence in the matching between these landmarks and the craniometric ones. The new proposal is tested over two instances of a real case solved by the Physical Anthropology lab at the University of Granada (Spain).*

Keywords— Craniofacial superimposition, fuzzy landmarks, genetic algorithms, image registration.

1 Introduction

Photographic supra-projection [1] is a forensic process where photographs or video shots of a missing person are compared with the skull that is found. By projecting both photographs on top of each other (or, even better, matching a scanned three-dimensional skull model against the face photo/video shot), the forensic anthropologist can try to establish whether that is the same person. To do so, an accurate 3D model of the skull is first demanded. Next, the matching of two sets of radiometric points (facial anthropometric landmarks in the subject photograph, and cranial anthropometric landmarks in the obtained skull model) is considered to guide the superimposition of the skull 3D model and the face photograph [1]. Then, a decision making stage starts by analyzing the different kinds of achieved matches between landmarks. Due to physiognomic characteristics, some of them will perfectly match, some will partially do so, and finally some others will not. After the whole process, the forensic expert must declare whether the analyzed skull corresponds to the missing person or not.

This procedure is very time consuming and there is no systematic methodology but every expert often applies a particular process. Hence, there is a strong interest in designing automatic methods to support the forensic anthropologist to put it into effect. In a previous proposal [2], we tackled the second stage of the process, i.e. the craniofacial superimposition, by means of evolutionary algorithms assuming no uncertainty was involved in the process. Next, in [3] we identified two sources of uncertainty in the problem. On the one hand, the *landmark matching uncertainty* (not yet modeled in any of our works) will refer to the imprecision involved in the matching of the landmarks that correspond to the two different objects: the face and the skull (as said, there is a clear partial matching situation). On the other hand, the *landmark location uncertainty* is related to the extremely difficult task of locating the landmarks [4] in an invariable place, with the accuracy required for this application. Indeed, every forensic anthropologist is prone to locate the landmarks in a slightly different position. The ambiguity may also arise from reasons like variation in shade distribution depending on light condition during photographing, unsuitable camera focusing, poor image quality, etc.

In our previous contribution [3], we tackled the landmark location uncertainty considering the cephalometric landmarks as rectangular zones of different sizes, instead of using crisp locations, taking inspiration from [5]. However, this is a too simple way to represent the underlying uncertainty since all the posible crisp points in the rectangle are equally likely to be the actual location, which is not so realistic.

In this contribution we will improve the previous approach by using fuzzy sets to model the uncertainty related to each cephalometric landmark location. In addition, we will also consider fuzzy distances to model the distance between each pair of craniometric and cephalometric landmarks. The resulting genetic fuzzy system [6] is tested on two superimposition problems derived from a real-world identification case solved by the Physical Anthropology lab at the University of Granada.

The structure of the contribution is as follows. Section 2 is devoted to review our previous proposal on evolutionary craniofacial superimposition. Our new proposal is described in Section 3. Then, Section 4 presents the experimental study. Finally, Section 5 collects some concluding remarks and future works.

2 Real-coded genetic algorithm for craniofacial superimposition

In [2] we first formulated this complex task in forensic identification as a numerical optimization problem related to the well known field of image registration (IR) [7]. Then we adapted three different evolutionary algorithms to solve it: two variants of a real-coded genetic algorithm (GA) and an evolution strategy.

A sensible way to design an automatic craniofacial superimposition procedure is through the use of a IR technique to properly align the 3D skull model and the 2D face photograph in a common coordinate frame. Taking a previous proposal [8] as a base, we modeled the required perspective transformation as a set of geometric operations involving translation, rota-

ISBN: 978-989-95079-6-8

tion, scaling and projection. That registration transformation is defined by 12 parameters which are encoded in a real-coded chromosome. The search for the best transformation parameters is guided by a number of pairings between craniometric and cephalometric landmarks located in the skull and the face, respectively.

Different definitions of the objective function were studied and the one achieving the best results was the mean error[1]:

$$ME = \frac{\sum_{i=1}^{N} ||f(cl^i) - fl^i||}{N} \quad (1)$$

where $||\cdot||$ is the 2D Euclidean distance, N is the number of considered landmarks (only four in Nickerson et al.'s proposal), cl^i corresponds to every 3D craniometric landmark, fl^i refers to every 2D facial landmark, f is the function which defines the geometric 3D-2D transformation, and $f(cl^i)$ represents the position of the transformed skull 3D landmark cl^i in the projection plane.

Among the three different evolutionary designs developed to solve the problem, a real-coded generational GA using tournament selection [9], Simulated Binary crossover (SBX) [10] and random mutation operators [11] was proposed and is the one considered in this contribution.

Later on, we introduced the uncertainty treatment in the location of cephalometric landmarks [3]. In that first approach, based on the work by Sinha [5], we considered the cephalometric landmarks as rectangular zones, instead of using crisp locations. The larger the rectangle that defines each landmark, the lower the contribution of the corresponding landmark to the fitness function. Finally, we calculated the Euclidean distances between craniometric and cephalometric landmarks, using the centroid of the rectangle related to the uncertain ones. Thus, once the centroid of the uncertain landmarks was considered, the problem of computing distances between a set of uncertain landmarks and a set of crisp ones became the problem of measuring a set of Euclidean distances between different pairs of crisp landmarks. That was just a first approach to model the location uncertainty, which did not take into account the inherent uncertainty involved when we are measuring distances between fuzzy and crisp points.

3 New proposal for the fuzzy location of cephalometric landmarks

The use of fuzzy landmarks in this contribution aims to face the landmark location uncertainty in the photograph of the missing person. Furthermore, the present approach also deals with the uncertainty involved when distances between crisp points and fuzzy ones are measured, as a consequence of the fact that craniometric landmarks are considered as crisp values[2]. In order to cope with the fact that each forensic expert could place each facial landmark in different positions in the 2D image, the higher the uncertainty related to a landmark, the broader the region where the forensic experts would locate that landmark in the photograph.

[1]Notice that, mean square error is not used because of its negative effect when image ranges are normalized in [0,1]

[2]As said, we have not considered the location uncertainty in craniometric landmarks due to the big resolution (one mm per pixel) and high quality of the 3D skull model.

In addition, by using fuzzy landmarks, those experts are able to locate a larger set of landmarks with the proper level of confidence (using fuzzy regions of different sizes). In contrast, following the classical crisp approach, they would only be able to mark the landmarks whose position they can determine precisely. Due to different reasons as the pose of the missing person, the quality of the image, or partial occlusions of landmark regions, it can be difficult to do so for many of the existing cephalometric landmarks. Hence, this approach presents an important advantage since the bigger the set of landmarks, the more complete the information will be to guide the genetic search of the best transformation to properly superimpose the skull on the photograph. In this way, it will also allow us to improve the performance of our previous proposal in some problematic instances with several specific poses (see Figures 4 and 8).

To ease the comprehension of our formulation to the reader, we review some required basic concepts from Fuzzy Sets Theory [12] as follows. For each $\alpha \in (0,1]$ the α-level set $\widetilde{A}_\alpha$ of a fuzzy set $\widetilde{A}, \mu_{\widetilde{A}} :\rightarrow [0,1]$, is $\widetilde{A}_\alpha = \{x \in X; \mu_{\widetilde{A}(x)} \geq \alpha\}$. Hence, the core $\widetilde{A}_1 = \{x \in X; \mu_{\widetilde{A}(x)} = 1\}$ of a fuzzy set is the subset of X whose elements have membership equal to 1. The support $\widetilde{A}_0$ is defined as the closure of the union of all its level sets, that is

$$\widetilde{A}_0 = \overline{\bigcup_{\alpha \in (0,1]} \widetilde{A}_\alpha}$$

3.1 Distance between a point and a set of points

Given a point x of $\mathbb{R}^n$ and a nonempty subset A of $\mathbb{R}^n$ we can define a distance $d : \mathbb{R}^n \times \mathbb{P}(\mathbb{R}^n) \rightarrow \mathbb{R}^+$ by:

$$d(x, A) = inf\{||x - a||; a \in A\}$$

for a certain norm $||.||$ on $\mathbb{R}^n$. Thus, $d(x, A) \geq 0$ and $d(x, A) = 0 \Rightarrow x \in A$.

3.2 Distance between a point and a fuzzy set of points

Now we can define the distance between a point x of $\mathbb{R}^n$ an a fuzzy set of points $A : \mathbb{R}^n \rightarrow [0,1]$ by:

$$d^*(x, \widetilde{A}) = \int_0^1 d(x, \widetilde{A}_\alpha) d\alpha$$

Lemma 3.1. *The distance from the point x to the fuzzy set $\widetilde{A}$ is lesser or equal than the distance to the core of $\widetilde{A}$ and greater or equal than the distance to the support of $\widetilde{A}_0$. That is,*

$$d(x, \widetilde{A}_1) \leq d^*(x, \widetilde{A}) \leq d(x, \widetilde{A}_0)$$

The proof is straight forward.

In case that we have discrete fuzzy set of points $\widetilde{A} = x_1/\alpha_1 + \cdots + x_m/\alpha_m$, the distance can be expressed by:

$$d^*(x, \widetilde{A}) = \frac{\sum_{i=1}^{m} d(x, \widetilde{A}_{\alpha_i}) * \alpha_i}{\sum_{i=1}^{m} \alpha_i}$$

Following the idea of metric spaces in [13] we will define a fuzzy landmark as a fuzzy convex set of points having a

ISBN: 978-989-95079-6-8

nonempty core and a bounded support. That is, all its α-levels are nonempty, bounded and convex sets.

In our case, since we are dealing with 2D photographs with an $x \times y$ resolution, we can define the fuzzy landmarks as 2D masks represented as a matrix m with $m_x \times m_y$ points. Each fuzzy landmark will have a different size depending on the imprecision on its localization but at least one pixel (i.e. crisp point related to a matrix cell) will have membership with degree one. An example of a 5×5 mask is:

	1	2	3	4	5
1	0.1	0.3	0.5	0.3	0.1
2	0.3	0.5	0.7	0.5	0.3
3	0.5	0.7	1	0.7	0.5
4	0.3	0.5	0.7	0.5	0.3
5	0.1	0.3	0.5	0.3	0.1

These masks are easily built starting from two triangular fuzzy sets $\widetilde{V}$ and $\widetilde{H}$ representing the approximate vertical and horizontal position of the landmark, thus becoming two-dimensional fuzzy sets. Each triangular fuzzy set $\widetilde{A}$, is defined by its center c and its offsets l, r as follows:

$$\widetilde{A}(x) = \begin{cases} 1 - \frac{|x-c|}{c-l}, & \text{if } l \leq x \leq c \\ 1 - \frac{|x-c|}{r-c}, & \text{if } c < x \leq r \\ 0, & \text{otherwise} \end{cases}$$

and the membership functions of the fuzzy landmarks are calculated using the product t-norm by:

$$\mu_{\widetilde{F}}(i,j) = \mu_{\widetilde{V}}(i) \cdot \mu_{\widetilde{H}}(j)$$

Now we can calculate the distance between a point (which will be the projection of the 3D craniometric landmark on the 2D face photo) and the fuzzy landmark (representing the position of the cephalometric landmark), as depicted in Figure 1. Note that the implemented distance between a point and a fuzzy set of points is quite similar to the one defined in [14].

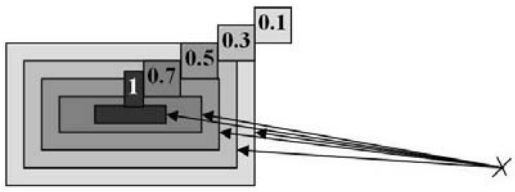


Figure 1: Distance between a point and a fuzzy point

If we denote as $d_i = d(x, \widetilde{F}_{\alpha_i})$ the distance from point x to the α-level set $\widetilde{F}_{\alpha_i})$, then the distance from the point to the fuzzy landmark $\widetilde{F}$, can be expressed by:

$$d^*(x, \widetilde{F}) = \frac{\sum_{i=1}^{m} d_i \cdot \alpha_i}{\sum_{i=1}^{m} \alpha_i}$$

In the example of Figure 1, taking $\{\alpha_1 = 0.1, \alpha_2 = 0.3, \alpha_3 = 0.5, \alpha_4 = 0.7, \alpha_5 = 1\}$ and assuming $\{d_1 = 4.5, d_2 = 5.4, d_3 = 6.3, d_4 = 7.3, d_5 = 9\}$, we calculate the distance as:

$$d^*(x, \widetilde{F}) = \frac{d_1 \cdot \alpha_1 + \cdots + d_5 \cdot \alpha_5}{\alpha_1 + \cdots + \alpha_5} = \frac{19.33}{2.6} = 7.43$$

Therefore, we have modified the previous definition of our genetic craniofacial superimposition techniques's fitness function (see section 2) as follows:

$$\textit{fuzzy ME} = \frac{\sum_{i=1}^{N} d^*(f(cl^i), \widetilde{F}^i)}{N} \quad (2)$$

where N is the number of considered landmarks, cl^i corresponds to every 3D craniometric landmark, f is the function which defines the geometric 3D-2D transformation, $f(cl^i)$ represents the position of the transformed skull 3D landmark cl^i in the projection plane, that is to say, a crisp point. $\widetilde{F}^i$ represents the fuzzy set points of each cephalometric landmark and finally, $d^*(f(C_i), \widetilde{F}^i)$ is the distance between a point and a fuzzy set points.

4 Experiments

After explaining the sources of uncertainty and our proposal to deal with the landmark location uncertainty, we will study its performance as follows. Section 4.1 presents the considered experimental design. Sections 4.2 and 4.3 are devoted to the analysis of one real-world case of study with two different photographs (two different poses).

4.1 Experimental design

Two different types of landmark sets for our case of study were provided by the forensic experts. The first type is the one classically used in the manual superimposition process. It is composed of crisp landmarks, those the forensic anthropologists can place in a single pixel. The second one is a set of fuzzy landmarks, that is to say, a region for each landmark. As said, in this second set, the forensic expert could place more landmarks than in the first one, due to the possibility of drawing bigger (in size) fuzzy sets of points. Notice that, in the former case, the forensic anthropologist would only identify those landmarks which are clearly located in a specific pixel without any doubt. According to different criteria concerning the features of both the considered landmarks and the photograph characteristics (pose, quality, etc.), forensic experts used areas of different size to determine the position of each landmark in that second set.

We compare the results of the genetic craniofacial superimposition based on crisp landmarks with those reached by using fuzzy location of cephalometric landmarks. In order to perform a significant and fair comparison between crisp and fuzzy approaches concerning the number of landmarks, we considered the following experimental design: two different sets of fuzzy landmarks are used, one with the same size (and, of course, the same specific landmarks) as the crisp set and another also including the additional landmarks identified thanks to the use of the fuzzy location approach.

The case study is a real-world one happened in Cádiz, Spain. The skull 3D model (327,641 points stored as x, y, z coordinates), see Figure 2, was acquired using Konica-Minolta 3D Lasserscanner VI-910. Two photographs[3] were provided by the family (see Figures 2 and 6). They were acquired at different moments and in different poses and conditions. Hence this case consists of two distinct superimposition

[3]Notice that we have processed the photographs to hide the subject identity following legal issues.

ISBN: 978-989-95079-6-8

problems. It was initially solved following a manual approach. The forensic experts tried to use both photographs but finally they were restricted to only one of them (pose 1 photograph), because they did not achieve a proper superimposition for the second due to the face pose.

Experiments consider the best performing GA parameter values in [2]: number of generations = 600, population size = 1,000, SBX η parameter = 1, mutation probability = 0.2, crossover probability = 0.9, and tournament size = 2. For all the tests, 30 runs of the GA are considered. Tables 1 and 2 show the minimum (m), maximum (M), mean (μ) and standard deviation (σ) of ME relative to the Euclidean distances between cephalometric landmarks and the transformed craniometric landmarks achieved by the GA. Notice that, in the case of crisp landmarks, the fitness function is directly defined as the "crisp" ME. However, when we use fuzzy landmarks the fitness function is not the ME, but a fuzzy distance between the transformed skull landmarks and the fuzzy regions (the average of the fuzzy distances). Owing to that, numerical results are not fully comparable. Nevertheless, we present the ME values in both cases as the only way to provide an error measurement.

4.2 Case study 1, pose 1

Figure 2 depicts this data set. In this first pose, the anthropologists identified nine crisp cephalometric landmarks. In addition, they marked these nine landmarks following a fuzzy approach, which also allowed them to identified a new set with five more (see Figure 3). Specially recall the large amount of fuzziness of the *vertex* landmark, located in the top part of the head. There is a strong uncertainty on its location due to the woman's hair and the forensic expert would have never trusted in that landmark in case a crisp location have to be defined. The corresponding craniometric landmarks were manually extracted from the skull 3D model.

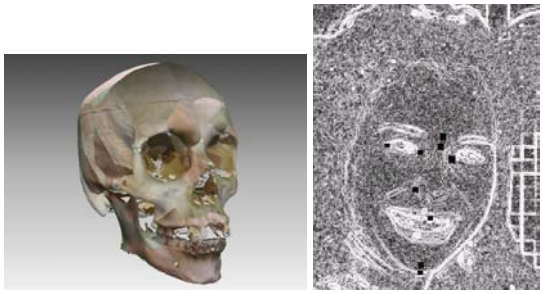


Figure 2: Case study 1, pose 1. 3D model of the skull (left) and photograph of the missing person with a set of nine crisp landmarks, represented by squares (right).

Table 1 presents the ME values for the obtained craniofacial superimpositions, distinguishing between crisp and fuzzy locations. It is important to remind that results are not fully comparable since the superimposition process using fuzzy landmarks does not minimize the ME but a different function (see Equations 1 and 2). Figures 4 and 5 present the best superimposition results of the crisp and the fuzzy approaches to allow a visual comparison.

On the one hand, referring to the numerical results, a very similar behavior is observed. Both approaches demonstrate a robust conduct. Notice that, in the case of the larger set of landmarks (fourteen fuzzy landmarks), Table 1 shows the ME corresponding just to the nine landmarks it shares with the smaller sets of landmarks. This higher ME value demonstrates that a larger error (i.e., distance between landmarks) value not necessarily means a worst superimposition (as we will see below). On the other hand, visual superimposition results show again a very similar behavior in both approaches for a smaller set of landmarks (Figure 4). These results are unacceptable for identification purposes. Finally, we can clearly identify the improvement in the superimposition when a larger set of fuzzy landmarks is used (Figure 5). Our forensic anthropologists confirmed the final superimposition result is good enough to be used in the final decision making stage of the photographic supra-projection process.

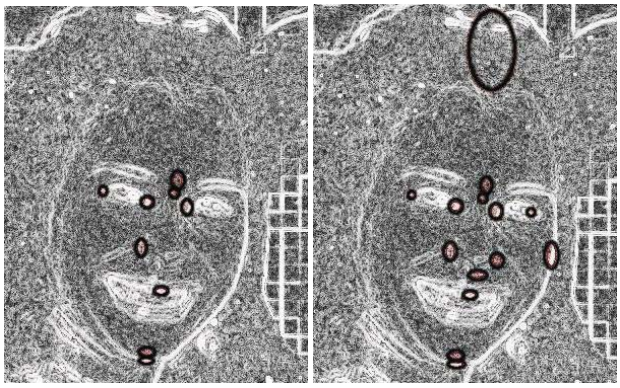


Figure 3: Case study 1, pose 1. Photograph of the missing person with two different sets of fuzzy landmarks: one of nine (left) and the other with fourteen (right) fuzzy landmarks. Landmarks are represented by ellipses.

Table 1: Case study 1, pose 1. Superimposition results.

Landmark set	ME			
	m	M	μ	σ
Nine crisp l.	0.0083	0.0086	0.0084	0.0000
Nine fuzzy l.	0.0084	0.0334	0.0095	0.0045
Fourteen fuzzy l. (ME over nine)	0.0122	0.0133	0.0124	0.0002



Figure 4: Case study 1, pose 1. Best superimposition results achieved using nine crisp (left) and nine fuzzy (right) landmarks, respectively.

ISBN: 978-989-95079-6-8



Figure 5: Case study 1, pose 1. Best superimposition result achieved using fourteen fuzzy landmarks.

4.3 Case study 1, pose 2

In this second pose, the anthropologists originally identified twelve crisp cephalometric landmarks following a crisp approach (see Figure 6). They also marked three more landmarks following a fuzzy approach. Figure 7 depicts both sets of fuzzy landmarks.



Figure 6: Case study 1, pose 2. Photograph of the missing person with the corresponding set of twelve crisp landmarks.

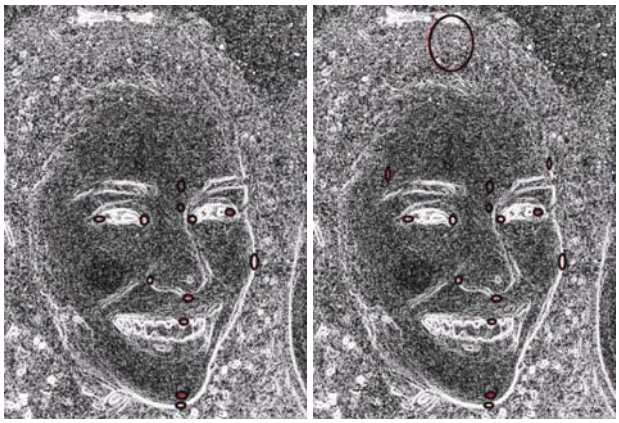


Figure 7: Case study 1, pose 2. Photograph of the missing person with two different sets of fuzzy landmarks, twelve (left) and fifteen (right).

Table 2 presents the ME figures for the obtained craniofacial superimpositions. In spite of the different fitness functions minimized, a very similar behavior is observed. As in the previous pose, the ME value of the larger set of landmarks was calculated just using the twelve landmarks that compose the smaller set. Both approaches demonstrate a robust conduct. Besides, it can be recognized how the fuzzy approach becomes more robust when considering a larger set of landmarks. On the other hand, Figure 8 presents the visual representations of the superimposition results which again show a very similar bad behavior in both approaches for a reduced set of landmarks. Finally, we can clearly see the improvement in the superimposition when a larger set of fuzzy landmarks is used (Figure 9, left). The final superimposition result is good enough to be used in the following decision making stage. In fact, this automatic superimposition is much better than that manually achieved by the forensic anthropologists (Figure 9, right), which was restricted by the computer vision tools they considered and by the lateral pose of the face photograph. The latter issues forced them to discard this photograph in the identification decision.

Table 2: Case study 1, pose 2. Superimposition results obtained using crisp and fuzzy locations

Landmark set	ME			
	m	M	μ	σ
Twelve crisp l.	0.0347	0.0354	0.0350	0.0002
Twelve fuzzy l.	0.0343	0.0613	0.0357	0.0048
Fifteen fuzzy l. (ME over twelve)	0.0382	0.0412	0.0391	0.0001



Figure 8: Case study 1, pose 2. Best superimposition results achieved using twelve crisp (left) and fuzzy (right) landmarks.

5 Concluding remarks and future works

We have proposed the use of fuzzy landmarks to tackle the uncertainty related to landmark location for a complex forensic identification task called craniofacial superimposition. We have used two different sets of landmarks provided by the forensic anthropologists for two photographs of the same identification case. We have compared the automatic superimpositions obtained by a GA considering crisp landmarks and those achieved using fuzzy landmarks. Promising results have been achieved following a fuzzy approach, showing that it is

ISBN: 978-989-95079-6-8



Figure 9: Case study 1, pose 2. Best superimposition result achieved using fifteen fuzzy landmarks (left) and best manual superimposition obtained by the forensic experts (right).

able to model the inherent uncertainty involved in the location of the landmarks. None of the achieved results following a crisp approach provided a better superimposition than the fuzzy approach. Thus, our proposal becomes more robust when tackling different locations of the landmarks.

By using fuzzy landmarks, forensic experts are able to place a larger landmark set with the proper level of confidence (using different sizes for the landmarks representing different levels of uncertainty in their actual location). In contrast, following the classical crisp approach, they are only able to mark those landmarks whose position they are able to determine precisely. As we expected, results show a much better performance of the GA when the number of landmarks is larger and that was only possible by using fuzzy location of landmarks.

As a drawback, the fuzzy approach implies more computational operations with the consequent increment in the run time required for the GA: from the 20-40 seconds per run using crisp landmarks to the 2-4 minutes using fuzzy landmarks. However, it is still a significantly short time if we compare it with the usual time needed by the forensic experts to perform a manual superimposition, around 24 hours.

Besides, these results need to be confirmed in a more extensive study, with a larger number of cases and using other different fuzzy distances [15]. As additional future work, we aim to determine the specific transformation parameters of the manual superimposition achieved by the forensic expert. We will use those parameters as the ground truth solution to measure the accuracy of the automatic superimposition processes. We will also make a poll between different forensic anthropologists in order to define the most appropriate shapes and sizes for the fuzzy landmarks in several face photographs corresponding to real-world previously solved cases. That poll will be also helpful in order to have different sets of crisp landmarks for comparison purposes. In addition, we are planning to introduce the landmark matching uncertainty (see Section 1). Finally, we aim to tackle the identification stage, i.e. the final decision making process, by using fuzzy logic, in order to assist the forensic expert in the final identification decision.

Acknowledgments

This work was partially supported by the Spain's Ministerio de Educación y Ciencia (ref. TIN2006-00829) and by the Andalusian Dpto. de Innovación, Ciencia y Empresa (ref. TIC1619), both including EDRF funding. We want to acknowledge all the team of the Physical Anthropology lab at the University of Granada (headed by Dr. Botella and Dr. Alemán) for their support during the data acquisition and validation processes. Dr. Guadarrama's work has been supported by the Spanish Department of Science and Innovation (MICINN) under program Juan de la Cierva JCI-2008-3531, and the European Social Fund.

References

[1] M. Y. Iscan. Introduction to techniques for photographic comparison. In M. Y. Iscan and R. Helmer, editors, *Forensic Analysis of the Skull*, pages 57–90. Wiley, 1993.

[2] O. Ibáñez, L. Ballerini, O. Cordón, S. Damas, and J. Santamaría. An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic identification. *Information Sciences,* in press*, http://dx.doi.org/10.1016/j.ins.2008.12.029*, 2009.

[3] O. Ibáñez, O. Cordón, S. Damas, and J. Santamaría. Craniofacial superimposition based on genetic algorithms and fuzzy location of cephalometric landmarks. In *Hybrid artificial intelligence systems*, number 5271 in LNAI, pages 599–607, 2008.

[4] J. Richtsmeier, C. Paik, P. Elfert, T.M. Cole, and F. Dahlman. Precision, repeatability and validation of the localization of cranial landmarks using computed tomography scans. *The Cleft Palate-Craniofacial Journal*, 32(3):217–227, 1995.

[5] P. Sinha. A symmetry perceiving adaptive neural network and facial image recognition. *Forensic Science International*, 98(1-2):67–89, November 1998.

[6] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena. Ten years of genetic fuzzy systems: Current framework and new trends. *Fuzzy Sets and Systems*, 141(1):5–31, 2004.

[7] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

[8] B. A. Nickerson, P. A. Fitzhorn, S. K. Koch, and M. Charney. A methodology for near-optimal computational superimposition of two-dimensional digital facial photographs and three-dimensional cranial surface meshes. *Journal of Forensic Sciences*, 36(2):480–500, March 1991.

[9] T. Blickle. Tournament selection. In T. Bäck, D. B. Fogel, and Z. Michalewicz, editors, *Handbook of Evolutionary Computation*, pages C2.3:1–C2.3:4. IOP Publishing Ltd and Oxford University Press, 1997.

[10] K. Deb and R. B. Agrawal. Simulated binary crossover for continuous search space. *Complex Systems*, 9:115–148, 1995.

[11] T. Bäck, D. B. Fogel, and Z. Michalewicz, editors. *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press, 1997.

[12] G. J. Klir and B. Yuan. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A. Zadeh*. World Scientific Publishing Co., Inc., 1996.

[13] P. Diamond and P. Kloeden. Metric topology of fuzzy numbers and fuzzy analysis. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy Sets*, The Handbooks of Fuzzy Sets, chapter 11, pages 583–637. Kluwer Academic, 2000.

[14] D. Dubois and H. Prade. On distance between fuzzy points and their use for pausible reasoning. In *International Conference on Systems, Man and Cybernetics*, pages 300–303, 1983.

[15] I. Bloch. On fuzzy distances and their use in image processing under imprecision. *Pattern Recognition*, 32:1873–1895, 1999.

ISBN: 978-989-95079-6-8

Incorporating Fuzziness in Spatial Susceptible-Infected Epidemic Models

Jan M. Baetens Bernard De Baets

KERMIT, Department of Applied Mathematics, Biometrics and Process Control
Ghent University, Coupure Links 653, 9000 Gent, Belgium
Email: jan.baetens@ugent.be, bernard.debaets@ugent.be

Abstract— *In this paper we propose a coupled-map lattice for modelling epidemic spread in a fuzzy setting. The presented model complies with the need for adequate modelling tools to describe and/or predict spatio-temporal phenomena, following the growing availability of spatio-temporal data. Furthermore, our approach does not rely on partial differential equations making it particularly suited to model epidemics in a fuzzy setting. It will be shown that the presented model allows to describe epidemic spread when the magnitude of the initial outbreak and/or the epidemic properties are only imprecisely known.*

Keywords— epidemic, discrete model, fuzzy initial condition, spatio-temporal dynamics

1 Introduction

Ordinary differential equations (ODEs) are widely used and well established to model various biological phenomena, as illustrated extensively in [1]. Partial differential equations (PDEs) are often resorted to if one is not merely interested in the process' temporal dynamics but also in the spatial patterns it generates [2]. Recently, several researchers have addressed the (numerical) solution of fuzzy ODEs (FODEs) [3, 4], endorsed by their potential importance in various scientific fields, for including imprecise information into well-established mathematical models (see [3, 5]). At present, the study of FODEs is still growing [6, 7], while thorough research on fuzzy PDEs is not yet carried out. Two distinct approaches have been developed within the theory of FODEs, differing only in whether they rely upon the notion of fuzzy derivatives or not [8]. This dichotomy and the immaturity of the theory of FODEs and FPDEs might hamper the widespread consideration of fuzziness within mathematical biology, despite often being faced with imprecise information. To overcome these barriers, we propose a coupled-map lattice (CML), developed by [9], to work with fuzzy initial conditions and/or parameters easily in a spatially explicit context. A short overview of fuzziness in discrete dynamical systems will be presented in the first section of this paper. In the second section we will introduce the CML which will be used to model spatio-temporal epidemic dynamics. In the third section we will deal with fuzzy initial conditions, while fuzzy parameters will be treated in the last section.

2 Fuzziness in discrete dynamical systems

A cellular automaton (CA), first introduced by von Neumann and Ulam as 'cellular spaces', and explored in detail by Wolfram [10, 11], is a mathematical model, discrete in all its senses, e.g. space is represented by an infinite lattice of cells, updates occur only at discrete time steps and the states can only take a finite number of values [10]. CA and CML are closely related, but in the latter states can take arbitrary values [12]. Although the first notion of a fuzzy CA dates back to the late 60s [13], literature on fuzzy CA is considered insufficient [14]. Fuzzy automata were first defined by Wee and Fu [13], but several alternative definitions have been proposed [14, 15]. Essentially, every state is attributed a membership value in a fuzzy CA [15], as such relaxing the condition of merely allowing discrete states. Recently, a review on fuzzy automata has been published [14], following the increasing number of articles published on the topic (see [16, 17, 18]). Despite the definition given in [13], fuzzy CA are often simply understood as spatial extensions of fuzzy rule-based models such as Mamdami-Assilian [19, 20] or Takagi-Sugeno models [21], e.g. CA in which the local transition function is a fuzzy rule-based model (see [22, 23, 24, 25, 26, 27]). To our knowledge, the notion of fuzziness in CML models has not been addressed yet.

3 Spatially explicit modelling of epidemics

3.1 The model

Several authors, including Kaneko [12] and Wolfram [11], have argued that CML and CA provide a suitable framework to deal with spatio-temporal dynamics. This is illustrated by the rich variety of such models that has been developed during the last decade for describing various spatial biological phenomena such as epidemics [28, 29, 30], population dynamics [31, 32], tumor growth [33, 34, 35], biofilm development [36, 37] and much other phenomena [38, 39].

Recently, Baetens and De Baets [9] proposed a generalized CML for modelling various biological phenomena, traditionally described by means of PDEs. Their model is general in two senses. Firstly, exploiting graph notations makes it independent from the spatial subdivisions used to discretize the space domain and secondly, it can serve as a basis for modelling various biological phenomena. With regard to an epidemic sweeping through a region which is subdivided into irregular polygons, and involving only non-reproducing susceptible and infected individuals, it can be written in a simplified form as

$$\begin{cases} S_j^{t+1} = S_j^t - S_j^t \displaystyle\sum_{P_k \in N(P_j)} w_{jk}\, G\left(\mathbf{I}_j, d_{jk}\right)\, I_k^t \\ I_j^{t+1} = I_j^t + S_j^t \displaystyle\sum_{P_k \in N(P_j)} w_{jk}\, G\left(\mathbf{I}_j, d_{jk}\right)\, I_k^t \end{cases} \tag{1}$$

where S_j^t, resp. I_j^t, represent the fraction of susceptible, resp. infected individuals within polygon P_j at the t-th time step

ISBN: 978-989-95079-6-8

and G is a function describing the effect of landscape and connectivity characteristics on the epidemic spread, d_{jk} is the distance measured on a graph between the vertices v_j and v_k, representing polygons P_j and P_k, $N(P_j)$ is the set of polygons of which the infected inhabitants can affect I_j^{t+1}, e.g. the neighbourhood of P_j defined as

$$N(P_j) = \{P_k \in P \mid d_{ij} \leq \epsilon\} \,, \tag{2}$$

where ϵ is the neighbourhood radius. Further, w_{jk} is a weighting factor, weighing the influence of every $P_k \in N(P_j)$ in the determination of S_j^{t+1}, and $\mathbf{I}_j$ contains information about P_j. This model may be regarded as a discrete analogue of the PDE-based SI-model (see [2]). Assuming $S_j^t + I_j^t = 1$, at all t, and for all P_j, we only have to keep track of one of the system's equations in order to follow its dynamics. From (1) it follows clearly that the system has a single fixed point $\left(S_j^*, I_j^*\right) = (0, 1)$.

Within the framework of this paper, only a rectangular lattice consisting of 101×101 equally-sized cells will be considered, since we rather want to focus on the fuzziness in (1) than on its general applicability as discussed in [9]. Moreover, we will assume that the region is spatially homogeneous, e.g. G does not depend on $\mathbf{I}_j$, reducing (1) to

$$I_j^{t+1} = I_j^t + S_j^t \sum_{P_k \in N(P_j)} w_{jk}\, H\left(d_{jk}\right)\, I_k^t \,, \tag{3}$$

where we introduced the function H, only dependent on d_{jk}, to distinguish from G, potentially dependent on both d_{jk} and $\mathbf{I}_j$. Further, we will define

$$H\left(d_{jk}\right) = \begin{cases} \nu_0, \text{ if } d_{jk} = 0, \\ \nu_1, \text{ if } d_{jk} = 1, \\ \quad \vdots \\ \nu_\epsilon, \text{ if } d_{jk} = \epsilon, \end{cases} \tag{4}$$

where ϵ is the neighbourhood radius and ν_ϵ is a measure for the epidemic's virulence, in such a way that

$$\sum_{P_k \in N(P_j)} w_{jk}\, H\left(d_{jk}\right) \leq 1, \quad \forall\, t,\, P_j \,, \tag{5}$$

assuring that $0 \leq I_j^t \leq 1$, at all t, and for all P_j. For this paper, let $\epsilon = 1$, $w_{jk} = \frac{1}{8}$ for all j, k and $j \neq k$ and $w_{jk} = 1$ if $j = k$. Consequently, P_j's eight nearest neighbours influence I_j^{t+1} to the same degree.

3.2 Fuzziness in the proposed model

In the remainder of this paper S_j^t and I_j^t are considered fuzzy intervals in $[0, 1]$, while the notations $\left[S_j^t\right]^\alpha$ and $\left[I_j^t\right]^\alpha$ will refer to their respective α-cuts defined by $\left[S_j^t\right]^\alpha = \left\{s \in [0,1] \mid S_j^t(s) \geq \alpha\right\}$ if $\alpha > 0$ and $\left[S_j^t\right]^0 = \text{cl}\left\{s \in [0,1] \mid S_j^t(s) > 0\right\}$ (the closure of the support) if $\alpha = 0$, and analogously for $\left[I_j^t\right]^\alpha$. In consequence, we can write $\left[S_j^t\right]^\alpha = \left[s_{j,1}^t(\alpha), s_{j,2}^t(\alpha)\right]$ and $\left[I_j^t\right]^\alpha = \left[i_{j,1}^t(\alpha), i_{j,2}^t(\alpha)\right]$. A triangular fuzzy interval A will be denoted as $A \equiv (l, u, r)$ where $[A]^0 = [l, r]$ and $[A]^1 = \{u\}$. Two sources of fuzziness in (3) will be examined more closely. A first concerns the initial conditions I_k^0 necessary to iteratively solve (3). As indicated by Beale [40], spatial epidemiological data are becoming increasingly available, providing the means to deduce I_k^0. Nonetheless, one has to be aware of the imprecision in the outbreak data which can be modeled with a fuzzy set approach. Analogously, the parameters in (4) might only be known imprecisely, and will be regarded as the second source of fuzziness. Since the mathematical functions in (3) are restricted to the basic arithmetic operations and involve no derivatives, it becomes relatively straightforward to obtain the fuzzy system's spatio-temporal dynamics. Nonetheless, one has to bear in mind the coupling between S_j^t and I_j^t through the condition $S_j^t + I_j^t = 1$, at all t, and for all P_j. As such, S_j^t and I_j^t cannot take values independently of each other and may be called interactive fuzzy variables [41]. Their joint possibility distribution C_j^t is not given by $S_j^t \times I_j^t$ as illustrated in Fig. 1(a), but can be defined as [42]

$$\left[C_j^t\right]^\alpha = \left\{l\left(s_{j,1}^t(\alpha), i_{j,2}^t(\alpha)\right) + (1-l)\left(s_{j,2}^t(\alpha), i_{j,1}^t(\alpha)\right) \mid l \in [0,1]\right\} \,, \tag{6}$$

for all $\alpha \in [0, 1]$, and is depicted in Fig. 1(b). The extension $f(S_j^t, I_j^t)$ is defined by [43]

$$f(S_j^t, I_j^t)\,(z) = \sup_{f(x,y)=z} C_j^t\,(x, y) \,. \tag{7}$$

Following the theory outlined in [44], S_j^t and I_j^t may be called completely negatively 'correlated'. In view of the existing interactivity we can write

$$\left[S_j^t + I_j^t\right]^\alpha = \left[s_{j,1}^t(\alpha) + i_{j,2}^t(\alpha), s_{j,2}^t(\alpha) + i_{j,1}^t(\alpha)\right] \,, \tag{8}$$

and

$$\left[S_j^t \cdot I_j^t\right]^\alpha = \left[\min\left(s_{j,2}^t(\alpha) \cdot i_{j,1}^t(\alpha), s_{j,1}^t(\alpha) \cdot i_{j,2}^t(\alpha)\right), \max\left(s_{j,2}^t(\alpha) \cdot i_{j,1}^t(\alpha), s_{j,1}^t(\alpha) \cdot i_{j,2}^t(\alpha)\right)\right] \,. \tag{9}$$

All simulations were performed in Mathematica 6.0 (Wolfram Research, Inc.) on a desktop PC with an Intel Dual Core 1.86 GHz processor. Although a quadratic lattice was used in this paper, the described simulations can easily be performed when irregular spatial subdivisions would be utilized.

4 Fuzzy initial conditions

In this section, we assume $\nu_0 = 0.5$ and $\nu_1 = 0.5$, meaning that the spread of an infection in a polygon P_j can be equally attributed to infected individuals living in P_j as to infected individuals residing in P_j's neighbourhood. First, we will consider the situation in which a disease outbreak was recorded in the center polygon P_c of the grid, but only imprecise information was available on the proportion of initially infected individuals. As such, we adopted the following initial condition

$$\begin{cases} I_j^0 = (0.2, 0.3, 0.4) \,, \text{ if } j = c \\ I_j^0 = 0 \qquad\qquad\quad \,, \text{ else.} \end{cases} \tag{10}$$

Figure 2 illustrates the CML evolution during the first time interval: starting from the initial condition (Fig. 2(a), 2(d)),

ISBN: 978-989-95079-6-8

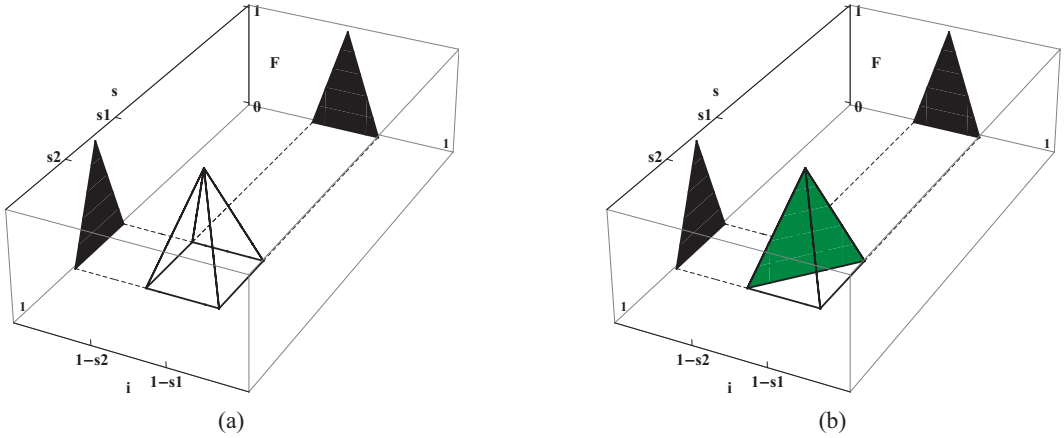


Figure 1: Non-interactive (a) and interactive (b) joint possibility distributions.

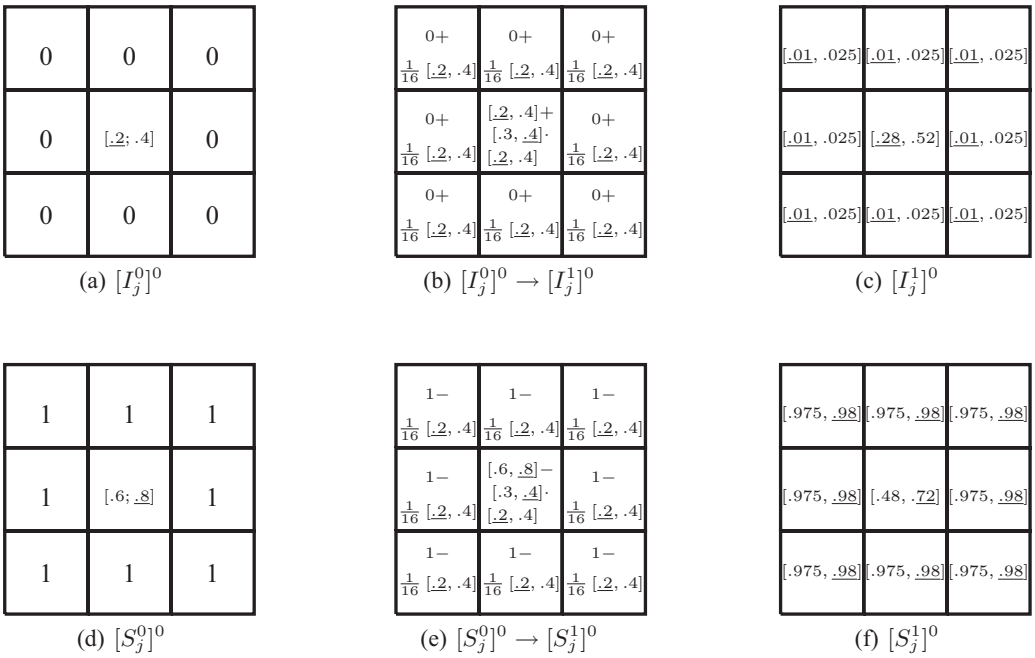


Figure 2: Evolution from $\left[I_j^0\right]^0$ to $\left[I_j^1\right]^0$ (a, b, c) and from $\left[S_j^0\right]^0$ to $\left[S_j^1\right]^0$ (e, d, f).

the lattice sites are updated according to (3), yielding the updated fuzzy states (Fig. 2(c), 2(f)). Since the neighbourhood was restricted to the nearest neighbours, it suffices to plot only the center polygon and its nearest neighbours to view all state changes during the first time interval. Underlining was used in order to visualize how the interaction between S_j^t and I_j^t was taken into account. For instance, $\left[I_j^1\right]^0$ was calculated as

$$\begin{aligned}\left[I_j^1\right]^0 &= \left[I_j^0\right]^0 + \left[s_{j,2}^0(0)\cdot\tfrac{1}{2}\,i_{j,1}^0(0),\; s_{j,1}^0(0)\cdot\tfrac{1}{2}\,i_{j,2}^0(0)\right]\\ &= [0.2, 0.4] + [0.8\cdot 0.1, 0.6\cdot 0.2]\\ &= [0.28, 0.52]\end{aligned}$$

Figure 3 shows the proportion of infected individuals after two, five, ten and fifteen time steps following the epidemic outbreak. For reasons of clarity, we limited the depicted spatial extent. The fill color of the polygons corresponds to the value of $\left[I_j^t\right]^1$, while the vertical axis indicates the membership value of all $i \in [0, 1]$. For each polygon, the fuzzy interval $I_j^t(i)$ is depicted at the center of the Y-direction, with the i-values marked along the X-direction, as indicated by the additional axis drawn at the back of the plots depicted in Fig. 3(a). This figure shows clearly that, although the epidemic was caused by an imprecise number of infected individuals residing in the center polygon, I_c^5's support was narrower than that of I_c^0, meaning that the imprecision of the model prediction decreased in this polygon during the considered time interval. This suggests that I_j^t tended to a crisp number as $t \rightarrow \infty$ which was verified experimentally by running the model for 150 time steps. This tendency could be expected since $(S_c^*, I_c^*) = (0, 1)$ is an asymptotic fixed point of the CML. Further, it is shown in this figure that the imprecision, initially only present in the outbreak data from the center polygon, propagated radially as a traveling wavefront giving rise to fuzziness in the model predictions at other locations. Though,

ISBN: 978-989-95079-6-8

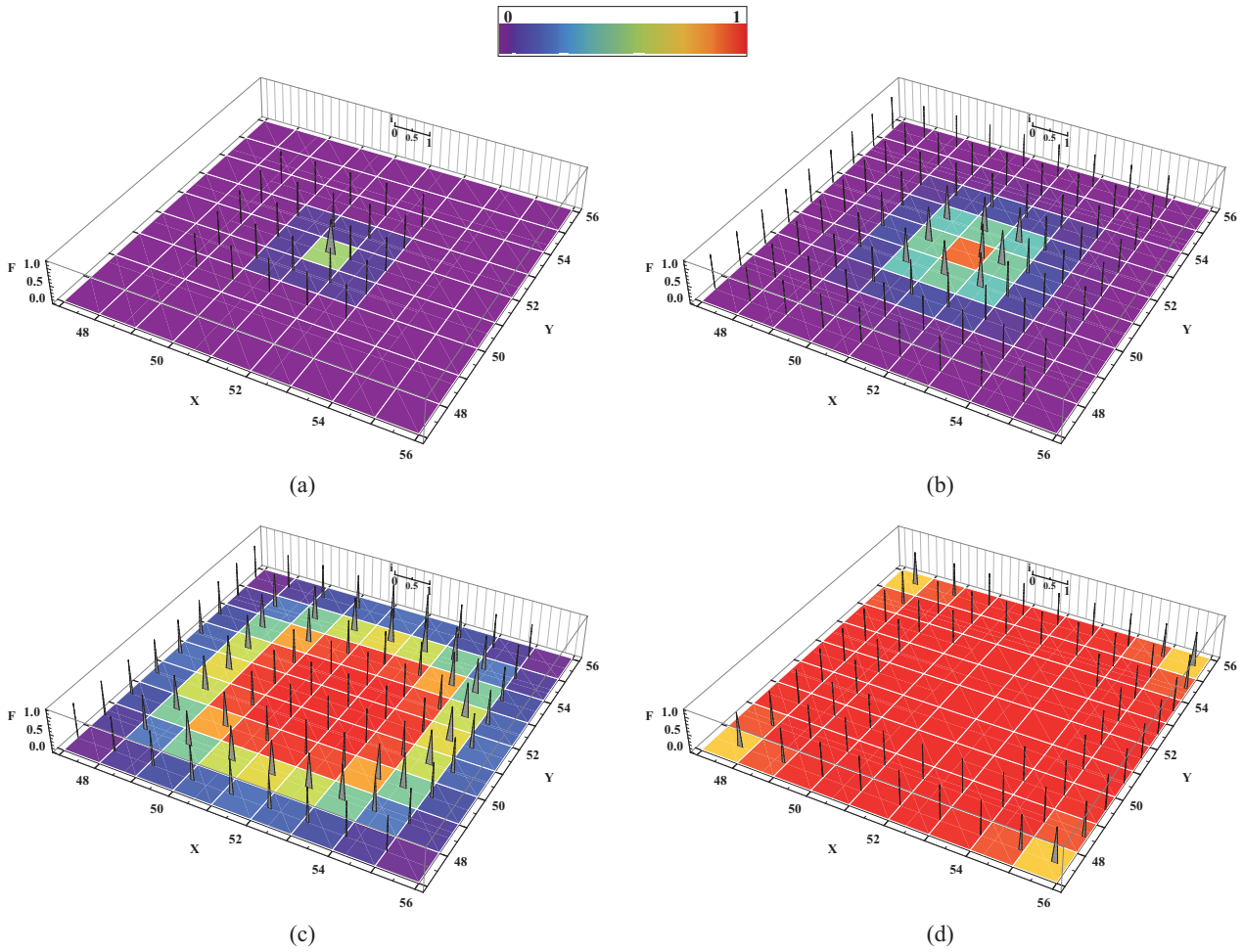


Figure 3: CML state two (a), five (b), ten (c) and fifteen (d) time steps after an outbreak was recorded in the center polygon.

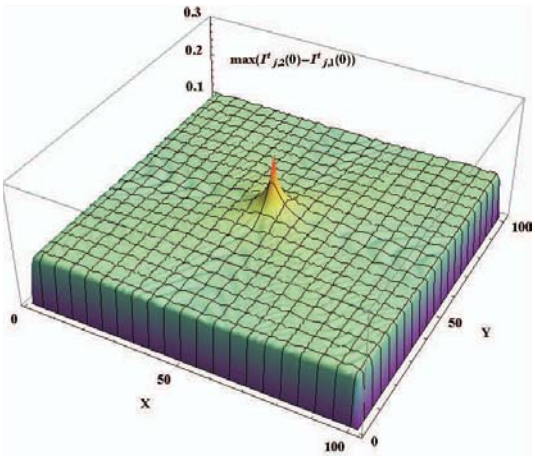


Figure 4: Maximum $[I_c^t]^0$ appeared during the first 125 time steps after an outbreak was recorded in the center polygon of the grid.

the degree of imprecision, expressed as the maximum width of I_j^t's support for all t, decreased asymptotically to 0.1 as d_{cj} increased (Fig. 4).

Also the situation of simultaneous epidemic outbreaks at different locations was studied more closely. For this setup, it can be expected that the imprecision linked to the outbreak data in each of the initially infected polygons might 'interact' when the epidemic waves meet. This was confirmed by the small roughnesses between the two peaks in Fig. 5 showing the maximum width of I_j^t's support during the first 50 time steps after an epidemic was initiated in two polygons which are at distance 14 from one another. Similar to the above-described situation, I_j^t tended to a crisp number as $t \to \infty$ and the CML evolved towards its fixed point.

5 Uncertain epidemic properties

In this section we again consider the model given by (3) with fuzzy initial conditions (10), but in addition we assume that ν_0 was not known accurately and could be described as a triangular fuzzy interval $\nu_0 = (0.2, 0.35, 0.5)$. Comparing Figs. 6 and 3 one clearly sees that I_j^t's support was wider when both the initial condition and ν_0 were only imprecisely known. As the wavefront propagated, the maximum width of the support spanned the entire unit interval in polygons further from P_c due to the successive non-interactive multiplication of ν_0 and $I_{P_i}^t$. Yet, I_j^t tended to a crisp number as $t \to \infty$ since the CML evolved to its fixed point.

6 Conclusions

In this paper it was shown that imprecise information, described using a fuzzy set approach, can be used easily within a spatially explicit model, based upon the coupled-map lattice

ISBN: 978-989-95079-6-8

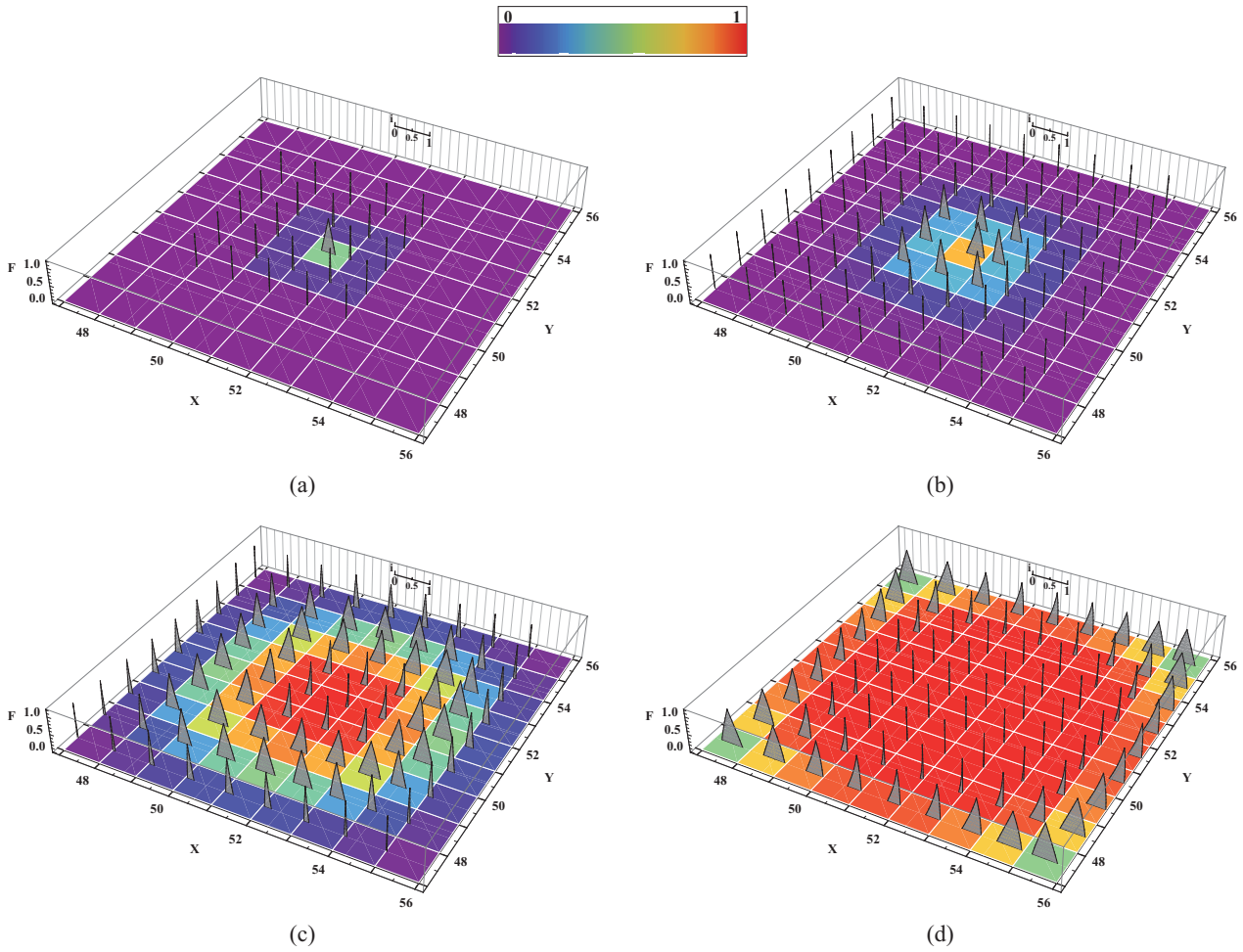


Figure 6: CML state two (a), five (b), ten (c) and fifteen (d) time steps after an outbreak was recorded in the center polygon.

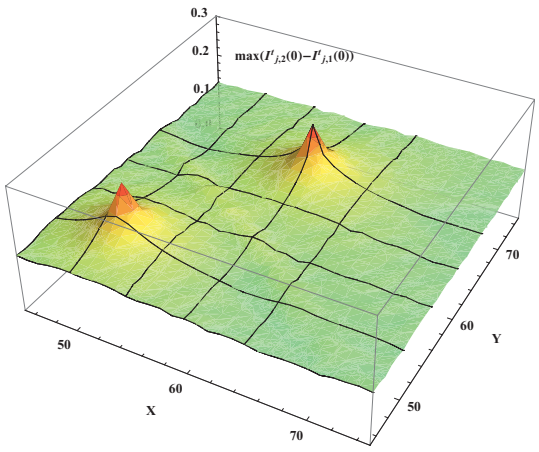


Figure 5: Maximum $[I_c^t]^0$ appeared during the first 125 time steps after an outbreak was recorded in two polygons, which are at distance 14 from one another.

paradigm. The presented modelling framework is perfectly suited to cope with the growing importance and availability of spatio-temporal data. Further work on this topic will consist of extending the presented model to cover also recovered individuals.

References

[1] J.D. Murray. *Mathematical Biology: I. An Introduction*. Springer, Berlin, Germany, second edition, 1993.

[2] J.D. Murray. *Mathematical Biology: II. Spatial Models and Biomedical Applications*. Springer, Berlin, Germany, third edition, 2007.

[3] M. Oberguggenberger and S. Pittschmann. Differential equations with fuzzy parameters. *Mathematical and Computer Modelling of Dynamical Systems*, 5:181–202, 1999.

[4] V. Lakshmikantham and R.N. Mohapatra. Theory of fuzzy differential equations and inclusions. In R.P. Agarwal and D. O'Regan, editors, *Series in Mathematical Analysis and Applications*, volume 6. Taylor & Francis, New York, United States, 2003.

[5] M. Guo and R. Li. Impulsive functional differential inclusions and fuzzy populations models. *Fuzzy Sets and Systems*, 138:601–615, 2003.

[6] J.J. Buckley and T. Feuring. Fuzzy differential equations. *Fuzzy Sets and Systems*, 110:43–54, 2000.

ISBN: 978-989-95079-6-8

[7] Y. Chalco-Cano and H. Román-Flores. On new solutions of fuzzy differential equations. *Chaos, Solitons and Fractals*, 38:112–119, 2008.

[8] D. Vorobiev and S. Seikkala. Towards the theory of fuzzy differential equations. *Fuzzy Sets and Systems*, 125:231–237, 2002.

[9] J.M. Baetens and B. De Baets. A generalized coupled-map lattice to model biological phenomena. *Mathematical Biology*, 2009. Submitted.

[10] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55:601–644, 1983.

[11] S. Wolfram, editor. *A New Kind of Science*. Wolfram Media Inc., Champaign, United States, 2002.

[12] K. Kaneko, editor. *Theory and Applications of Coupled Map Lattices*. John Wiley & Sons Ltd., Chichester, United Kingdom, 1193.

[13] W.G. Wee and K.S. Fu. A formulation of fuzzy automata and its application as a model of learning systems. *IEEE Transactions on Systems Science and Cybernetics*, 5:215–223, 1969.

[14] M. Doostfatemeh and S.C. Kremer. New directions in fuzzy automata. *International Journal of Approximate Reasoning*, 38:175–214, 2005.

[15] J.N. Mordeson and D.S. Malik. *Fuzzy automata and languages, theory and applications*. Chapman and Hall/CRC, Boca Raton, United States, 2002.

[16] G. Cattaneo, P. Flocchini, C. Quaranta Vogliotti, and N. Santoro. Cellular automata in fuzzy backgrounds. *Physica D*, 105:105–120, 1997.

[17] P. Flocchini, F. Geurts, and N. Santoro. CA-like error propagation in fuzzy CA. *Parallel Computing*, 23:1673–1682, 1997.

[18] D. Dunne and A. Mingarelli. On the dynamics of some exceptional fuzzy cellular automata. In *7th International Conference on Cellular Automata for Research and Industry, ACRI 2006*, pages 78–87, Perpignan, France, September 2006.

[19] S. Assilian. *Artificial Intelligence in the Control of Real Dynamical Systems*. PhD thesis, London University.

[20] E.H. Mamdani. Application of fuzzy algorithms for control of simple dynamic plant. *Proceedings of the IEEE*, 121:1585–1588, 1974.

[21] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15:116–131, 1985.

[22] C. Bone, S. Dragicevic, and A. Roberts. A fuzzy-constrained cellular automata model of forest insect manifestations. *Ecological Modelling*, 192:107–125, 2006.

[23] F. Wu. Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. *Journal of Environmental Management*, 5:293–308, 1998.

[24] M. da Silva Peixoto, L. Carvalho de Barros, and R. Carlos Bassanezi. A model of cellular automata for the spatial and temporal analysis of citris sudden death with the fuzzy parameter. *Ecological Modelling*, 214:45–52, 2008.

[25] M. Mraz, N. Zimic, and J. Virant. Intelligent bush fire spread prediction using fuzzy cellular automata. *Journal of Intelligent and Fuzzy Systems*, 7:203–207, 1999.

[26] Q. Chen and A.E. Mynett. Modelling algal bloom in the Dutch coastal waters by integrated numerical and fuzzy cellular automata approaches. *Ecological Modelling*, 199:73–81, 2006.

[27] L.A. Mentelas, P. Prastacos, and T. Hatzichristos. Modeling urban growth using fuzzy cellular automata. In *Proceedings of the 11th AIGLE Conference (CD-ROM)*, Girona, Spain, 2008.

[28] S. Hoya White, A. Martín del Rey, and G.Rodríguez Sanchez. Modeling epidemics using cellular automata. *Applied Mathematics and Computation*, 186:193–202, 2007.

[29] J. Milne and S C. Fu. Epidemic modelling using cellular automata. In *Proceedings of the 1st Australian Conference on Artificial Life (ACAL'03)*, pages 43–57, Canberra, December 2003.

[30] R.J. Doran and S.W. Laffan. Simulating the spatial dynamics of foot and mouth disease outbreaks in feral pigs and livestock in Queensland, Australia, using a susceptible-infected-recovered cellular automata model. *Preventive Veterinary Medicine*, 70:133–152, 2005.

[31] A.K. Dewdney. Sharks and fish wage an ecological war on the toroidal planet. *Scientific American*, pages 14–22, December 1984.

[32] H. Baltzer, P.W. Braun, and W. Köhler. Cellular automata models for vegetation dynamics. *Ecological Modelling*, 107:113–125, 1998.

[33] A. Qi, X. Zheng, C. Du, and B. An. A cellular automaton model of cancerous growth. *Journal of Theoretical Biology*, 161:1–12, 1993.

[34] D.G. Mallet and L.G. De Pillis. A cellular automata model of tumor-immune system interactions. *Journal of Theoretical Biology*, 239:334–350, 2006.

[35] L. Preziosi. *Cancer Modelling and Simulation*. Chapman & Hall, Boca Raton, United States, 2003.

[36] C. Picioreanu, M.C.M. van Loosdrecht, and J.J. Heijnen. Mathematical modeling of biofilm structure with a hybrid differential-discrete cellular automaton approach. *Biotechnology and Bioengineering*, 58:101–116, 1998.

[37] G. Pizarro, D. Griffeath, and D.R. Noguera. Quantitative cellular automaton model for biofilms. *Journal of Environmental Engineering*, 127:782–789, 2001.

[38] L.B. Kier, P.G. Seybold, and C. Cheng. *Modelling Chemical Systems using Cellular Automata*. Springer, Dordrecht, The Netherlands, 2005.

[39] J.L. Schiff. *Cellular Automata: A Discrete View of the World*. John Wiley & Sons Ltd., Chichester, United Kingdom, 2008.

[40] L. Beale, J.J. Abellan, S. Hodgson, and L. Jarup. Methodologic issues and approaches to spatial epidemiology. *Environmental Health Perspectives*, 116:1105–1110, 2008.

[41] D. Dubois and H. Prade. *Possibility Theory: an Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, United States, 1988.

[42] R. Fullér and P. Majlender. On interactive fuzzy numbers. *Fuzzy Sets and Systems*, 143:355–369, 2004.

[43] P. Majlender. *A Normative Approach to Possibility Theory and Soft Decision Support*. PhD thesis, Åbo Akademi University, Finland, 2004.

[44] C. Carlsson, R. Fullér, and P. Majlender. On possibilistic correlation. *Fuzzy Sets and Systems*, 155:425–445, 2005.

ISBN: 978-989-95079-6-8

Recurrent Neural Kalman Filter Identification and Indirect Adaptive Control of a Continuous Stirred Tank Bioprocess

Ieroham S. Baruch[1] Carlos R. Mariaca-Gaspar[1]

1 Department of Automatic Control, CINVESTAV-IPN,
Ave. IPN No 2508, Col. Zacatenco, A.P. 14-740,
07360 Mexico City, MEXICO
Email: {baruch,cmariaca}@ctrl.cinvestav.mx

Abstract— *The aim of this paper is to propose a new Kalman Filter Recurrent Neural Network (KFRNN) topology and a recursive Levenberg-Marquardt (L-M) algorithm of its learning capable to estimate states and parameters of a highly nonlinear Continuous Stirred Tank Bioreactor (CSTR) in noisy environment. The estimated parameters and states obtained by the proposed KFRNN identifier are used to design an indirect adaptive sliding mode control scheme. The obtained simulation results of the real-time neural identification and control of a CSTR model, taken from the literature, exhibited fast convergence, noise filtering, and low mean squared error of reference tracking. A 20 runs comparative validation experiment in noisy environment is also done. It gives some priority of the L-M learning over the BP one.*

Keywords— Backpropagation learning, continuous stirred tank bioreactor, indirect adaptive sliding mode control, Kalman filter recurrent neural network, Levenberg-Marquardt learning.

1 Introduction

The Recent advances in understanding of the working principles of artificial neural networks has given a tremendous boost to identification, prediction and control tools of nonlinear systems, [1], [2]. The main network property namely the ability to approximate complex non-linear relationships without prior knowledge of the model structure makes them a very attractive alternative to the classical modeling and control techniques, [3]. It is particularly useful in applications where the complexity of the data or tasks makes the design of such a functions by hand impractical. Among several possible network architectures the ones most widely used are the Feedforward NN (FFNN) and Recurrent NN (RNN). In a FFNN the signals are transmitted only in one direction, starting from the input layer, subsequently through the hidden layers to the output layer, which requires applying a tap delayed global feedbacks and a tap delayed inputs to achieve a Nonlinear Autoregressive Moving Average (NARMAX) neural dynamic plant model. A RNN has local feedback connections to some of the previous layers. Such a structure is suitable alternative to the FFNN when the task is to model dynamical systems. Its main advantage is the reduced complexity of the network structure. However, the analysis of state of the art in the area of classical RNN-based modelling and control has also shown some of their inherent limitations as follows: 1. The RNN input vector consists of a number of past system inputs and outputs and there is not a systematic way to define the optimal number of past values, and usually, the method of trials and errors is performed, [2]. 2. The RNN model is naturally formulated as a discrete model with fixed sampling period, therefore, if the sampling period is changed, the neural network has to be trained again; 3. It is assumed that the plant order is known, which represents a quite strong modelling assumption in general, [1]; 4. The managing of noisy input/output plant data required to augment the filtering capabilities of the identification RNNs, [4]. Driven by these limitations, a new Kalman Filter Recurrent Neural Network (KFRNN) topology and the recursive Backpropagation (BP) type learning algorithm in vector-matrix form was derived and its convergence was studied, [5]. But the recursive BP algorithm, applied for KFRNN learning, is a gradient descent first order learning algorithm which not permits to augment the precision and to accelerate the learning. So, the aim of the paper is to use a second order learning algorithm for the KFRNN like the Levenberg-Marquardt (L-M) algorithm is, [4, 6]. The KFRNN with L-M learning will be applied for Continuous Stirred Tank Reactor (CSTR) model identification, [7]. In [8, 9] a comparative study of linear, nonlinear and neural-network-based adaptive controllers for a CSTR is done. The papers proposed to use the neuro-fuzzy and adaptive nonlinear control systems design, applying FFNNs (multilayer perceptron and radial basis functions NN). The proposed control gives a good adaptation to the nonlinear plants dynamics, better with respect to the other methods of control, but the used FFNNs have a great complexity, and the plant order has to be known. The application of KFRNNs, [5], together with the recursive L-M, [4], could avoid these problems improving the learning and the precision of the plant states and parameters estimation. Here the obtained from the KFRNN identifier parameters and states will be used to design an indirect adaptive sliding mode control scheme, [5].

2 Topology and learning of the KFRNN

2.1 KFRNN Topology

A Kalman Filter Recurrent Neural Network model and its learning algorithm of dynamic Backpropagation-type, together with the explanatory figures and stability proofs, are described in [5]. The KFRNN topology (see Fig. 1) is described by the following vector-matrix equations:

ISBN: 978-989-95079-6-8

$$X(k+1) = A_1 X(k) + BU(K) - Dy(k) \quad (1)$$

$$Z(k) = \mathrm{G}[X(k)] \quad (2)$$

$$V_1(k) = CZ(k) \quad (3)$$

$$V(k+1) = A_2 V(k) + V_1(k) \quad (4)$$

$$Y(k) = \mathrm{F}[V(k)] \quad (5)$$

$$A_1 = block\text{-}diag\,(A_{1,i})\,;\ |A_{1,i}| < 1$$
$$A_2 = block\text{-}diag\,(A_{2,i})\,;\ |A_{2,i}| < 1 \quad (6)$$

Where: Y, X, and U are, respectively, output, state and input vectors with dimensions l, n, m; A_1 and A_2 are (nxn) and (lxl) block-diagonal local feedback weight matrices; $A_{1,i}$ and $A_{2,i}$ are i-th diagonal block of A_1 and A_2 with (nxn) and (lxl) dimensions, respectively. Equation (6) represented the local stability conditions, imposed on all blocks of A_1 and A_2; B and C are (nxm) and (lxn) input and output weight matrices; D is a (nxl) global output feedback weight matrix; G(.), F(.) are vector-valued sigmoid or hyperbolic tangent-activation functions, Z, V_1, V are vector variables with corresponding dimensions; the integer k is a discrete-time variable.

2.2 Backpropagation learning of the KFRNN

The general BP learning algorithm is given by the following equation:

$$W(k+1) = W(k) + \eta \Delta W(k) + \alpha \Delta W(k-1) \quad (7)$$

Where: W is the weight matrix, being modified (A_1, A_2, B, C, D); ΔW is the weight matrix correction (ΔA_1, ΔA_2, ΔB, ΔC, ΔD); η and α are learning rate parameters. Applying the diagrammatic method, [10], and using the block- diagram of the KFRNN topology (see Fig. 1), we could design an error predictive adjoint KFRNN (see Fig. 2). Following this adjoint KFRNN block diagram we could obtain the next matrix KFRNN weight updates:

$$\Delta C(k) = E_1(k) Z^T(k) \quad (8)$$

$$\Delta A_2(k) = E_1(k) V^T(k) \quad (9)$$

$$E_1(k) = F'[Y(k)]E(k);\ E(k) = Y_p(k) - Y(k) \quad (10)$$

$$\Delta B(k) = E_3(k) U^T(k) \quad (11)$$

$$\Delta A_1(k) = E_3(k) X^T(k) \quad (12)$$

$$\Delta D(k) = E_3(k) Y^T(k) \quad (13)$$

$$E_3(k) = G'[Z(k)]E_2(k);\ E_2(k) = C^T E_1(k) \quad (14)$$

$$\Delta vA_1(k) = E_3(k) \oplus X(k);\ \Delta vA_2(k) = E_1(k) \oplus V(k) \quad (15)$$

Where: ΔA_1, ΔA_2, ΔB, ΔC, ΔD are weight corrections of the of the learned matrices A_1, A_2, B, C, D, respectively; E is an error vector of the output KFRNN layer, where Yp is a desired target vector and Y is a KFRNN output vector, both with dimensions l; X is a state vector, and E1, E2, E3 are error vectors, shown on Fig. 2; F'(.), G'(.) are diagonal

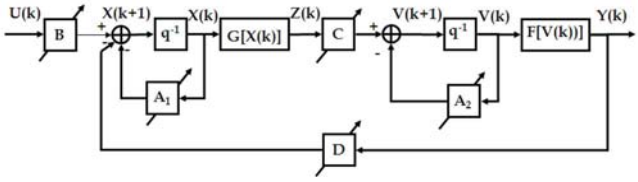


Figure 1: Block-diagram of the KFRNN topology.

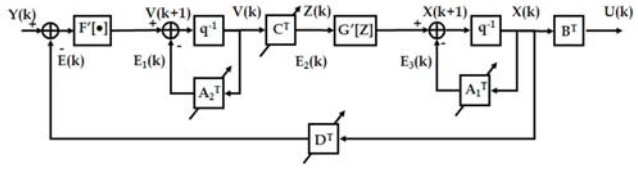


Figure 2: Block-diagram of the adjoint KFRNN topology.

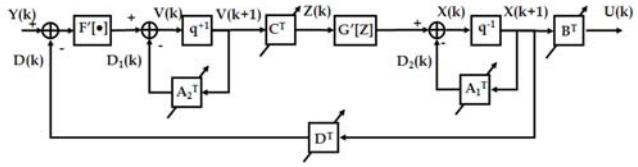


Figure 3: Block-diagram of the adjoint KFRNN used for the L-M algorithm.

Jacobean matrices with appropriate dimensions. The diagonal elements of these matrices are derivatives of the activation functions. The equations (12), (9) represented the weight update of the feedback weight matrices A_1, A_2, as full (nxn), (lxl) matrices. The equation (15) gives the learning solution as an element by element vector product when these matrices are diagonal vA_1, vA_2 which is our case.

2.3 Recursive Levenberg-Marquardt KFRNN learning

The general recursive L-M algorithm of learning, [4, 6], is given by the following equations:

$$W(k+1) = W(k) + P(k)\nabla Y[W(k)]E[W(k)] \quad (16)$$

$$Y[W(k)] = g[W(k), U(k)] \quad (17)$$

$$E^2[W(k)] = \{Y_p(k) - g[W(k), U(k)]\}^2 \quad (18)$$

$$DY[W(k)] = \frac{\partial}{\partial W} g[W, U(k)]\Big|_{W=W(k)} \quad (19)$$

Where: W is a general weight matrix (A_1, A_2, B, C, D), under modification; P is the covariance matrix of the weights estimates, being updated; DY is a nw-dimensional gradient vector; Y is the KFRNN output vector which depended on the updated weights and the input; E is an error vector; Yp is the plant output vector which is in fact the target vector. Using the same KFRNN adjoint block diagram (see Fig.2), we could obtain the values of the gradients DY for each updated weight, propagating the value D(k) = I through it and following the block diagram of Fig. 3, we could apply equation (19) for each element of the weight matrices (A_1, A_2, B, C, D) to be updated. The corresponding gradient components are as follows:

$$DY[C_{ij}(k)] = D_{1,i}(k) Z_j(k) \quad (20)$$

$$DY[A_{2ij}(k)] = D_{1,i}(k) V_j(k) \quad (21)$$

ISBN: 978-989-95079-6-8

$$D_{1,i}(k) = F_i'[Y_i(k)] \tag{22}$$

$$DY[A_{1ij}(k)] = D_{2,i}(k) X_j(k) \tag{23}$$

$$DY[B_{ij}(k)] = D_{2,i}(k) U_j(k) \tag{24}$$

$$DY[D_{ij}(k)] = D_{2,i}(k) Y_j(k) \tag{25}$$

$$D_{2,i}(k) = G_i'[Z_i(k)] C_i D_{1,i}(k) \tag{26}$$

So the Jacobean matrix could be formed as:

$$DY[W(k)] = [DY(C_{ij}(k)), DY(A_{2ij}(k)), DY(B_{ij}(k)) \\ DY(A_{1ij}(k)), DY(D_{ij}(k))] \tag{27}$$

The P(k) matrix is computed recursively by the equation:

$$P(k) = \alpha^{-1}(k)\{P(k-1) - \\ -P(k-1)\Omega[W(k)] S^{-1}[W(k)] \Omega^T[W(k)] P(k-1)\} \tag{28}$$

Where: the S(.), and Ω(.) matrices are given as follows:

$$S[W(k)] = \alpha(k)\Lambda(k) + \Omega^T[W(k)] P(k-1)\Omega[W(k)] \tag{29}$$

$$\Omega^T[W(k)] = \begin{bmatrix} & \nabla Y^T[W(k)] & \\ 0 \quad \cdots & 1 & \cdots \quad 0 \end{bmatrix}; \\ \Lambda(k)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix}; 10^{-4} \le \rho \le 10^{-6}; \\ 0.97 \le \alpha(k) \le 1; 10^3 \le P(0) \le 10^6 \tag{30}$$

The matrix Ω(.) has dimension (nwx2), where the second row has only one unity element (the others are zero). The position of that element is computed by:

$$i = k \bmod (nw) + 1; k > nw \tag{31}$$

Next the given up topology and learning are applied for CSTR system identification and control.

3 Indirect Adaptive Control Scheme (Sliding Mode Control)

The indirect adaptive control using the RTNN as plant identifier has been described in, [11]. Later the proposed indirect control has been derived as a Sliding Mode Control (SMC) and applied for control of unknown hydrocarbon biodegradation processes, [12]. The block diagram of the indirect adaptive control scheme is shown on Fig. 4. It contains identification and state estimation KF RNN and a sliding mode controller. The stable nonlinear plant is identified by a KFRNN model with topology, given by equations (1)-(6) learned by the stable BP-learning algorithm, given by equations (7)-(15), or using the second order LM-learning algorithm, given by equations (16)-(31). The simplification and linearization of the neural identifier equations (1)-(2), omitting the DY(.) term, leads to the next

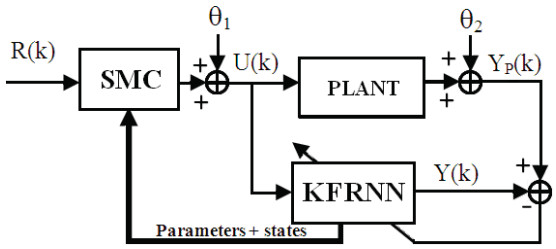


Figure 4: Block diagram of the closed-loop system containing KF RNN identifier and a SMC.

local linear plant model, extracted from the complete KF RNN model:

$$X(k+1) = A_1 X(k) + BU(k) \tag{32}$$

$$Z(k) = HX(k); \; H = CG'(Z) \tag{33}$$

Where G'(.) is the derivative of the activation function and L = M, is supposed.

In [13], the sliding surface is defined with respect to the state variables, and the SMC objective is to move the states form an arbitrary space position to the sliding surface in finite time. In [14], the sliding surface is also defined with respect to the states but the states of the SISO systems are obtained from the plant outputs by differentiation. In [15], the sliding surface definition and the control objectives are the same. The equivalent control systems design is done with respect to the plant output, but the reachability of the stable output control depended on the plant structure. In [12], the sliding surface is derived directly with respect to the plant outputs which facilitated the equivalent SMC systems design. Let us define the following SS as an output tracking error function:

$$S(k+1) = E(k+1) + \sum_{i=1}^{P} \gamma_i E(k-i+1); \; |\gamma| < 1 \tag{34}$$

Where: S(.) is the Sliding Surface Error Function (SSEF) defined with respect to the plant output; E(.) is the systems output tracking error; γ_i are parameters of the desired stable SSEF; p is the order of the SSEF. The tracking error and their iterate are defined as:

$$E(k) = R(k) - Z(k); E(k+1) = R(k+1) - Z(k+1) \tag{35}$$

Where R(k), Z(k) are L-dimensional reference and output vectors of the local linear plant model. The objective of the sliding mode control systems design is to find a control action which maintains the systems error on the sliding surface which assure that the output tracking error reaches zero in P steps, where P < N. So, the control objective is fulfilled if:

$$S(k+1) = 0 \tag{36}$$

ISBN: 978-989-95079-6-8

Now, let us to iterate (33) and to substitute (32) in it so to obtain the input/output local plant model, which yields:

$$Z\left(k+1\right) = FX\left(k+1\right) = F\left[AX\left(k\right) + BU\left(k\right)\right] \quad (37)$$

From (34), (35), and (36) it is easy to obtain:

$$R\left(k+1\right) - Z\left(k+1\right) + \sum_{i=1}^{P} \gamma_i E\left(k-i+1\right) = 0 \quad (38)$$

The substitution of (37) in (38) gives:

$$R\left(k+1\right) - FAX\left(k\right) + FBU\left(k\right) - \sum_{i=1}^{P} \gamma_i E\left(k-i+1\right) \quad (39)$$

As the local approximation plant model (32), (33), is controllable, observable and stable (see [5]), the matrix A_1 is diagonal, and L = M, then the matrix product (HB), representing the plant model static gain, is nonsingular, and the plant states X(k) are smooth non-increasing functions. Now, from (39) it is easy to obtain the equivalent control capable to lead the system to the sliding surface which yields:

$$U_{eq}\left(k\right) = \left(FB\right)^{-1}\left[-FAX\left(k\right) + R\left(k+1\right) + \sum_{i=1}^{P} \gamma_i E\left(k-i+1\right)\right] \quad (40)$$

Following [13], the SMC avoiding chattering is taken using a saturation function instead of sign one. So the SMC takes the form:

$$U^*(k) = \begin{cases} U_{eq}\left(k\right) & \text{if } \left\|U_{eq}\left(k\right)\right\| < U_0 \\ -U_0 U_{eq}\left(k\right) / \left\|U_{eq}\left(k\right)\right\| & \text{if } \left\|U_{eq}\left(k\right)\right\| \geq U_0 \end{cases} \quad (41)$$

The SMC substituted the multi-input multi-output coupled high order dynamics of the linearized plant with desired decoupled low order one.

4 Description of the CSTR bioprocess

The CSTR model given in [16, 17] was chosen as a realistic example for application of the KFRNN and the SMC for solution of system identification and control problems. The CSTR is described by the following continuous time, nonlinear, system of ordinary differential equations:

$$\frac{dC_A(t)}{dt} = \frac{Q}{V}\left(C_{Af} - C_A(t)\right) - k_0 C_A(t) \exp\left(-\frac{E}{RT(t)}\right) \quad (42)$$

$$\frac{dT(t)}{dt} = \frac{Q}{V}(T_f - T(t)) + \frac{(-\Delta H)C_A(t)}{\rho C_p}\exp\left(\frac{E}{RT(t)}\right) + \frac{\rho_c C_{pc}}{\rho C_p V} Q_c(t)\left[1 - \exp\left(\frac{-hA}{Q_c(t)\rho_c C_{pc}}\right)\right](t_{ef} - T(t)) \quad (43)$$

It suffices to know that within the CSTR two chemicals are mixed and react to produce a product compound A at a concentration CA(t), with the temperature of the mixture being T(t). The reaction is exothermic and producing heat which slows down the reaction. By introducing a coolant flow-rate Qc(t), the temperature can be varied and hence the product concentration controlled. CAf is the inlet feed concentration, Q is the process flow-rate, Tf and Tef are the inlet feed and coolant temperatures, respectively, all of which are assumed constant at nominal values. Likewise k0, E/R, V, ΔH, ρ, Cpc, Cp and ρc are thermodynamic and chemical constant relating to this particular problem. Numerical values for the parameters and nominal operating conditions of this model are given in Table 1.

The quantities Qc0, T0, and CA0 shown in Table 1 are steady values for steady operating point in the CSTR. The objective is to control the product compound A, manipulating the variable $Q_c(t)$. The operating values of the variables are obtained from the papers [16], [17]. The authors of these papers examined the performances of neural networks and of fuzzy-neural networks based control systems, respectively.

5 Simulation Results

The proposed indirect adaptive SMC is applied for the given up CSTR plant. The Fig. 5 and Fig. 7 showed the comparison between the reference and the plant output signals applying the BP and L-M algorithms of learning for the KFRNN, respectively. Detailed comparative graphical simulation results of CSTR KFRNN plant identification by means of the BP and the L-M learning are given in Fig.6 and Fig.8. A 10% white noise is added to the plant inputs and outputs and the behavior of the plant identification and control has been studied accumulating some statistics of the final MSE% (ξ_{av}) for BP and L-M learning, which results are given in Table 3 and Table 4 for 20 runs. The mean average cost for all runs (ε) of control and the standard deviation (σ) with respect to the mean value and the deviation (Δ) are presented in Table 2 for BP and L-M algorithms of learning.

Table 1: CSTR parameters and operating conditions

$Q = 100 \quad (L/\min)$	$E/R = 9.95\times10^3 \quad (K)$
$C_{Af} = 1.0 \quad (mol/L)$	$-\Delta H = 2\times10^5 \quad (cal/mol)$
$T_f = 350 \quad (K)$	$\rho\cdot\rho_c = 1000 \quad (g/L)$
$T_{cf} = 350 \quad (K)$	$C_p C_{pc} = 1 \quad (cal/gK)$
$V = 100 \quad (L)$	$Q_{c0} = 103.41 \quad (L/\min)$
$hA = 7\times10^5 \quad (cal/\min K)$	$T_0 = 440.2 \quad (K)$
$k_0 = 7.2\times10^{10} \quad (1/\min)$	$C_{A0} = 0.0836 \quad (mol/L)$

Table 2: Standard deviation and mean average values using BP and L-M algorithms of learning.

BP Algorithm	LM Algorithm
ε = 1.0997	ε =1.0370
σ = 0.0295	σ =0.0232

ISBN: 978-989-95079-6-8

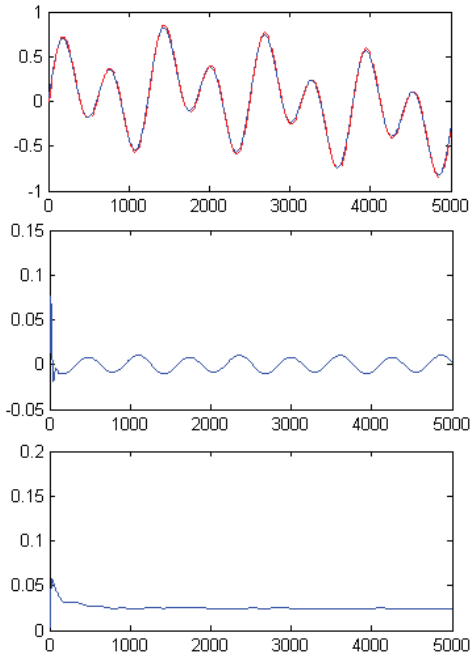


Figure 5: Detailed graphical simulation results of CSTR plant Sliding Mode Indirect Control using BP KFRTNN learning a) comparison between the plant output and the reference signal; b) control signal; c) MSE% of control.

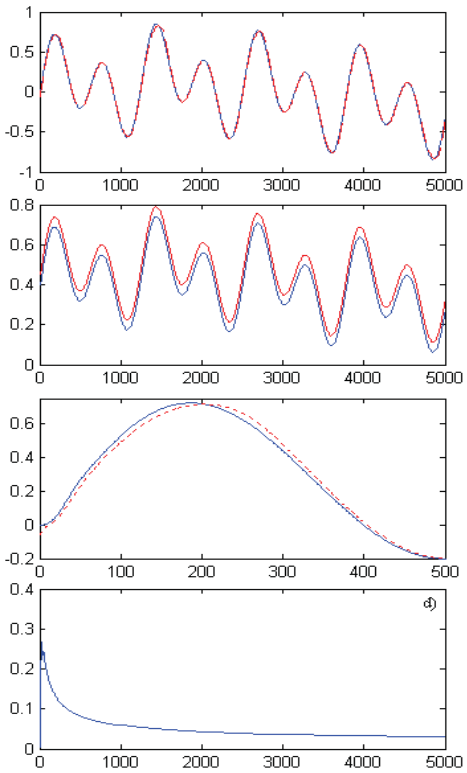


Figure 6: Graphical results of identification using BP KFRNN learning. a) Comparison of the plant output (continuous line) and KFRNN output (pointed line); b) state variables; c) comparison of the plant output (continuous line) and KFRNN output (pointed line) in the first instants; d) MSE% of identification

The values are computed using the following formulas:

$$\varepsilon = \frac{1}{n}\sum_{k=1}^{n}\xi_{av_k}, \quad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\Delta_i^2}, \quad \Delta = \xi_{av} - \varepsilon \tag{44}$$

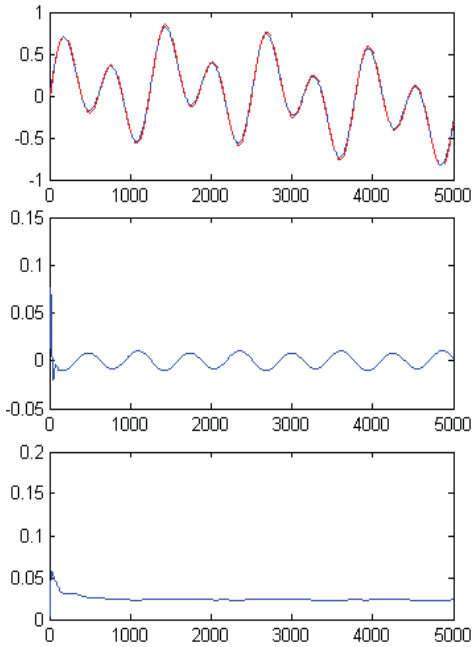


Figure 7: Detailed graphical simulation results of CSTR plant Sliding Mode Indirect Control using L-M KFRTNN learning a) comparison between the plant output and the reference signal; b) control signal; c) MSE% of control.

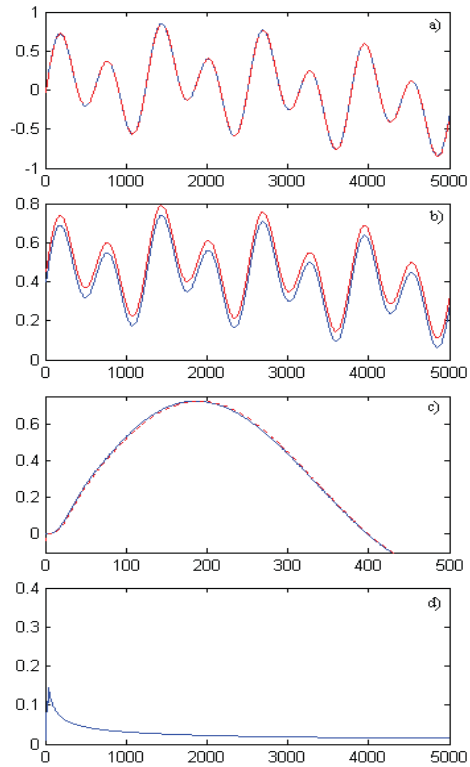


Figure 8: Graphical results of identification using L-M KFRNN learning. a) Comparison of the plant output (continuous line) and KFRNN output (pointed line); b) state variables; c) comparison of the plant output (continuous line) and KFRNN output (pointed line) in the first instants; d) MSE% of identification

ISBN: 978-989-95079-6-8

Table 3: Final MSE% of control for 20 runs of control program using the Backpropagation algorithm.

No.	**1**	**2**	**3**	**4**	**5**
MSE%	1.1096	1.1530	1.1331	1.0660	1.1375
No.	**6**	**7**	**8**	**9**	**10**
MSE%	1.0701	1.1171	1.1202	1.1311	1.0899
No.	**11**	**12**	**13**	**14**	**15**
MSE%	1.0964	1.1007	1.1139	1.0685	1.0502
No.	**16**	**17**	**18**	**19**	**20**
MSE%	1.1330	1.1039	1.0724	1.0609	1.0665

Table 4: Final MSE% of control for 20 runs of control program using the Levenberg-Marquardt algorithm.

No.	**1**	**2**	**3**	**4**	**5**
MSE%	1.0296	1.0462	1.0472	1.0670	1.0134
No.	**6**	**7**	**8**	**9**	**10**
MSE%	1.0078	1.0396	1.0678	1.0017	1.0609
No.	**11**	**12**	**13**	**14**	**15**
MSE%	1.0019	1.0364	1.0135	1.0501	1.0175
No.	**16**	**17**	**18**	**19**	**20**
MSE%	1.0654	1.0096	1.03652	1.0627	1.0660

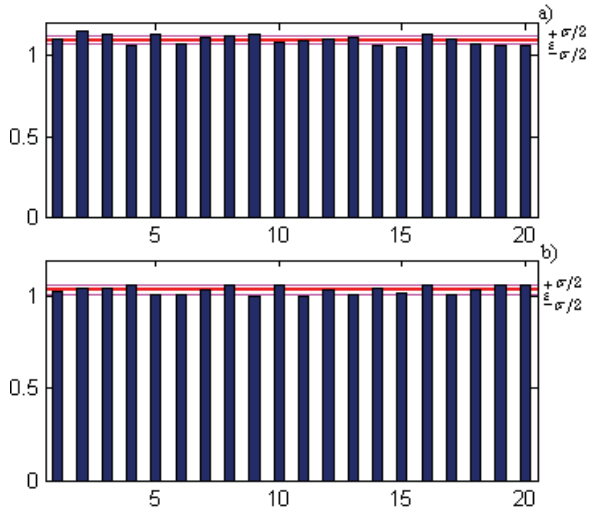


Figure 9: Comparison between the final MSE% for 20 runs of control program: a) using BP algorithm, b) using L-M algorithm.

6 Conclusions

The paper proposed a new KFRNN model for systems identification and states estimation of nonlinear plants. The KFRNN is learned by the second order recursive learning algorithm of Levenberg-Marquardt. The estimated parameters and states of the RNN model are used for indirect (SM) adaptive trajectory tracking control systems design. The applicability of the proposed neural control system is confirmed by simulation results with a CSTR plant. The results showed good convergence of both L-M and BP learning algorithms. The L-M algorithm of learning is more precise (see Table 2) but more complex then the BP one.

Acknowledgment

The Ph.D. student Carlos-Roman Mariaca-Gaspar is thankful to CONACYT, Mexico for the fellowship received during his studies at CINVESTAV-IPN, Mexico.

References

[1] K.S. Narendra, and K. Parthasarathy, Identification and Control of Dynamic Systems Using Neural Networks, *IEEE Transactions on Neural Networks,* 1(1): 4 -27, 1990.

[2] K.J. Hunt, D. Sbarbaro, R. Zbikowski, and P.J. Gawthrop, Neural Network for Control Systems - A Survey, *Automatica*, 28(6): 1083-1112, 1992.

[3] S.A. Pao, S.M. Phillips, and D.J. Sobajic, Neural Net Computing and Intelligent Control Systems, *International Journal of Control*, 56: 263-289, 1992.

[4] I. S. Baruch, S. Escalante, and C. R. Mariaca-Gaspar, Identification, Filtering and Control of Nonlinear Plants by Recurrent Neural Networks Using First and Second Order Algorithms of Learning, *Dynamics of Continuous, Discrete and Impulsive Systems, International Journal, Series A: Mathematical Analysis, Special Issue on Advances in Neural Networks-Theory and Applications*, Watam Press, Canada, ISSN 1201-3390, 14: 512-521, 2007.

[5] I.S. Baruch, C.R. Mariaca-Gaspar, and J. Barrera-Cortes, Recurrent Neural Network Identification and Adaptive Neural Control of Hydrocarbon Biodegradation Processes, *In: Recurrent Neural Networks, Xiaolin Hu, and P. Balasubramaniam (Eds.)*, ISBN 978-3-902613-28-8, I-Tech Education and Publishing KG, Vienna, Austria, Chapter 4: 61-88, 2008.

[6] L.S. Ngia, and J. Sjöberg, Efficient Training of Neural Nets for Nonlinear Adaptive Filtering Using a Recursive Levenberg-Marquardt Algorithm, *IEEE Transactions on Signal Processing*, 48: 1915-1927, 2000.

[7] L. D. Schmidt, *The Engineering of Chemical Reactions,* New York: Oxford University Press, ISBN 0-19-510588-5, 1998.

[8] Shi-Rong Liu, and Jin-Shou Yu, Robust Control Based on Neuro Fuzzy Systems for a Continuous Stirred Tank Reactor, *In: Proc. of the First International Conference on Machine Learning and Cybernetics,* Beijing, 4-5 November 2002, 1483-1488, 2002.

[9] T. Zhang, and M. Guay. Adaptive Nonlinear Control of Continuously Stirred Tank Reactor Systems, *In: Proc. of the American Control Conference, Arlington*, 25-27 June 2001, 1274-1279, 2001.

[10] E. Wan, and F. Beaufays, Diagrammatic Method for Deriving and Relating Temporal Neural Networks Algorithms, *Neural Computations*, 8: 182-201, 1996.

[11] I.S. Baruch, J.M.Flores, F. Thomas, and R. Garrido, Adaptive Neural Control of Nonlinear Systems, *In: Artificial Neural Networks-ICANN 2001, LNCS 2130, G. Dorffner, H. Bischof, K. Hornik (Eds.)*, ISBN 3-540-42486-5, Springer, Berlin, 930-936, 2001.

[12] I.S. Baruch, C.R. Mariaca-Gaspar, I. Cruz-Vega, and J. Barrera-Cortes, Sliding Mode Control of a Hydrocarbon Degradation in Biopile System Using Recurrent Neural Network Model, *In: MICAI 2007: Advances in Artificial Intelligence, LNAI 4827, A. Gelbukh, A.F. Kuri-Morales (Eds.)*, ISBN-10 3-540-76630-8, Springer, Berlin Heidelberg New York, 1184-1194, 2001.

[13] K. D. Young, V.I. Utkin, and U. Ozguner, A Control Engineer's Guide to Sliding Mode Control, *IEEE Transactions on Control Systems Technology*, 7 (3): 328-342, 1999.

[14] A. Levent, Higher Order Sliding Modes, Differentiation and Output Feedback Control, *International Journal of Control, Special Issue Dedicated to Vadim Utkin on the Occasion of his 65th Birthday (Guest editor: Leonid M. Fridman),* ISSN 0020-7179, 76(9/10): 924-941, 15 June-10 July 2003.

[15] C. Eduards, S.K. Spurgeon, and R.G. Hebden, On the Design of Sliding Mode Output Feedback Controllers, *International Journal of Control, Special Issue Dedicated to Vadim Utkin on the Occasion of his 65th Birthday (Guest editor: Leonid M. Fridman),* ISSN 0020-7179, 76(9/10): 893-905, 15 June-10 July 2003.

[16] G. Lightbody, and G. W. Irwin, Nonlinear Control Structures Based on Embedded Neural Systems Models, *IEEE Transactions on Neural Networks*, 8: 553-557, 1997.

[17] A. Aoyama, F. J. Doyle, and V. Venkatasubramanian, Control-Affine Fuzzy Neural Networks Approach for Nonlinear Process Control, *Journal of Process Control*, 8: 375-386, 1995.

A Multi-granular Linguistic Promethee Model

Nesrin Halouani[1], Luis Martínez[2], Habib Chabchoub[1], Jean-Marc Martel[3], Jun Liu[4]

[1] University of Economic Sciences and Management, Sfax, Tunisia,
[2] University of Jaén, Spain,
[3] University of Laval, Canada,
[4] University of Ulster, Nothern Ireland, UK
{halouani.nesrin@gmail.com, martin@ujaen.es}

Abstract— *In Multi-criteria Decision Making (MCDM) problems dealing with qualitative criteria and uncertain information the use of linguistic values is suitable for the experts in order to express their judgments. It is common that the group of experts involved in such problems have different degrees of knowledge about the criteria, so we propose a multi-granular linguistic framework such that each expert can provide his/her evaluations in different linguistic term sets according to his/her knowledge. MCDM problems have been solved in the literature by using different methods, in this contribution we focus on PROMETHEE method and our proposal consists of developing tools and operators for the PROMETHEE method to deal with multi-granular linguistic information.*

Keywords— Multi-criteria decision making, linguistic hierarchies, Promethee.

1 Introduction

Decision-making is a common human activity and its multidimensional nature of real world decision problems is well addressed by multi-criteria decision aid (MCDA). The focal point of interest within the methodological framework of MCDA is the analysis and the modelling of the multiple decision makers' preferences. This special characteristic of MCDA implies that a comprehensive model of a decision situation cannot be developed, but instead the model should be developed to meet the expert's requirements. However, sometimes tradeoffs between some criteria may also be too difficult to define for the experts, and they can be then reluctant to express any measurable opinions.

Different methods belonging to MCDA are the following [1]: the *Electre* family developed by Roy and his co-workers, *Promethee* (Brans, Mareshal, Vincke), *Oreste* (Pastijin and Leyson), *Melchior* (Leclerq), *Qualifex* (Paelink), *Regime* (Hinloopen, Nijkamp, Rietvald), *Macbetch* (Bana e Costa, Vansnick), *Ahp* (Saaty), *Topsis* (Hwang and Yoon). Often these methods require a group of experts to express their preferences over the criteria involved in the decision process.

In the real-world, many decision problems are characterized by two overarching concerns: to consider conflict between the criteria of the problem, and to take into account the uncertainty inherent in decision making that depends on the outcome of unknown future events.

To deal with these concerns, the Promethee method is used. Promethee [2] is a popular decision method that has been successfully applied in the selection of the final solution of a problem. It generates a ranking of available alternatives, according to the expert's preferences, and the best ranked one is considered the favourite final solution.

In this paper, we focus on decision under uncertainty because is one of the most frequent situations in practical decision making, namely in planning activities in many fields. Traditional studies of such issues are conducted by using probabilistic tools and techniques. However, it is not difficult to see in many problems that aspects related to imprecision or vagueness clearly have a non probabilistic character since they are related to imprecision of meanings. Usually, when we deal with certain knowledge in a quantitative setting the information provided by the experts is expressed by means of numerical precise values. However, when we work in a qualitative setting, that is, with vague or imprecise knowledge, it could not be estimated with an exact numerical value. Then, a more realistic approach may be to use linguistic assessments instead of numerical values [11].

The use of linguistic variables makes experts' evaluations more flexible and reliable, but implies processes of computing with words (CW). The main problem that presents the traditional linguistic approaches to carry out the CW processes is the loss of information and hence a lack of precision in the final results. Different linguistic computational models have been developed Semantic model [4], Symbolic [5] or the 2-tuple one [3] that provides a model to deal with CW processes in a precise way.

Our aim in this contribution is focused on MCDM problems where different experts can have different degree of knowledge about the criteria so they can use different linguistic term sets to provide their information defining a multi-granular linguistic context. Again, the main problem is to carry out the CW processes in such a context, in order to overcome this drawback, Herrera and Martinez [6] have developed a model based on linguistic hierarchies. Thus, the CW processes in such contexts can be carried out without loss of information.

Accordingly, a flexible and realistic multi-granular hierarchical linguistic approach based on Promethee method is presented in this paper. The main advantage of this approach is to tackle the uncertainty of both performance of criteria and experts' knowledge without loss of information.

The structure of the paper is the following one. Basic concepts about Promethee method are introduced in section 2. A brief linguistic background is presented in section 3. An aggregation process for multi-granular linguistic information in PROMETHEE is proposed in section 4. In section 5, we

ISBN: 978-989-95079-6-8

apply this approach to an example. The conclusions are pointed out in section 6.

2 The Promethee Method

The PROMETHEE method (Preference Ranking Organization METHod for Enrichment Evaluation) is a multicriteria decision-making method, belonging to the family of outranking methods [2]. It is a ranking method quite simple in conception and application compared to other methods for multicriteria analysis. It is well adapted to problems where a finite number of alternatives are ranked considering several conflicting criteria. The evaluation table is the starting point of this method. In this table, the alternatives are evaluated according to different criteria. The implementation of Promethee requires two additional types on information, namely: information about the relative importance, w_j, (i.e. the weights) of the criteria considered, and information on the expert's preference modeling, which it uses when comparing the contribution of the alternatives in terms of each separate criterion.

The Promethee method encompasses two phases: (i) the aggregation of information about the alternatives and the criteria, (ii) the exploitation of the outranking relation for decision aid.

The aggregation phase requires that each point of view would be associated with a generalized criterion to assess the preference for an alternative a_i with regards to a_k as a function of $P_j(a_i,a_k) = H_j\left(f_j(a_i) - f_j(a_k)\right)$. A generalized criterion is thus a function $H_j\left(f_j(a_i) - f_j(a_k)\right)$ which is null when $\left(f_j(a_i) - f_j(a_k)\right)$ is negative, non-decreasing with $\left(f_j(a_i) - f_j(a_k)\right)$ varying between 0 and 1. Six different types of generalized criteria (for a further description see [2]) are proposed to experts, in each case at most two parameters from these thresholds q, p and s have to be fixed. Indifference threshold, q, is the largest deviation to consider as negligible on that criterion. It is a small value with regards to the scale of measurement. Preference threshold, p, is the smallest deviation to consider decisive in the preference of one alternative over another. It is a large value with respect to the scale of measurement. Gaussian threshold, s, is only used with the Gaussian preference function. It is usually fixed as intermediate value between an indifference and a preference threshold.

The outranking relation can be then represented by an oriented graph. The value of each arc is the multi-criteria preference index $\pi(.,.)$, which is defined for all ordered pairs of alternatives. These indices that may take any value in the interval $[0,1]$ define a fuzzy outranking relation. For each $(a_i,a_k) \in A \times A$, Promethee permits the computation of the following quantities for alternatives a_i and a_k:

$$\pi(a_i,a_k) = \frac{\sum_{j=1}^{n} w_j P_j(a_i,a_k)}{\sum_{j=1}^{n} w_j} \quad (1)$$

$$\phi^+(a_i) = \sum_{a_k \in A} \pi(a_i,a_k), \quad (2)$$

$$\phi^-(a_i) = \sum_{a_k \in A} \pi(a_k,a_i),$$

$$\phi(a_i) = \phi^+(a_i) - \phi^-(a_i)$$

For each alternative a_i, belonging to the set A of alternatives, $\pi(a_i,a_k)$ is an overall preference index of a_i over a_k. The leaving flow $\phi^+(a_i)$ defines the strength of the alternative a_i, how much a_i dominates all the other alternatives of A. Symmetrically, the entering flow $\phi^-(a_i)$ defines the weakness of the alternative, how much a_i is dominated by all the other alternatives of A. $\phi(a_i)$ represents a value function, whereby a higher value reflects a higher attractiveness of alternative a_i. We call $\phi(a_i)$ the net flow of alternative a_i.

According to Promethee I, alternative a_i is better than a_k if the leaving flow of a_i ($\phi^+(a_i)$) is greater than the leaving flow of a_k ($\phi^+(a_k)$) and the entering flow of a_i ($\phi^-(a_i)$) is smaller than the entering flow of a_k ($\phi^-(a_k)$).

Equality in $\phi^+(a_i)$ and $\phi^-(a_i)$ indicates indifference between the two alternatives. In the case where the leaving flows indicate that a_i is better than a_k, while the entering flows indicate the reverse, a_i and a_k are considered incomparable. Therefore, the Promethee I provide a partial ranking of the alternatives.

In Promethee II, the net flow $\phi(a_i)$ is used in order to obtain a complete ranking of all alternatives. The alternative with the higher net flow is better.

3 Linguistic Background

Due to the fact that, our proposal consists in dealing with MCDM problems defined in multi-granular linguistic contexts that implies processes of CW, here we review briefly the 2-tuple linguistic representation model and the linguistic hierarchies structure that are necessary concepts to achieve our aim.

3.1 The 2-tuple Fuzzy Representation Model

This model was presented in [3] for overcoming the drawback of the loss of information presented by the classical linguistic computational models [12], i.e., (i) the semantic model [4], and (ii) the symbolic one [5].

The 2-tuple fuzzy linguistic representation model is based

ISBN: 978-989-95079-6-8

on the symbolic method and takes as the base of its representation the concept of Symbolic Translation.

Definition 1. *The Symbolic Translation of a linguistic term* $s_i \in S = \{s_0,...,s_g\}$ *is a numerical value assessed in [-0.5, 0.5) that supports the "difference of information" between an amount of information* $\beta \in [0, g]$ *and the closest value in {0,..., g} that indicates the index of the closest linguistic term in S (*s_i*), being [0,g] the interval of granularity of S.*

From this concept a new linguistic representation model is developed, which represents the linguistic information by means of 2-tuples (s_i,α_i), $s_i \in S$ and $\alpha_i \in$ [-0.5, 0.5).

Definition 2. *Let* $S = \{s_0,...,s_g\}$ *be a linguistic term set and* $\beta \in [0,g]$ *a value supporting the result of a symbolic aggregation operation. Then the 2-tuple that expresses the equivalent information to* β *is obtained with the following function:*

$$\Delta : [0, g] \to S \times [-0.5,.0.5) \qquad (3)$$

$$\Delta(\beta) = (s_i, \alpha), \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-0.5, 0.5) \end{cases}$$

where round(·) is the usual round operation, s_i has the closest index label to "β" and "α" is the value of the symbolic translation.

Proposition 1. Let $S = \{s_0,...,s_g\}$ be a linguistic term set and (s_i, α_i) be a linguistic 2-tuple. There is always a Δ^{-1} function, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in$ [0, g] in the interval of granularity of *S*.

Proof. It is trivial, we consider the following function:

$$\Delta^{-1} : S \times [-0,5, 0.5) \to [0, g] \qquad (4)$$

$$\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$$

Remark 1. From Definitions 1 and 2 and Proposition 1, it is obvious that the conversion of a linguistic term into a linguistic 2-tuple consists of adding a value 0 as symbolic translation: $s_i \in S \Rightarrow (s_i, 0)$.

This model has a computational technique based on the 2-tuples [3]:

- Aggregation of 2-tuples
  The aggregation of linguistic 2-tuples consist of obtaining a value that summarizes a set of values, therefore, the result of the aggregation of a set of 2-tuples must be a linguistic 2-tuple. We can find several 2-tuple aggregation operators in [5] based on classical aggregation operators as the arithmetic mean and weighted mean operators.
- Comparison of 2-tuples
  The comparison of information represented by 2-tuples is carried out according to an ordinary lexicographic order.
  Let (s_k, α_1) and (s_l, α_2) be two 2-tuples represented two assessments:
  - If $k < l$ then (s_k, α_1) is smaller than (s_l, α_2);
  - If $k=l$ then
    1) If $\alpha_1 = \alpha_2$ then (s_k, α_1) and (s_l, α_2) represent the same value;
    2) If $\alpha_1 < \alpha_2$ then (s_k, α_1) is smaller than (s_l, α_2);
    3) If $\alpha_1 > \alpha_2$ then (s_k, α_1) is bigger than (s_l, α_2).
- Negation Operator of a 2-tuple
  The negation operator over 2-tuples is defined as:

$$Neg(s_i, \alpha) = \Delta\ (g - \Delta^{-1}(s_i, \alpha)) \qquad (5)$$

where g+1 is the cardinality of *S*.

3.2 Linguistic Hierarchies

The Linguistic Hierarchies were introduced in [6] in order to accomplish processes of CW with multi-granular linguistic information in a precise way. A Linguistic Hierarchy is a set of levels, where each level represents a linguistic term set with different granularity to the remaining levels. Each level is denoted as *l*(*t, n*(*t*)):

- *t* a number that indicates the level of the hierarchy
- *n*(*t*) the granularity of the term set of the level *t*

We assume that levels containing linguistic terms are triangular shaped, symmetrical and uniformly distributed. In addition, the linguistic term sets have an odd number of linguistic terms being the middle one the value of indifference.

The levels belonging to a linguistic hierarchy are ordered according to their granularity, i.e., for two consecutive levels *t* and *t*+1, *n*(*t*+1) > *n*(*t*). Therefore, the level *t*+1 is a refinement of the previous level *t*.

From the above concepts, we define a linguistic hierarchy, LH, as the union of all levels *t*:

$$LH = \bigcup_t l(t, n(t)) \qquad (6)$$

Given an LH, we denote as $S^{n(t)}$ the linguistic term set of LH corresponding to the level *t* of LH characterized by a granularity of uncertainty *n*(*t*):

$$S^{n(t)} = \{s_0^{n(t)},...,s_{n(t)-1}^{n(t)}\} \qquad (7)$$

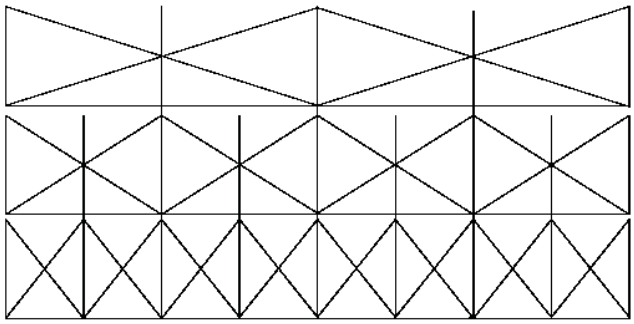
Generically, we can say that the linguistic term set of level *t* + 1 is obtained from its predecessor as:

$$l(t, n(t)) \to l(t+1, 2 \bullet n(t) - 1) \qquad (8)$$

A graphical example of a linguistic hierarchy can be seen in Figure 1.

ISBN: 978-989-95079-6-8

Fig. 1. Linguistic Hierarchy with term sets of 3,5 and 9 terms.



In [6] different transformation functions between labels of different levels were developed without loss of information. To understand how these functions are working, there were defined transformation functions between two consecutive levels and afterwards between any levels of the hierarchy, those transformation functions use the linguistic 2-tuple computational model. Here, we present the transformation function between any levels.

Definition 3. *Let* $LH = \bigcup_t l(t, n(t))$ *be a linguistic hierarchy whose linguistic term sets are denoted as* $S^{n(t)} = \{s_0^{n(t)}, ..., s_{n(t)-1}^{n(t)}\}$, *and let us consider the 2-tuple linguistic representation. The transformation function from a linguistic label in level t to a label in level t' is defined as*:

$$TF_{t'}^{t} : l(t, n(t)) \to l(t', n(t')) \quad \textbf{(9)}$$

$$TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)}) = \Delta_{n(t')}\left(\frac{\Delta_{n(t)}^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \bullet (n(t')-1)}{n(t)-1}\right)$$

Proposition 2. *The transformation function between linguistic terms in different levels of the linguistic hierarchy is bijective:*

$$TF_{t}^{t'}(TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)})) = (s_i^{n(t)}, \alpha^{n(t)}) . \quad \textbf{(10)}$$

4 A Linguistic Multi-Granular Promethee model

The model developed in this paper is the result of the integration between the aggregation operators of the PROMETHEE method and the multi-granular linguistic model to combine multiple experts' assessments defining a multi-granular linguistic framework. So, each expert can express his/her evaluations in the suitable scale according to his/her knowledge in the table 1.

Table 1. Multiple experts' assessments scheme.

				Experts					
		e_1			e_l			e_E	
Alt.	c_1	c_i	c_n	c_1	c_i	c_n	c_1	c_i	c_n
a_1	C_{11}^1	C_{1j}^1	C_{1n}^1	C_{11}^2	C_{1j}^2	C_{1n}^2	C_{11}^E	C_{1j}^E	C_{1n}^E
a_j	C_{j1}^1	C_{ij}^1	C_{jn}^1	C_{j1}^2	C_{ij}^2	C_{jn}^2	C_{j1}^3	C_{ij}^3	C_{jn}^3
a_m	C_{m1}^1	C_{mj}^1	C_{mn}^1	C_{m1}^2	C_{mj}^2	C_{mn}^2	C_{m1}^3	C_{mj}^3	C_{mn}^3

Being $A = \{a_1, ..., a_j, ..., a_m\}$ a set of alternatives, $C = \{c_1, ..., c_i, ..., c_n\}$ a set of criteria, $E = \{e_1, ..., e_l, ..., e_E\}$ a set of experts. The assessments C_{ij}^e provided by the experts, e_e, can be assessed in a linguistic term sets of the linguistic hierarchy that can have different granularity of uncertainty. Therefore $C_{ij}^e \in S^e$ and $S^e \in LH$.

The proposed model to deal with MCDM problems defined in multi-granular linguistic contexts consists of three steps. Due to the use of multi-granular linguistic information, the aggregation step is divided in two steps:

- *Normalization step*. The multi-granular linguistic information is expressed in a unique linguistic expression domain.
- *Aggregation step*. The unified information expressed in a unique linguistic term set is aggregated.
- *Exploitation step*. The collective preference values are ordered in a decreasing way and the solution set is composed of the best alternative/s.

4.1 Normalization step

At the beginning, a linguistic term set to unify the multi-granular linguistic information must be selected called S_T. Any linguistic term set to do it can be chosen because the transformations between levels in a LH are bijective (see Proposition 2). In order to reduce the number of computations, the linguistic term set that the most of experts express their preferences in it shall be chosen.

Let us suppose that $S^e = S^{n(t')}$ and $S_T = S^{n(t'')}$, so a transformation function between linguistic terms in different levels of the linguistic hierarchy is obtained as follows:

$$TF_{t''}^{t'}(C_{ij}^e) = (s_{ij}^e, \alpha) . \quad \textbf{(11)}$$

with, $s_{ij}^e \in S_T$.

4.2 Aggregation step

In this step, two types of preference indexes, individual ($\pi^e(a_i, a_k)$) and collective ($\pi(a_i, a_k)$), will be computed. The individual preference index which translates the intensity of the preference of the alternative a_i compared to the alternative a_k according to the point of view of each expert is expressed as:

$$\pi^e : A \times A \to S^{n(t)} \times [-0.5, 0.5) \quad \textbf{(12)}$$

$$\pi^e(a_i, a_k) = \Delta\left[\frac{\sum_{j=1}^{n} \Delta^{-1}\left(w_j^e . P_j^e(a_i, a_k)\right)}{\Delta^{-1}(W^e)} . n(t)\right] = \left(s_i^{e,n(t)}, \alpha_i^e\right)$$

w_j^e is the weight assigned by each expert to each criterion, $W^e = \sum_{j=1}^{n} w_j^e$, $P_j^e(a_i, a_k)$ is a preference function among the six types of functions proposed in [2], a_i and a_k are two alternatives belonging to A, $n(t)$ is the cardinality of the chosen linguistic term set, $\left(s_i^{e,n(t)}, \alpha_i^e\right)$ is a linguistic

ISBN: 978-989-95079-6-8

2-tuple, $i,k \in \{1,...,m\}$ alternatives, $j \in \{1,...,n\}$ criteria, $e \in \{1,...,E\}$ experts.

Thus, the collective preference index corresponds to the aggregation of the linguistic values computed. The aggregation operator could be different in each problem. Different aggregation operators were defined in [3] to deal with linguistic 2-tuples, for the sake of simplicity in this case we have chosen the arithmetic mean. The expression of the collective preference index is:

$$\pi : A \times A \to S^{n(t)} \times [-0.5, 0.5) \quad \textbf{(13)}$$

$$\pi(a_i, a_k) = \Delta\left[\sum_{e=1}^{E} \frac{1}{E} \Delta^{-1}\left(\pi^e(a_i, a_k)\right)\right] = \left(s_i^{n(t)}, \alpha_i\right).$$

With, $E = \{e_1,...,e_E\}$ is the set of experts and $\left(s_i^{n(t)}, \alpha_i\right)$ is a 2-tuple.

4.3 Exploitation step

The exploitation step generates a solution set of alternatives (the best ones) for the decision problem. To do so, this step uses a total ranking of the alternatives in a decreasing way according to a choice function. Different choices functions have been proposed in the literature [8-10]. In this paper, a choice function that computes the dominance degree for each alternative, a_i, over the other alternatives is used as follows:

$$\Lambda(a_i) = \frac{1}{m-1} \sum_{k,i=1,k\neq i}^{m} \pi(a_i, a_k). \quad \textbf{(14)}$$

Then, the best alternative(s) are in the head of ranking should be chosen as solution set of alternatives.

5 Numerical Example

In this section, we present an investment example to show the integration between the aggregation operators of PROMETHEE method and the linguistic hierarchies.

5.1 Input Data

An investment company wants to invest a sum of money in the best option. There is a panel with four possible alternatives $A = \{a_1,...,a_4\}$ of investment possibilities. a_1 is a car industry, a_2 is a food company, a_3 is a computer company, and a_4 is an arms industry. The investment company chooses four experts $E = \{e_1,...,e_4\}$ from four consultancy departments: risk analysis, growth analysis, social-political analysis, and environmental impact analysis departments respectively, to construct a decision group throughout a set of three criteria $C = \{c_1, c_2, c_3\}$ being, c_1 profit, c_2 pollution, and c_3 employment.

These experts use different linguistic term sets from the LH (showed in Fig. 1) to provide their preferences over the alternative set as following: e_1 provides his preferences in $l(3,9)$, e_2 provides his preferences in $l(2,5)$, e_3 provides his preferences in $l(1,3)$, and e_4 provides his preferences in $l(3,9)$.

After a deep study, each expert provides the following preference values:

Table 2. Input data of each expert.

		e1			e2			e3			e4	
	c1	c2	c3	c1	c2	c3	c1	c2	c3	c1	c2	c3
a1	S_6^9	S_3^9	S_2^9	S_4^5	S_2^5	S_1^5	S_2^3	S_1^3	S_1^3	S_8^9	S_1^9	S_3^9
a2	S_7^9	S_2^9	S_4^9	S_3^5	S_3^5	S_3^5	S_2^3	S_1^3	S_1^3	S_5^9	S_2^9	S_1^9
a3	S_8^9	S_5^9	S_5^9	S_3^5	S_1^5	S_2^5	S_2^3	S_1^3	S_2^3	S_7^9	S_3^9	S_5^9
a4	S_8^9	S_6^9	S_1^9	S_4^5	S_3^5	S_2^5	S_2^3	S_2^3	S_2^3	S_8^9	S_5^9	S_2^9
wj	S_8^9	S_6^9	S_4^9	S_3^5	S_4^5	S_1^5	S_1^3	S_2^3	S_1^3	S_7^9	S_1^9	S_5^9
Type	II	III	IV	II	III	IV	II	III	IV	II	III	IV
pj	S_4^9	S_2^9	S_5^9	S_3^5	S_3^5	S_2^5	S_2^3	S_1^3	S_2^3	S_5^9	S_2^9	S_4^9
qj	S_2^9	S_1^9	S_3^9	S_1^5	S_2^5	S_1^5	S_1^3	S_0^3	S_1^3	S_3^9	S_1^9	S_3^9

5.2 Normalization Step

In this example, the linguistic term set $l(3,9)$ shall be chosen to unify the multi-granular linguistic information, since the most of experts have expressed their preferences in it.

5.3 Aggregation Step

This step is carried out by the computation of both individual and collective preference indexes. The obtained results are shown in the tables below:

Table 3. Individual preference index of each expert.

π^1	a1	a2	a3	a4
a1	-	$(S_3^9, 0)$	$(S_0^9, 0)$	$(S_0^9, 0)$
a2	$(S_1^9, 0)$	-	$(S_0^9, 0)$	$(S_0^9, 0)$
a3	$(S_8^9, 0)$	$(S_3^9, 0)$	-	$(S_1^9, 0)$
a4	$(S_7^9, 0)$	$(S_3^9, 0)$	$(S_3^9, 0)$	-

Table 4. Collective preference index.

π	a_1	a_2	a_3	a_4
a_1	-	$(S_3^9, -0.24)$	$(S_1^9, 0.4)$	$(S_0^9, 0)$
a_2	$(S_3^9, 0.27)$	-	$(S_3^9, 0.26)$	$(S_0^9, 0.25)$
a_3	$(S_2^9, 0.14)$	$(S_0^9, 0.75)$	-	$(S_3^9, -0.27)$
a_4	$(S_4^9, 0.12)$	$(S_4^9, 0.5)$	$(S_2^9, 0.33)$	-

ISBN: 978-989-95079-6-8

5.4 Exploitation Step

This step provides a total ranking of the alternatives in a decreasing way (Table 6) according to a choice function (Table 5).

Table 5. Dominance degree.

$\Lambda(a_1)=(S_1^9,0.04)$	$\Lambda(a_2)=(S_2^9,-0.3)$
$\Lambda(a_3)=(S_1^9,0.4)$	$\Lambda(a_4)=(S_3^9,-0.3)$

Where the degrees are computed as:

$$\Lambda(a_1)=\frac{1}{3}\sum_{k,i=1,k\neq 1}^{3}\pi(a_1,a_k)=\frac{1}{3}\left(\pi(a_1,a_2)+\pi(a_1,a_3)+\pi(a_1,a_4)\right)$$
$$=\frac{1}{3}\left(\left(S_3^9,-0.24\right)+\left(S_1^9,0.4\right)+\left(S_0^9,0\right)\right)$$
$$=\left(S_1^9,0.4\right)$$

Table 6. Alternatives ranking.

a_4	a_2	a_3	a_1

According to the ranking of the alternatives, the company should choose the alternative, a_4, for its investment.

6 Conclusions

In order to manage multi-granular linguistic information in MCDM problems, we extended aggregation operators of PROMETHEE method for combining the linguistic values by the direct computation on labels. In this paper, a multi-criteria, multi-expert method has been presented to obtain the overall linguistic value without loss of information, taking into account the particular nature of the criteria and the specific differences among the experts through the aggregation process. The proposed model is computationally simple and quick.

Acknowledgments

This paper has been partially supported by the R&D project TIN2006-02121, P08-TIC-3548 and FEDER funds.

References

[1] Maystere, L.Y., Pictet, J., Simos, J.: Méthodes Multicritères ELECTRE Lausanne. Presses Polytechniques et Universitaires, Romandes (1994)

[2] Brans, J.P., Mareschal, B., Vincke, P.: PROMETHEE: A New Family of Outranking Methods in Multicriteria Analysis. In: Brans (eds.): Operational Research'84, Elsevier Science Publishers, North Holland, (1984) 408-421

[3] Herrera, F., Martinez, L.: A 2-Tuple Fuzzy Linguistic Representation Model For Computing with Words. IEEE Transactions on Fuzzy Systems, Vol. 8. (2000) 746-752

[4] Degani, R., Bortolan, G.: The Problem of Linguistic Approximation in Clinical Decision Making. International Journal of Approximate Reasoning, Vol. 2. (1988) 143-162

[5] Delgado, M., Verdegay, J.L., Vila, M.A.: On Aggregation Operations of Linguistic Labels. International Journal of Intelligent Systems, Vol. 2. (1993) 351-370

[6] Herrera, F., Martinez, L.: A Model Based on Linguistic 2-Tuples for Dealing with Multigranularity Hierarchical Linguistic Contexts in Multiexpert Decision-Making. IEEE Transactions on Systems, Man and Cybernetics, Vol. 31. (2001) 33-49

[7] Herrera, F., Martinez, L., Sanchez, P.J. : Managing Non-Homogeneous Information in Group Decision Making. European Journal of Operational Research, Vol. 166. (2005) 115-132

[8] Arrow, K.J.: Social Choice and Individual Values. Yale University Press, New Haven, CT (1963)

[9] Orlovsky, S.A.: Decision-Making with a Fuzzy Preference Relation. Fuzzy Sets and Systems, Vol. 1. (1978) 155-167

[10] Roubens, M.: Some Properties of Choice Functions Based on Valued Binary Relations. European Journal of Operational Research, Vol. 40. (1989) 309-321.

[11] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. Part I, Information Sciences 8 (1975) 199-249, Part II, Information Sciences 8 (1975) 301-357, Part III, Information Sciences 9 (1975) 43-80.

[12] F. Herrera, L. Martínez, The 2-tuple Linguistic Computational Model. Advantages of its linguistic description, accuracy and consistency. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2001, Vol 9 pp. 33-48.

ISBN: 978-989-95079-6-8

Generating Minimum Dispersion Densities from an Interval-Valued Fuzzy Set

Carol L. Walker[1] Elbert A. Walker[1] Ronald R. Yager[2]

1. Department of Mathematical Sciences
Las Cruces, New Mexico, USA
2. Machine Intelligence Institute, Iona College
New Rochelle, New York, USA
Email: hardy@nmsu.edu, elbert@nmsu.edu, yager@panix.com

***Abstract*—Let f be an interval-valued fuzzy subset of a finite set U of size n. This yields n closed intervals inside the unit interval. Picking a point in each interval and dividing by the sum of the points gives rise to a probability density on the set of intervals. For a given measure of dispersion, such as entropy, the problem is to pick points that maximize (or minimize) this dispersion. The particular case considered here is to pick points that minimize the sum of the squares of probabilities in the density. We provide an algorithm for picking such points.**

***Keywords*—interval-valued fuzzy set; minimum dispersion density; probability density; sum of squares**

I. Introduction

A problem that arises in several situations is this. Given a family of probability densities, choose one from that family that has the largest value for some measure of dispersion. Some instances involving entropy are discussed in [1], [2], [3], [4], [5], [6], [7]. But there are other measures of dispersion, for which a general formulation is provided in [8].

The problem we consider is, given n subintervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_n, b_n]$ of $[0, 1]$, choose $x_i \in [a_i, b_i]$ such that for an n-tuple $\mathbf{x} = (x_1, ..., x_n)$ the value

$$H(\mathbf{x}) = \sum_j \left(\frac{x_j}{\sum_i x_i} \right)^2 \tag{1}$$

is minimum. This means that among all the possible choices $\mathbf{x}$ for the n-tuple of x_i's, this quantity is minimum.

One motivation for this problem is the following. Let U be a finite set and f a function from U into the set of closed intervals of $[0, 1]$. That is, $f : U \to \{[a, b] : 0 \leq a \leq b \leq 1\}$. This yields a finite set of closed intervals $\{[a_i, b_i] : i = 1, 2, \ldots, n\}$. If not all the $a_i = 0$, than choosing $x_i \in [a_i, b_i]$ yields a probability density $P([a_i, b_i]) = x_i / \Sigma x_j$ on $\{[a_i, b_i] : i = 1, 2, \ldots, n\}$. Maximizing, (or minimizing) a dispersion measure gives a canonical way to pick out a point x_i in each interval $[a_i, b_i]$, a possible first step toward defuzzifying the interval-valued fuzzy set f.

Several remarks are in order.

1) It should be noted that the intervals do not have to be distinct, and that an interval can be a single point. Intervals such as $[0, 0]$ and $[0, a]$ are allowed, as are $[1, 1]$ and $[b, 1]$. And of course, an interval may be contained in another.
2) If all the intervals are $[0, 0]$, then each $x_i = 0$ and each probability is $\frac{0}{0}$, which is undefined so *we assume this is not the case*. Thus at least one of the intervals is not equal to $[0, 0]$, and there is at least one $S = \sum_{i=1}^n x_i$ that is positive. In this case, the interval $[0, 0]$ has associated probability 0, and its contribution to the sum of the squares is 0. Thus, in developing an algorithm for finding the x_i's that give minimum dispersion, *we can assume that no interval is* $[0, 0]$, that is, $b_i > 0$ for all i.
3) If the intersection of the intervals is non-empty, then choosing all the x_i to be any point $x_i = x > 0$ in that intersection yields minimum dispersion, namely
$$\sum_{j=1}^n \left(\frac{x}{\sum_{i=1}^n x} \right)^2 = \frac{1}{n} \tag{2}$$
In particular, the solution may not be unique. In fact, in this case, if the intersection is not a single point, then there are uncountably many solutions, namely any positive point in the intersection. That $\frac{1}{n}$ is the minimum possible dispersion is a special case of the Cauchy-Schwarz inequality, which says that
$$\left(\sum_{i=1}^n x_i y_i \right)^2 \leq \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right) \tag{3}$$
Letting each $y_i = 1$, we get
$$\left(\sum_{i=1}^n x_i \right)^2 \leq \left(\sum_{i=1}^n x_i^2 \right) n \tag{4}$$
and
$$\frac{1}{n} \leq \sum_{j=1}^n \left(\frac{x_j}{(\sum_{i=1}^n x_i)} \right)^2 \tag{5}$$
4) If $a_i = 0$ for all i, then since we are assuming $b_i > 0$ for all i, the intersection of the intervals contains a point $x > 0$, and we are in case 3 above. In particular, we can assume that $a_i > 0$ for some i. In this situation, $S = \sum x_i$ is always positive, and *there are x_i that minimize*

ISBN: 978-989-95079-6-8

$H(\mathbf{x}) = \sum_{j=1}^{n}\left(\frac{x_j}{\sum_{i=1}^{n} x_i}\right)^2$ since H is a continuous function on the compact space $\prod_{i=1}^{n}[a_i, b_i]$.

5) An x_i does not have to be an end-point of its interval. For example, if $n = 3$, and the three intervals are disjoint, then there will be a unique solution, namely the right endpoint of the left most interval, the left endpoint of the right most interval, and generally some interior point of the middle interval. We will see an explicit example of this below.

II. A Reformulation

Based on the discussion above, we are in the following situation. We have a finite set $[a_1, b_1], [a_2, b_2], ..., [a_n, b_n]$ of closed subintervals of the unit interval $[0, 1]$, all the b_i are positive, and at least one a_i is positive. We wish to choose $\mathbf{x} = (x_1, x_2, ..., x_n)$ with each $x_i \in [a_i, b_i]$ such that $H(\mathbf{x}) = \sum_{j=1}^{n}\left(\frac{x_j}{\sum_{i=1}^{n} x_i}\right)^2$ is minimum. In fact, we will provide an algorithm that will give an exact solution in terms of the end points of the intervals $[a_i, b_i]$. First we prove a lemma that provides a reformulation into a somewhat simpler problem. We term $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ a **solution** if $H(\mathbf{x})$ is minimum.

Lemma 1: Let $x_i \in [a_i, b_i]$ with $x_1 \leq x_2 \leq \cdots \leq x_k < x_{k+1} \leq \cdots \leq x_n$, and let $S = \sum_{i=1}^{n} x_i$. Then there exists $h > 0$ such that

$$\left(\frac{x_1 + h}{S + h}\right)^2 + \sum_{j=2}^{n}\left(\frac{x_j}{S + h}\right)^2 < \sum_{j=1}^{n}\left(\frac{x_j}{S}\right)^2 \tag{6}$$

Proof: Letting $f(h) = \left(\frac{x_1+h}{S+h}\right)^2 + \sum_{j=2}^{n}\left(\frac{x_j}{S+h}\right)^2$, we have

$$\begin{aligned} f'(h) &= 2\left(\frac{x_1 + h}{S + h}\right)\left(\frac{S + h - (x_1 + h}{(S + h)^2}\right) \\ &\quad + \sum_{j=2}^{n} 2\left(\frac{x_j}{S + h}\right)\left(\frac{-x_j}{(S + h)^2}\right) \end{aligned} \tag{7}$$

Now $f'(h)$ is negative if and only if

$$\begin{aligned} &(x_1 + h)(S - x_1) - \sum_{j=2}^{n} x_j^2 \\ &= (x_1 + h)\left(\sum_{j=2}^{n} x_j\right) - \sum_{j=2}^{n} x_j^2 \\ &= \left(x_1 \sum_{j=2}^{n} x_j\right) - \sum_{j=2}^{n} x_j^2 + h\sum_{j=2}^{n} x_j \end{aligned} \tag{8}$$

is negative. Since $x_1 < x_{k+1}$, $\left(x_1 \sum_{j=2}^{n} x_j\right) - \sum_{j=2}^{n} x_j^2$ is negative, and hence h can be chosen positive and small enough so that $f'(h)$ is negative. The lemma follows. ■

If $x_1 < b_1$, then h can be chosen so that also $x_1 + h \in [a_1, b_1]$. Thus we have the following.

Corollary 2: If $x_i \in [a_i, b_i]$ such that $x_1 \leq x_2 \leq ... \leq x_k < x_{k+1} \leq ... \leq x_n$ is a solution, then $x_1 = b_1$.

Note that the proof uses the fact that x_1 is the smallest x_i. If $x_2 = x_1$, then of course $x_2 = b_2$.

For intervals $[a_i, b_i]$, let $R = \min\{b_i\}_{i=1}^{n}$ and $L = \max\{a_i\}_{i=1}^{n}$. Note that if $L \leq R$, then $[L, R]$ is the intersection of the intervals $[a_i, b_i]$, and picking for each x_i the same element x in $[L, R]$ yields a solution $\mathbf{x}$ with $H(\mathbf{x}) = \frac{1}{n}$, the minimum possible value. Also, from the lemma above, we see that for $\mathbf{x}$ with at least two distinct entries, $H(\mathbf{x}) > \frac{1}{n}$. Thus *we may assume that* $R < L$. Otherwise the problem is trivial. If $R < L$, then any $\mathbf{x}$ has at least two distinct entries, and by our corollary, all $x_i \geq b_1 = R$. Similarly, we may show that all x_i are $\leq L$. This means that the intervals $[a_i, b_i]$ may be replaced by the intervals $[a_i \vee R, b_i \wedge L]$. This turns out to be a significant reduction of the original problem. For example, it already provides us with two elements of a solution $\mathbf{x}$. The smallest and the largest entry in any solution will be R and L, respectively, since two intervals are $[R, R]$ and $[L, L]$. Here is our reformulation.

Reformulation: Let $0 < R < L \leq 1$, and let $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_n, b_n]$ be subintervals of $[R, L]$ with $a_1 = b_1 = R$ and $a_n = b_n = L$. Find $x_i \in [a_i, b_i]$ which minimizes $\sum_{i=1}^{n}\left(\frac{x_i}{\sum_{j=1}^{n} x_j}\right)^2$.

The reformulated problem has the same solution (or perhaps solutions) as the original problem. In the subintervals of $[R, L]$, one is $[R, R]$, and one is $[L, L]$. Call these $[a_1, b_1]$ and $[a_n, b_n]$ respectively. So $x_1 = a_1 = b_1$, and $x_n = a_n = b_n$, and we have two of the required x_i immediately. We assume from now on that we are in this reformulated situation. There are some additional lemmas we need.

Lemma 3: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution and if $a_k < b_k = L$, then $x_k < b_k$. Similarly, if $R = a_k < b_k$, then $a_k < x_k$.

Proof: Of course, $k < n$. Let $x_i \in [a_i, b_i]$, $i = 1, 2, \ldots, n$, and suppose that $a_k < b_k = b_n = x_k = L$. Let $S = \sum_{j=1}^{n} x_j$ and consider

$$f(h) = \sum_{i=1, i\neq k}^{n}\left(\frac{x_i}{S - h}\right)^2 + \left(\frac{x_k - h}{S - h}\right)^2 \tag{9}$$

Then a calculation shows that

$$f'(h) = \frac{2}{(S - h)^3}\left(\sum_{\substack{i=1\\ i\neq k}}^{n} x_i^2 - x_k \sum_{\substack{i=1\\ i\neq k}}^{n} x_i + h\sum_{\substack{i=1\\ i\neq k}}^{n} x_i\right) \tag{10}$$

If $h = 0$, then $f'(h) < 0$ since $x_i \leq x_k$ for all i, and $x_1 = R < x_k$. For h sufficiently small and positive, $f'(h) < 0$. A similar proof works for $R = a_k < b_k$. ■

Lemma 4: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, then for no two distinct entries $x_i < x_j$ in $\mathbf{x}$ can it be that $a_i < b_i$, $a_j < b_j$, and $x_i \in [a_i, b_i)$ and $x_j \in (a_j, b_j]$.

Proof: Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, suppose that $x_1 < x_2$, and suppose that $x_1 \in [a_1, b_1)$ and $x_2 \in (a_2, b_2]$. Consider the density given by $(x_1 + h, x_2 - h, x_3, \ldots, x_n)$. Since $x_1 + h +$

ISBN: 978-989-95079-6-8

$x_2 - h + \sum_{i=3}^{n} x_i = \sum_{i=1}^{n} x_i = S$, we have

$$f(h) = \left(\frac{x_1+h}{S}\right)^2 + \left(\frac{x_2-h}{S}\right)^2 + \sum_{j=3}^{n}\left(\frac{x_j}{S}\right)^2 \quad (11)$$

and its derivative with respect to h is

$$f'(h) = \frac{2}{S}\left(\frac{x_1+h}{S}\right) - \frac{2}{S}\left(\frac{x_2+h}{S}\right) = \frac{2}{S^2}(x_1 - x_2) \quad (12)$$

This latter quantity is negative whenever $x_2 - h > x_1 + h$. Since $x_1 \in [a_i, b_i)$ and $x_2 \in (a_2, b_2]$ we can so choose h, keeping x_1 in $[a_1, b_1]$ and x_2 in $[a_2, b_2]$, and decreasing H. Thus $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ cannot be a solution. ■

This lemma just says that if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, and $x_i < x_j$, then x_i cannot be moved to the right and x_j to the left, keeping them in their original intervals, and keeping $x_i < x_j$. That is, if $x_i < x_j$ is part of a solution, then either $x_i = b_i$ or $x_j = a_j$.

Corollary 5: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, then no two distinct entries in $\mathbf{x}$ can be in the interior of their intervals.

Corollary 6: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, and x_i is in the interior of its interval, then for every $x_j < x_i$, $x_j = b_j$ and for every $x_j > x_i$, $x_j = a_j$.

As mentioned earlier, the function

$$H(\mathbf{x}) = \sum_{j=1}^{n}\left(\frac{x_j}{\sum_{i=1}^{n} x_i}\right)^2 \quad (13)$$

assumes a minimum since it is continuous on a compact space. Whatever that minimum, it consists of some endpoints of some of the intervals $[a_i, b_i]$ and interior points of the rest. But this corollary says that all those interior points must be equal. *So any solution consists of endpoints of some of the intervals, and a common interior point of the remainder of the intervals.*

We need one more lemma.

Lemma 7: Let $0 < k < n$. Suppose that $[a, b]$ is the intersection of the intervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_k, b_k]$, $a < b$, and $x_i \in [a_i, b_i]$, $i = k+1, \ldots, n$. Letting $S = kx + \sum_{i=k+1}^{n} x_i$, an $x \in [a, b]$ that minimizes

$$f(x) = \left(k\left(\frac{x}{S}\right)^2 + \sum_{i=k+1}^{n}\left(\frac{x_j}{S}\right)^2\right) \quad (14)$$

is either an endpoint a or b, or an interior point x such that $f'(x) = 0$. The solution to $f'(x) = 0$ is the point

$$x_0 = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i} \quad (15)$$

which may or may not be in the interval $[a, b]$. If $x_0 < a$, then f assumes its minimum on $[a, b]$ at a. If $x_0 > b$, then f assumes its minimum on $[a, b]$ at b. Otherwise, f assumes its minimum on $[a, b]$ at x_0.

Proof: The function f has a minimum at a, b, or at a point in the interior of the interval $[a, b]$ where the derivative of f is zero. Since the derivative of S is k, the derivative of $f(x)$ with respect to x is

$$\begin{aligned} f'(x) &= 2k\left(\frac{x}{S}\right)\left(\frac{S-kx}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right) \quad (16)\\ &= 2k\left(\frac{x}{S}\right)\left(\frac{S-kx}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right)\\ &= 2k\left(\frac{x}{S}\right)\left(\frac{\sum_{i=k+1}^{n} x_i}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right)\\ &= \left(\frac{2k}{S^3}\right)\left(x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2\right) \end{aligned}$$

This implies that $f'(x) = 0$ only when

$$x = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i} \quad (17)$$

The sign of the derivative $f'(x)$ is the same as the sign of $x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2$. If $x < \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, then

$$\begin{aligned} x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2 &< \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2\\ &= \sum_{i=k+1}^{n} x_i^2 - \sum_{i=k+1}^{n} x_i^2 = 0 \quad (18) \end{aligned}$$

Similarly, if $x > \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, then $x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2 > 0$. Thus, $f(x)$ is decreasing when $x < \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, increasing when $x > \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$,. and reaches its minimum at $x = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$. The lemma follows. ■

Note that this lemma only minimizes H *given the* x_i, $i = k+1, k+2, \ldots, n$, and note that we have assumed that $0 < k < n$.

III. An Algorithm

Corollary 5 says that any solution is an endpoint of some set of intervals and a common interior point of the intersection of the rest of the intervals. So we search out all such situations and choose the one with minimum H. The goal is to search out all such situations efficiently. Here is an algorithm that yields a solution $\mathbf{x}$ with $H(\mathbf{x})$ minimum. In the next section, we give some examples illustrating this algorithm.

An Algorithm:

1) If $\cap_{i=1}^{n}[a_i, b_i] = [a, b] \neq \varnothing$, then let each x_i be any element in $[a, b]$.
2) If $\cap_{i=1}^{n}[a_i, b_i] = \varnothing$, replace $[a_i, b_i]$ by $[a_i \vee R, b_i \wedge L]$, and call these new intervals again $[a_i, b_i]$. (This is the reformulation mentioned earlier, which says we may assume $0 < R < L \leq 1$ with $[a_1, b_1]$, $[a_2, b_2]$, ...,

ISBN: 978-989-95079-6-8

$[a_n, b_n]$ subintervals of $[R, L]$, with $a_1 = b_1 = R$, and $a_n = b_n = L$.)

3) Find the intersection of all distinct pairs of intervals from $[a_i, b_i]$, $i = 2, 3, ..., n-1$, such that these intersections are proper closed intervals; that is, are of the form $[a, b]$ with $a < b$. Let $\mathcal{I}$ be the set of these intersections together with all the proper intervals $[a_i, b_i]$. (The intersection of any nonempty family $\mathcal{S}$ of the intervals $[a_i, b_i]$ is either a member of $\mathcal{S}$ or can be obtained as the intersection of two members of $\mathcal{S}$.)
4) For each interval $I \in \mathcal{I}$, form the family $\mathcal{S}_I$ of all intervals $[a_i, b_i]$ containing I, so for each intersection $[a, b]$, $\mathcal{S}_{[a,b]}$ is the family of all the intervals in $\{[a_i, b_i] : i = 2, 3, ..., n-1\}$ containing $[a, b]$. Note that $\mathcal{S}_{[a,b]}$ has a multiplicity, namely the number of (not necessarily distinct) intervals containing $[a, b]$.
5) For those intervals $[a_i, b_i]$ not in $\mathcal{S}_I$ and such that $a_i < b_i$, choose an endpoint in the interior of $[R, L]$. If $a_i < a_j < b_i < b_j$, do not choose a_j and b_i, as such endpoints cannot be part of a solution minimizing H by Lemma 4. If no possible choices of endpoints are possible, proceed to another $\mathcal{S}_I$. Choose all endpoints that are of the form $[a, a]$, which include $[a_1, b_1]$ and $[a_n, b_n]$. This gets for each $[a_i, b_i] \notin \mathcal{S}_I$, an endpoint x_i, except for those I for which no proper selection of endpoints is possible.
6) For each $I \in \mathcal{I}$, with each possible set of endpoints gotten from step 5, use Lemma 7 to calculate the resulting H. There may be several sets of possible endpoints for a given $I \in \mathcal{I}$.
7) Repeat for all $I \in \mathcal{I}$. Now choose from the resulting candidates the one that gives minimum H.

Note that the resulting solution given by this algorithm is an exact expression in the endpoints of the given intervals. For all we know, there may be more than one solution, even when $R < L$, but we suspect not. And, of course, there may be a more efficient algorithm.

IV. Examples

We give three examples illustrating the algorithm. The first illustrates the simplest nontrivial case.

Example 8: Consider the three intervals $I_1 = [0.25, 0.33]$, $I_2 = [0.5, 0.67]$, $I_3 = [0.75, 1]$.

1) The intersection of the intervals is empty.
2) $R = 0.33$ and $L = 0.75$. Replacing the intervals by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} &[0.33, 0.33] \\ &[0.5, 0.67] \\ &[0.75, 0.75] \end{aligned} \tag{19}$$

3) $\mathcal{I}$ has only one element, namely $[0.5, 0.67]$.
4) The only interval containing $[0.5, 0.67]$ is $[0.5, 0.67]$ itself.
5) The only endpoints to be chosen are 0.33 and 0.75.
6) Lemma 7 says that the minimum H is the minimum value of

$$\left(\frac{0.33}{1.08 + x}\right)^2 + \left(\frac{x}{1.08 + x}\right)^2 + \left(\frac{0.75}{1.08 + x}\right)^2 \tag{20}$$

for $x \in [0.5, 0.67]$. Setting the derivative equal to 0 gives

$$x = \frac{0.33^2 + 0.75^2}{0.33 + 0.75} = 0.621\,67 \tag{21}$$

which is in $[0.5, 0.67]$. So by Lemma 7 the solution is $\mathbf{x} = (0.33, 0.621\,67, 0.75)$. Calculating, we get $H(0.33, 0.621\,67, 0.75) = 0.365\,33$. Notice that $0.36533 > \frac{1}{n} = \frac{1}{3}$.
7) There was only one I to deal with.

Example 9: Consider the four intervals $I_1 = [0.25, 0.33]$, $I_2 = [0.5, 0.67]$, $I_3 = [0.69, 0.72]$, $I_4 = [0.75, 1]$.

1) The intersection of the intervals is empty.
2) $R = 0.33$ and $L = 0.75$. Replacing the intervals by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} &[0.33, 0.33] \\ &[0.5, 0.67] \\ &[0.69, 0.72] \\ &[0.75, 0.75] \end{aligned} \tag{22}$$

3) $\mathcal{I}$ has two elements, namely $[0.5, 0.67]$ and $[0.69, 0.72]$.
4) The only interval containing $[0.5, 0.67]$ is $[0.5, 0.67]$ itself and the only endpoints to choose with this interval are 0.33, 0.69, and 0.75.
5) Lemma 7 says that the minimum of H for $[0.5, 0.67]$ occurs for

$$x_0 = \frac{0.33^2 + 0.69^2 + 0.75^2}{0.33 + 0.69 + 0.75} = 0.648\,305 \tag{23}$$

since $0.648\,305$ is in $[0.5, 0.67]$. So by Lemma 7 the minimum solution for $[0.5, 0.67]$ occurs for $\mathbf{x} = (0.33, 0.648\,305, 0.69, 0.75)$. Calculating, we get $H(0.33, 0.648\,305, 0.69, 0.75) = 0.268\,082$.
6) The only interval containing $[0.69, 0.72]$ is $[0.69, 0.72]$ itself and the only endpoints to choose with this interval are 0.33, 0.67, and 0.75.
7) Lemma 7 says that the minimum H for this interval occurs for

$$x_0 = \frac{0.33^2 + 0.67^2 + 0.75^2}{0.33 + 0.67 + 0.75} = 0.640\,171 \tag{24}$$

if this is in the interval, but it is to the left of the interval, and thus by Lemma 7 the minimum occurs for $x = 0.69$. So the solution is $\mathbf{x} = (0.33, 0.67, 0.69, 0.75)$. Calculating, we get $H(0.33, 0.67, 0.69, 0.75) = 0.268\,14$.
8) Notice that $0.268\,14 > 0.268\,082 > \frac{1}{n} = 0.25$. There are no more intervals to check, so the minimum is $H(\mathbf{x}) = 0.268\,082$, which occurs for $\mathbf{x} = (0.33, 0.648\,305, 0.69, 0.75)$.

From the two preceding examples, it is easy to see what happens for any family of *mutually disjoint* intervals: for each

ISBN: 978-989-95079-6-8

proper interval $[a_i, b_i]$, choose all the right endpoints to the left of the interval and all the left endpoints to the right of the interval. Use the point in $[a_i, b_i]$ determined by Lemma 7 and compute the resulting $H(\mathbf{x})$. Then compare these and find the minimum.

In the case the intervals are not all mutually disjoint, the situation is a bit more complicated and the more subtle points in the algorithm come into play. The next example is more illustrative of the algorithm.

Example 10: Consider the seven intervals

$$\begin{aligned} J_1 &= [0.2, 0.3] \\ J_2 &= [0.35, 0.5] \\ J_3 &= [0.15, 0.375] \\ J_4 &= [0.1, 0.3] \\ J_5 &= [0.35, 0.7] \\ J_6 &= [0.25, 0.5] \\ J_7 &= [0.4, 0.8] \end{aligned} \tag{25}$$

1) The intersection of the intervals is empty.
2) $R = 0.3$ and $L = 0.4$. Replacing $[a_i, b_i]$ by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} I_1 &= [0.3, 0.3] \\ I_2 &= [0.35, 0.4] \\ I_3 &= [0.3, 0.375] \\ I_4 &= [0.3, 0.3] \\ I_5 &= [0.35, 0.4] \\ I_6 &= [0.3, 0.4] \\ I_7 &= [0.4, 0.4] \end{aligned} \tag{26}$$

3) We get the new interval $I_8 = [0.35, 0.375]$, and $\mathcal{I} = \{I_2, I_3, I_6, I_8\}$. ($I_2 = I_5$, so we need not list I_5.)
4) For each interval $I \in \mathcal{I}$, forming the family $\mathcal{S}_I$ of all intervals containing I yields
   a) $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$
   b) $\mathcal{S}_{I_3} = \{I_3, I_6\}$
   c) $\mathcal{S}_{I_6} = \{I_6\}$
   d) $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$
5) For each of these four families there happens to be only one choice for endpoints for each of the intervals not in $\mathcal{S}_I$.
   a) $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$ pairs with $\mathcal{P}_{I_2} = \{P_1 = 0.3,\ P_3 = 0.375,\ P_4 = 0.3,\ P_7 = 0.4\}$
   b) $\mathcal{S}_{I_3} = \{I_3, I_6\}$ pairs with $\mathcal{P}_{I_3} = \{P_1 = 0.3,\ P_2 = 0.35,\ P_4 = 0.3,\ P_5 = 0.35,\ P_7 = 0.4\}$
   c) $\mathcal{S}_{I_6} = \{I_6\}$ pairs with $\mathcal{P}_{I_6} = \{P_1 = 0.3,\ P_2 = 0.35,\ P_3 = 0.375,\ P_4 = 0.3,\ P_5 = 0.35,\ P_7 = 0.4\}$. But by Lemma 4, the endpoints P_2 and P_3 cannot both be part of a solution, so we discard this option.
   d) $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$ pairs with $\mathcal{P}_{I_8} = \{P_1 = 0.3,\ P_4 = 0.3,\ P_7 = 0.4\}$
6) For each of these three remaining options $\mathcal{S}_{I_2}$, $\mathcal{S}_{I_3}$, and $\mathcal{S}_{I_8}$ we compute the values

$$f_I(x) = k_I \left(\frac{x}{S_I}\right)^2 + \sum_{x_j \in \mathcal{P}_I} \left(\frac{x_j}{S_I}\right)^2 \tag{27}$$

where k_I is the number of intervals in $\mathcal{S}_I$, $S_I = k_I x + \sum_{x_j \in \mathcal{P}_I} x_j$, and x is either an endpoint of the intersection of $\mathcal{S}_I$ that lies in the interior of $[R, L]$, or

$$x = \frac{\sum_{x_i \in \mathcal{P}_I} x_i^2}{\sum_{x_i \in \mathcal{P}_I} x_i}. \tag{28}$$

   a) For $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$ we have $I_2 = [0.35, 0.4]$ and $\mathcal{P}_{I_2} = \{0.3, 0.375, 0.3, 0.4\}$. Then

$$\begin{aligned} x &= \frac{0.3^2 + 0.375^2 + 0.3^2 + 0.4^2}{0.3 + 0.375 + 0.3 + 0.4} \\ &= 0.349\,55 \notin [0.35, 0.4] \end{aligned} \tag{29}$$

   By Lemma 7, since $0.349\,55 < 0.35$, on $I_2 = [0.35, 0.4]$, the minimum is assumed at 0.35. We have

$$3\,(0.35) + 0.3 + 0.375 + 0.4 = 2.\,125 \tag{30}$$

   and so the value of H for this set of points is

$$\begin{aligned} f_{I_2}(0.35) &= 3\left(\frac{0.35}{2.\,125}\right)^2 + \left(\frac{0.3}{2.\,125}\right)^2 \\ &+ \left(\frac{0.375}{2.\,125}\right)^2 + \left(\frac{0.4}{2.\,125}\right)^2 \\ &= 0.167\,89 \end{aligned} \tag{31}$$

   b) For $\mathcal{S}_{I_3} = \{I_3, I_6\}$, we have $I_3 \cap I_6 = [0.3, 0.375]$ and $\mathcal{P}_{I_3} = \{0.3, 0.35, 0.3, 0.35, 0.4\}$, and

$$\begin{aligned} x &= \frac{0.3^2 + 0.35^2 + 0.3^2 + 0.35^2 + 0.4^2}{0.3 + 0.35 + 0.3 + 0.35 + 0.4} \\ &= 0.344\,12 \in [0.3, 0.375] \end{aligned} \tag{32}$$

   By Lemma 7, since $0.344\,12 \in [0.3, 0.375]$, on $[0.3, 0.375]$, the minimum is assumed at $0.344\,12$.which has multiplicity 2. We have

$$2\,(0.344\,12) + 0.3 + 0.35 + 0.3 + 0.35 + 0.4 = 2.\,388\,2 \tag{33}$$

   and so the value of H for this set of points is

$$\begin{aligned} f_{I_3}(0.344\,12) &= 2\left(\frac{0.344\,12}{2.\,388\,2}\right)^2 + \left(\frac{0.3}{2.\,388\,2}\right)^2 \\ &+ \left(\frac{0.35}{2.\,388\,2}\right)^2 + \left(\frac{0.3}{2.\,388\,2}\right)^2 \\ &+ \left(\frac{0.35}{2.\,388\,2}\right)^2 + \left(\frac{0.4}{2.\,388\,2}\right)^2 \\ &= 0.144\,09 \end{aligned} \tag{34}$$

ISBN: 978-989-95079-6-8

d. For $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$ we have $I_2 \cap I_3 \cap I_4 \cap I_6 = [0.35, 0.375]$ and $\mathcal{P}_{I_8} = \{0.3, 0.3, 0.4\}$, and

$$x = \frac{0.3^2 + 0.3^2 + 0.4^2}{0.3 + 0.3 + 0.4} = 0.34 \notin [0.35, 0.375] \tag{35}$$

By Lemma 7, since $0.34 < 0.35$, the minimum is assumed at 0.35,which have multiplicity 4. We have

$$4\,(0.35) + 0.3 + 0.3 + 0.4 = 2.4 \tag{36}$$

and so the value of H for this set of points is

$$\begin{aligned} f_{I_8}(0.35) &= 4\left(\frac{0.35}{2.4}\right)^2 + \left(\frac{0.3}{2.4}\right)^2 \\ &+ \left(\frac{0.3}{2.4}\right)^2 + \left(\frac{0.4}{2.4}\right)^2 \\ &= 0.144\,10 \end{aligned} \tag{37}$$

Note that the values computed are all $> \frac{1}{n} = \frac{1}{7}$, as they should be. Comparing the values 1.6789, 1.4409, and 1.4410, we see that H assumes its minimum as indicated below.

$$\begin{aligned} 0.3 &\in [0.3, 0.3] \\ 0.35 &\in [0.35, 0.4] \\ 0.344\,12 &\in [0.3, 0.375] \\ 0.3 &\in [0.3, 0.3] \\ 0.35 &\in [0.35, 0.4] \\ 0.344\,12 &\in [0.3, 0.4] \\ 0.4 &\in [0.4, 0.4] \end{aligned} \tag{38}$$

In terms of the original intervals, we have

$$\begin{aligned} 0.3 &\in [0.2, 0.3] \\ 0.35 &\in [0.35, 0.5] \\ 0.344\,12 &\in [0.15, 0.375] \\ 0.3 &\in [0.1, 0.3] \\ 0.35 &\in [0.35, 0.7] \\ 0.344\,12 &\in [0.25, 0.5] \\ 0.4 &\in [0.4, 0.8] \end{aligned} \tag{39}$$

V. Comments

We have not shown for $R < L$ that there is a unique solution. This seems to be an interesting technical problem but may yield to more sophisticated analytical techniques. We also have not shown that our algorithm is the fastest possible. This was done in [7] for the case of computing entropy under interval uncertainty, and it may be possible to do something similar for minimizing the sum of squares.

The algorithm provided to find a solution is tedious, but could easily be programmed. It is emphasized that each member of a solution is an exact expression in terms of the endpoints of the given intervals.

There are many measures of dispersion, and in [8], several are discussed. Also, a general definition of dispersion is formulated.

References

[1] J. Abellan and S. Moral. Maximum of entropy for credal sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11:587-597, 2003.

[2] E. Jaynes. *Probability Theory: The Logic of Science*. Cambridge Univ. Press, 2003.

[3] A. Meyerowitz, F. Richman, E. Walker. Calculating maximum-entropy probability densities for belief functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:377-389, 1994.

[4] H. Nguyen and E. Walker. On decision-making using belief functions. In *Advances in the Dempster-Shafer Theory of Evidence*, R. Yager, M. Fedrizzi, and J. Kacprzyk, Eds., John Wiley & Sons: New York, 311-330, 1994.

[5] H. Nguyen and E. Walker. *A First Course in Fuzzy Logic*, 3rd ed. Chapman & Hall/CRC: Boca Raton, Florida, 2006.

[6] C. Walker, E. Walker, and R. Yager. Generating a Maximum Entropy Probability Density from an Interval-Valued Fuzzy Set. *2008 Proceedings of the NAFIPS International Conference*, New York, 2224-27, May 2008.

[7] G. Xiang, M. Ceberio, and V. Kreinovich. Population Variance and Entropy under Interval Uncertainty: Linear-Time Algorithms. *Reliable Computing*, 13(6):467-488, 2007.

[8] R. Yager. On the dispersion measure of OWA operators. *Technical Report #MII-2901*, Machine Intelligence Institute, Iona College.

ISBN: 978-989-95079-6-8

Approximate Reasoning in Surgical Decisions

Elisabeth Rakus-Andersson

School of Engineering, Department of Mathematics and Science, Blekinge University of Technology
Karlskrona, Sweden
Email: Elisabeth.Andersson@bth.se

Abstract—*Approximate reasoning is one of the most effective fuzzy systems. The compositional rule of inference founded on the logical law modus ponens is furnished with a true conclusion, provided that the premises of the rule are true as well. Even though there exist different approaches to an implication, being the crucial part of the rule, we modify the early implication proposed by Zadeh* [1] *in our practical model concerning a medical application. The approximate reasoning system presented in this work considers evaluation of a risk in the situation when physicians weigh necessity of the operation on a patient. The patient's clinical symptom levels, pathologically heightened, indicate the presence of a disease possible to recover by surgery. We wish to evaluate the extension of the operation danger by involving particularly designed fuzzy sets in the algorithm of approximate reasoning.*

Keywords—Approximate reasoning, compositional rule of inference, Zadeh's implication, operation risk, symptom levels, parametric membership functions.

1 Introduction

The technique of approximate reasoning, earliest evolved by Zadeh [1, 2] quickly found many adherents who differentiated the foundations of the theory. Especially, the changes concerned the implication IF…THEN…ELSE…, which constitutes an important factor of the reasoning system. In [3, 4, 5] we can trace the discussion revealing definitions of the implication generated by Kleene and Denies, Willmot, Mamdani and Assilian, Larsen, Gödel et al. The trials of inserting individually created operations on fuzzy sets discern the approaches mentioned above. Even the item of compositional rule of inference was debated from separate points of views [6, 7, 8, 9, 10]. We can mention the Yager conception [10] and the Sugeno design [3] as the most original modifications of the initial version of the rule.

For a practitioner an applicable meaning of approximate reasoning is essential, especially in technique and natural sciences where vagueness of input and output is often expected. Although some technical trials of applications are remarkable, it can happen coincidentally to counterpart the approximate reasoning in medicine. The only contribution in the topic, found by the author in [11], is a discussion of the model employing a pharmacological example.

Since members of surgical staff make decisions about operations on severely-ill patients with the highest care then we wish to support these verdicts by results coming from reasoning systems. We adopt Zadeh's approach to the rule [1, 2, 12], which is slightly modified by us and based on Lukasiewicz's definition of the fuzzy implication [1, 6, 12]. We still find this rule to be the most appealing for the reason of simply performed operations and clearly interpretable results. Then we build an own original apparatus accommodated to medical assumptions. Particular fuzzy sets that contain input data and output effects are designed in compliance with the physician's hint. The discussion about how to find the objective of reasoning, i.e. operation risk, is accomplished in Section 2. Fuzzy sets, taking place in the model, are furnished with appropriate membership degrees in Section 3. Section 4, added as a presentation of efficiency of the algorithm, reveals some risks in cancer surgery.

2 Adoption of approximate reasoning to operation decisions

For patients, who suffer from e.g. cancer, decisions concerning their operations are made with the highest thoughtfulness. In the later or the last stage of the disease the possibility to cure the patient totally of cancer by operating him/her for tumors is rather little. As a physician does not want the patient to run the risk to suffer even more after an unnecessary operation, he ought to judge thoroughly the consequences of the surgery.

We intend to involve approximate reasoning to support mathematically the extraction of a proper decision when discerning the operation danger. The most decisive clinical symptoms found in an individual patient will be taken into consideration to evaluate the risk.

Let us ponder a logical compound statement

$$\begin{aligned}&\text{IF}\,(p\,\text{AND}\,((\text{IF}\,p\,\text{THEN}\;q)\,\text{ELSE}\,(\text{IF}\,(\text{NOT}\;p)\\&\text{THEN}\,(\text{NOT}\;q)))\,)\,\text{THEN}\;q\end{aligned}\tag{1}$$

whose primitive statements p and q are included in the equivalent form of (1) derived as

$$p\wedge((p\rightarrow q)\wedge(\neg p\rightarrow\neg q))\rightarrow q.\tag{2}$$

The logical joint ELSE is interpreted in (2) as the conjunction $\wedge$ in compliance with the suggestions made by Lukasiewicz and Zadeh [1, 6].

The logical statement (2) is a tautology, which can be easily confirmed by the method of truth tables. We also prove that thesis q in (2) will become true if the premises p and $(p\rightarrow q)\wedge(\neg p\rightarrow\neg q)$ constitute true statements as well. In order to accomplish the last proof we utilize the method of denying the truth of the thesis q. Let $v(p)$ and $v(q)$ denote the truth values of p and q according to the convention of binary logic. If, on behalf of the proof, we

ISBN: 978-989-95079-6-8

assume that the thesis q is not true then $v(q) = 0$. From the previous assumption $v((p \to q) \wedge (\neg p \to \neg q)) = 1$ if $v(p \to q) = 1$ and $v(\neg p \to \neg q) = 1$. But $v(q) = 0$, which suggests that $v(p) = 0$ as well to warrant $v(p \to q) = 1$. On the other hand we have already assumed that premise p is true. As the suggestion $v(q) = 0$ leads to the contradiction "p is false" against "p is true" then we will accept $v(q) = 1$.

In accordance with the extended law *modus ponens* proposed by Zadeh [1, 2] we interpret (2) as a sentence

IF
$p`$ (premise)
AND (3)
(IF p THEN q) ELSE (IF (NOT p) THEN (NOT q)) (premise)
THEN
$q`$ (thesis)

provided that the semantic meaning of p and $p`$ ($q$ and $q`$ respectively) is very close.

Let p be visualized by a fuzzy set P in the universe X and let q be expressed by another fuzzy set Q in the universe of discourse Y. Analogously, the fuzzy set $P` \subset X$ constitutes a mathematical formalization of the primitive statement $p`$ whereas $Q` \subset Y$ replaces formally the sense of $q`$. The modus ponens rule thus becomes

IF
$p` = P`$ (premise)
AND (4)
(IF $p = P$ THEN $q = Q$) ELSE (IF (NOT $p = CP$) THEN (NOT $q = CQ$)) (premise)
THEN
$q` = Q`$ (thesis)

The sets CP and CQ are complements of P and Q.

When making a feedback to the medical task previously outlined, we wish to use a technique of accommodating actual theoretical assertions to concrete formulations letting us evaluate the operation decision in some grades of risk.

Let S denote a symptom possessing the most decisive power in the evaluation of the operation risk. We regard S as either the complex qualitative symptom or the symptom whose intensity is assimilated with level codes. These codes, determined for both descriptions of S's complexion, form the universe X = "*symptom levels*" = $\{1,\ldots,k,\ldots,n\}$. Let us assume that level 1 is associated with the slightly heightened symptom values whereas level n indicates their critical status.

The statement $p`$

$p`$ = "symptom S is found in patient on level k"

is now addressed to a fuzzy set $P`$ introduced by

$$P` = \frac{\mu_{P`}(1)}{1} + \cdots + \frac{\mu_{P`}(k)}{k} + \cdots + \frac{\mu_{P`}(n)}{n}. \quad (5)$$

The sentence p built by

"p = "rising levels of S are essential for operation risk"

is dedicated to a fuzzy set P given by

$$P = \frac{\mu_P(1)}{1} + \cdots + \frac{\mu_P(k)}{k} + \cdots + \frac{\mu_P(n)}{n}. \quad (6)$$

Another category of elements, constituting a content of the universe Y, is determined in the model as risk grades. We set risk grades in Y = "*operation risk grades*" = $\{L_0$ = "*none*", L_1 = "*little*", L_2 = "*moderate*", L_3 = "*great*", L_4 = "*total*"$\}$, on condition that Y is experimentally restricted to five risk grades only.

For sentence q

q = "operation risk exists for patient"

a creation of a fuzzy set Q is supported by

$$Q = \frac{\mu_Q(L_0)}{L_0} + \frac{\mu_Q(L_1)}{L_1} + \frac{\mu_Q(L_2)}{L_2} + \frac{\mu_Q(L_3)}{L_3} + \frac{\mu_Q(L_4)}{L_4}. \quad (7)$$

At last, we define $q`$ containing the final risk judgment as a statement

$q`$ = "patient runs estimated risk of being operated",

where *risk* is graded by membership degrees of the corresponding fuzzy set Q' proposed as

$$Q` = \frac{\mu_{Q`}(L_0)}{L_0} + \frac{\mu_{Q`}(L_1)}{L_1} + \frac{\mu_{Q`}(L_2)}{L_2} + \frac{\mu_{Q`}(L_3)}{L_3} + \frac{\mu_{Q`}(L_4)}{L_4}. \quad (8)$$

In the next paragraph we accomplish the discussion about an apparatus providing us with membership degrees of sets (5)–(8).

Due to *modus ponens* rule (4) we set all decision data in the scheme

IF
"symptom S is found in patient on level k" = $P`$ (premise)
AND
(IF "rising levels of S are essential for operation risk" = P THEN "operation risk exists for patient" = Q) ELSE (IF ("rising levels of S are not essential for operation risk" = CP THEN operation risk does not exists for patient = CQ) (premise)
THEN
"patient runs estimated risk of being operated" = $Q`$ (thesis)

In conformity with [1, 6, 12] we first prognosticate a mathematical expression of the implication

ISBN: 978-989-95079-6-8

(IF "rising levels of *S* are essential for operation risk" = *P* THEN "operation risk exists for patient" = *Q*) ELSE (IF ("rising levels of *S* are not essential for operation risk" = *CP* THEN operation risk does not exists for patient = *CQ*)

performed as matrix *R*. Even though several approaches to membership functions of implications were made [1, 2, 3, 5, 6, 8, 12] we still feel attracted by the Lukasiewicz [6, 12] conception of fuzzy implication *R* with a membership function derived as

$$\mu_R(k,L_l) = 1 \wedge ((1-\mu_P(k)) + \mu_Q(L_i)) \\ \wedge (\mu_P(k) + (1-\mu_Q(L_i))), \tag{9}$$

$k = 1,\ldots,n$, $l = 0,\ldots,4$, for all $x \in X$ and all $y \in Y$.

The membership degrees of set $Q`$ will be visualized after composing set $P`$ with relation R due to Zadeh's compositional rule [1]

$$Q` = P' \circ R \tag{10}$$

designated by the membership function

$$\mu_{Q`}(L_l) = \max_{k \in X}(\min(\mu_{P`}(k), \mu_R(k,L_l))) . \tag{11}$$

The comparisons of magnitudes of membership degrees in set $Q`$ yield indications referring to judgments of the risk grades after consideration of symptom level *k* verified in the patient.

As the operations of maximum and minimum have a tendency to filter the input data, which sometimes does not result in a clear-cut decision, then we will propose another set of composition operations in (10). In accordance with [13] we propose

$$Q` = P' \underset{\cdot +}{\circ} R \tag{12}$$

assisted by membership degrees

$$\mu_{Q`}(L_l) = \frac{\sum_{k=1}^{n} \mu_{P`}(k) \cdot \mu_R(k,L_l)}{\sum_{k=1}^{n} \mu_R(L_l)} . \tag{13}$$

To be able to apply (13) we ought to prove that the value of the quotient $\mu_{Q`}(L_l)$ is a number belonging to the interval [0, 1]. To verify this we first notice that $\mu_{P`}(k) \cdot \mu_R(k,L_l) \le \mu_R(k,L_l)$ since both $\mu_{P`}(k)$ and $\mu_R(k,L_l)$ are less than one for all *k* and *l*, $k = 1,\ldots,n$, $l = 0,\ldots,4$. This causes the value of a product to be lesser than the values of both factors. We thus conclude that the numerator is less than or equal to the denominator, which guarantees that the entire value of the quotient is a member of [0, 1]; therefore it can be approved as a membership degree of L_l coming from the support of $Q`$.

We also notice that the sum placed in the denominator of the quotient never becomes equal to zero, since almost all risk grades will be designed as positive quantities. This assumption prohibits membership degrees of the risk grades from being undefined structures.

Values $\mu_R(k,L_l)$ are adaptable to be treated as weights of level importance assigned to a distinct risk. These, as the entries of matrix *R* are invariants in the system promoting the same diagnostic model, contrary to information concerning different patients that is changeable. And, additionally, we can prove that operation (13) satisfies the criteria of OWA operators [13].

3 Mathematical design of data sets

The decision model designed in Section 2 includes operations on fuzzy sets furnished with symbolically established membership degrees. In the current paragraph we put some life into theoretical symbols by assigning to them mathematical structures. The set $P`$ a.k.a. (5) now gets assigned

$$P` = \frac{\mu_{P`}(1)}{1} + \cdots + \frac{\mu_{P`}(k)}{k} + \cdots + \frac{\mu_{P`}(n)}{n} \\ = \cdots + \frac{\frac{n-2}{n}}{k-2} + \frac{\frac{n-1}{n}}{k-1} + \frac{1}{k} + \frac{\frac{n-1}{n}}{k+1} + \frac{\frac{n-2}{n}}{k+2} + \cdots \tag{14}$$

for the k^{th} symptom level certified in the patient examined.

Another set *P*, concerning the same symptom levels in the support, is found by (6) and modified as

$$P = \frac{\mu_P(1)}{1} + \cdots + \frac{\mu_P(k)}{k} + \cdots + \frac{\mu_P(n)}{n} \\ = \frac{\frac{1}{n}}{1} + \cdots + \frac{\frac{k}{n}}{k} + \cdots + \frac{\frac{n}{n}}{n}, \tag{15}$$

due to the previously made assumptions, which suggest the tendency to ascending values of the membership degrees in *P*.

The set *Q* is more sophisticated to design as a fuzzy set whose support consists of other fuzzy sets L_l, $l = 0,\ldots,4$, commonly defined in a symbolic risk reference set $Z = [0, 1]$. We also intend to determine the membership degrees of *Q* as some characteristic quantities from [0, 1]. Evaluation of these numbers is founded on a procedure involving a linguistic variable

"*operation risk grades*" = {L_0 = "*none*", L_1 = "*little*", L_2 = "*moderate*", L_3 = "*great*", L_4 = "*total*"},

experimentally restricted to five risk grades only.

We first fuzzify the expressions concerning the items of the list to continue further with their defuzzification in order to attach numerical equivalents to the words from the list. Each word assists now a fuzzy set L_l, $l = 0, 1, 2, 3, 4$, whose constraint is grounded on an *s*-class mapping defined for *z* in $Z = [0, 1]$ as [14]

ISBN: 978-989-95079-6-8

$$\mu_{L_l}(z)=\mu_{L_0(l)}(z)=\begin{cases}\text{left}(\mu_{L_0(l)}(z))= \\ \text{right}(\mu_{L_0(l)}(z))= \\ s(z,\alpha_{L_0},\beta_{L_0},\gamma_{L_0},l\cdot h), & \text{for} \quad z\le\gamma_{L_0}, \\ 1-s(z,\alpha_{L_0}+h,\beta_{L_0}+h,\gamma_{L_0}+h,l\cdot h) & \text{for} \quad z>\gamma_{L_0}. \end{cases} \tag{16}$$

We clarify the fact that formulas of all membership functions are derived from only one predetermined subject defining $\mu_{L_0}(z)$. The equality $\mu_{L_l}(z)=\mu_{L_0(l)}(z)$ reveals that $\mu_{L_l}(z)$ is dependent on a parameter l equal to level number l, l = 0,…,4. The h unit determines a distance between α_{L_l} and $\alpha_{L_{l+1}}$ (respectively β_{L_l} and $\beta_{L_{l+1}}$ or γ_{L_l} and $\gamma_{L_{l+1}}$) for symmetric functions s.

We prepare constraints for L_0, which are affected by $\alpha_{L_0}=-0.25$, $\beta_{L_0}=-0.125$ and $\gamma_{L_0}=0$ as

$$\text{left}(\mu_{L_0}(z))=\begin{cases}2\left(\frac{z-(-0.25)}{0-(-0.25)}\right)^2 \\ \text{for} \quad -0.25\le z<-0.125, \\ 1-2\left(\frac{z-0}{0-(-0.25)}\right)^2 \\ \text{for} \quad -0.125\le z<0, \end{cases} \tag{17}$$

and

$$\text{right}(\mu_{L_0}(z))=\begin{cases}1-2\left(\frac{z-0}{0.25-0}\right)^2 \\ \text{for} \quad 0\le z<0.125, \\ 2\left(\frac{z-0.25}{0.25-0}\right)^2 \\ \text{for} \quad 0.125\le z<0.25. \end{cases} \tag{18}$$

By inserting in (17) and (18) the current value l, l = 0,…,4, and the distance h, casually determined as h = 0.25, we obtain a formula of the left branch of L_l

$$\text{left}(\mu_{L_l}(z))= \begin{cases}2\left(\frac{z-(-0.25+l\cdot 0.25)}{0-(-0.25)}\right)^2 \\ \text{for} \quad -0.25+l\cdot 0.25\le z<-0.125+l\cdot 0.25, \\ 1-2\left(\frac{z-(0+l\cdot 0.25)}{0-(-0.25)}\right)^2 \\ \text{for} \quad -0.125+l\cdot 0.25\le z<0+l\cdot 0.25, \end{cases} \tag{19}$$

and a function shaping its right branch

$$\text{right}(\mu_{L_l}(z))= \begin{cases}1-2\left(\frac{z-(0+l\cdot 0.25)}{0.25-0}\right)^2 \\ \text{for} \quad 0+l\cdot 0.25\le z<0.125+l\cdot 0.25, \\ 2\left(\frac{z-(0.25+l\cdot 0.25)}{0.25-0}\right)^2 \\ \text{for} \quad 0.125+l\cdot 0.25\le z<0.25+l\cdot 0.25. \end{cases} \tag{20}$$

Figure 1 collects plots of L_0–L_4 in conformity with different values of l included in (19) and (20).

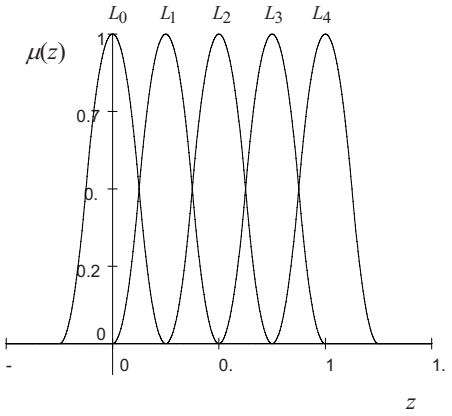


Figure 1: The terms of "*operation risk grades*" as fuzzy sets L_0–L_4

Actually, we have an intension to emphasize the meaning of parametric nature of the L_l membership functions, which deprives the model of many distinct formulas. Apart from this advantage we focus on generating the functions that represent elegant structures mathematically expressed.

In the process of defuzzification we consider only z-values for which the sets L_0–L_4 get the status of normal sets, i.e., z = 0, z = 0.25, z = 0.5, z = 0.75 and z = 1. For these, another fuzzy set "*numerical operation risk*" is projected by developing its membership function in the form of

$$\mu_{\text{"numerical operation risk"}}(z)=\begin{cases}2\left(\frac{z-0}{1-0}\right)^2 \\ \text{for} \quad 0\le z<0.5, \\ 1-2\left(\frac{z-1}{1-0}\right)^2 \\ \text{for} \quad 0.5\le z<1. \end{cases} \tag{21}$$

Via the selected z-quantities above, we tie their membership degrees calculated by means of (21) to expressions from the list in order to establish relations between words and their numerical replacements. Therefore, the set Q finally obtains a shape of

$$Q=\frac{0}{L_0}+\frac{0.125}{L_1}+\frac{0.5}{L_2}+\frac{0.875}{L_3}+\frac{1}{L_4}. \tag{22}$$

We now wish to demonstrate the action of approximate reasoning accustomed to the judgment of surgical risk.

4 Risks grades in cancer surgery

In patients, who suffer from cancer as the recognized diagnosis, one of the symptoms, namely, *CRP* (*C*-reactive proteins) is carefully measured and discussed with a view to make a decision about accomplishing a successful operation. The heightened values of *CRP* (measured in milligrams per liter) are theoretically discerned in four levels stated as

1 = "*almost normal*" for $CRP < 10$,
2 = "*heightened*" if $10 \le CRP \le 20$,
3 = "*very heightened*" if $20 \le CRP \le 25$,

ISBN: 978-989-95079-6-8

4 = "*dangerously heightened*" for $CRP > 25$.

Due to (15) set P is expressed as

$$P = \frac{0.25}{1} + \frac{0.5}{2} + \frac{0.75}{3} + \frac{1}{4} \tag{23}$$

in $X = \{1,\ldots,4\}$.

Suppose that an individual patient examined reveals the *CRP*-value to be 23. *CRP* is thus classified in level 3 and set P` characteristic of the patient is stated in the form of

$$P` = \frac{0.5}{1} + \frac{0.75}{2} + \frac{1}{3} + \frac{0.75}{4} \ . \tag{24}$$

according to (14).

The sets (23) and (22) together with

$$\mathrm{C}P = \frac{0.75}{1} + \frac{0.5}{2} + \frac{0.25}{3} + \frac{0}{4} \tag{25}$$

and

$$\mathrm{C}Q = \frac{1}{L_0} + \frac{0.875}{L_1} + \frac{0.5}{L_2} + \frac{0.125}{L_3} + \frac{0}{L_4} \tag{26}$$

generate matrix R with the entries computed in compliance with (9). R is expanded as a two-dimensional table

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} L_0 & L_1 & L_2 & L_3 & L_4 \end{array} \\ \begin{bmatrix} 0.75 & 0.875 & 0.75 & 0.375 & 0.25 \\ 0.5 & 0.625 & 1 & 0.625 & 0.5 \\ 0.25 & 0.375 & 0.75 & 0.875 & 0.75 \\ 0 & 0.125 & 0.5 & 0.875 & 1 \end{bmatrix} \end{array} \tag{27}$$

which, inserted in (10) for P` determined by (24), provides us with

$$Q` = \frac{0.5}{L_0} + \frac{0.625}{L_1} + \frac{0.75}{L_2} + \frac{0.875}{L_3} + \frac{0.75}{L_4} . \tag{28}$$

By interpreting the meaning of (28) we understand that there exists a risk when considering an operation in patient whose *CRP*-index is evaluated on the third level. The most possible risk is evaluated as "*great*" according to the highest quantity of the membership degree. The total danger of accomplishing the surgical operation is evaluated as essential with the membership degree 0.75.

Even the results of implementing (13) given as

$$Q` = \frac{0.66}{L_0} + \frac{0.69}{L_1} + \frac{0.75}{L_2} + \frac{0.795}{L_3} + \frac{0.725}{L_4} . \tag{29}$$

fully confirm the risk extension judged by (28).

We hope that the classical model of approximate reasoning, modified by us and adapted to the problem of operation decision can constitute its complementary solution, especially when a decision of saving somebody's life via surgery is crucial.

5 Conclusions

Via the way of forming the text of this paper we have already come to substantial conclusions. We only summarize that we have used approximated reasoning to introduce the own initial interpretation of the system to approximate the operation risk concerning patients with rising values of a biological index. The formulas of membership degrees and membership functions have been expanded by applying a formal mathematical design. We expect that the study makes a contribution in the domain of mathematical models projected for medical applications.

In future works we wish to examine a model consisted of several symptoms that are divided in different numbers of levels. The symptoms should be included in the pattern simultaneously, which may expose some internal interactions among them. In other words, the operation risk will be a criterion that can employ many data factors. We count on finding some helpful remarks in [15] to implement an algorithm supporting the method newly planned.

References

[1] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision process. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. MSC-3(1): 28-44, 1973.

[2] L. A. Zadeh. A theory of approximate reasoning. *Machine Intelligence*, 9: 149-174, 1979.

[3] R. Fuller. On fuzzy reasoning schemes. http:// www.abo. fi/~ rfuller/ pgs97.pdf.

[4] E. Hüllermeier**.** *Case based approximate reasoning*, Berlin Heidelberg: Springer, 2007.

[5] P. M. van den Broek. Efficient algorithms for approximate reasoning. *Proceedings of the 6th International Conference on Neural Information Processing, ICONIP 1999*, Perth, Western Australia, 1999.

[6] J. F. Baldwin and B. W. Pilsworth. A model of fuzzy reasoning through multi-valued logic and set theory. *Int. J. of Man-Machine Studies*, 11: 351-380, 1979.

[7] C. Cornelis, M. de Cock and E. Kerre. Efficient approximate reasoning with positive and negative information. *Lecture Notes in Artificial Intelligence*, 3214: 779-785, 2004.

[8] G. Gerla. *Fuzzy logic: mathematical tools for fuzzy reasoning*. Kluwer, 2001.

[9] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. of Man-Machine Studies*, 7: 1-13, 1975.

[10] R. Yager. An approach to inference in approximate reasoning. *Int. J. of Man-Machine Studies*, 13: 323-338, 1980.

[11] K. Kilic, B. A. Sproule, I. B. Türksen and C. A. Naranjo. A fuzzy system modeling algorithm for data analysis and approximate reasoning. https://research.sabanciuniv. edu/418/1/robotics _kemal. DOC.R.

[12] J. Kacprzyk. *Fuzzy sets in system analysis*. Warsaw: PWN, 1986.

[13] E. Rakus-Andersson. *Fuzzy and rough techniques in medical diagnosis and medication*. Berlin Heidelberg: Springer, 2007.

[14] E. Rakus-Andersson. *S*-truncated functions and rough sets in approximation and classification of polygons. *Proc. of Modeling Decisions for Artificial Intelligence – MDAI 2005,* Tsukuba, CD-ROM, paper nr 049, Consejo Superior de Investigaciones Cientificas, 2005.

ISBN: 978-989-95079-6-8

[15] O. Wolkenhauer. Fuzzy inference engines, composition and individual rule based on non-linear mappings. www. Systemsbiology .umist.ac. uk/dokumente/ dataengineering/ infengines .pdf

ISBN: 978-989-95079-6-8

Chaos induced by turbulent and erratic functions

H. Román-Flores[1] Y. Chalco-Cano[2] G. Nunes-Silva[3]

1.Instituto de Alta Investigación, Universidad de Tarapacá
Arica, Chile
2.Departamento de Matemática, Universidad de Tarapacá
Arica, Chile
3.Departamento de Ciência da Computação e Estatística, Universidade Estadual Paulista
São José do Rio Preto - SP, Brasil
Email: {hroman,ychalco}@uta.cl, gsilva@ibilce.unesp.br

***Abstract*— *Let (X,d) be a compact metric space and $f:X\to X$ a continuous function and consider the hyperspace $(\mathcal{K}(X),H)$ of all nonempty compact subsets of X endowed with the Hausdorff metric induced by d. Let $\bar{f}:\mathcal{K}(X)\to\mathcal{K}(X)$ be defined by $\bar{f}(A)=\{f(a)/a\in A\}$ the natural extension of f to $\mathcal{K}(X)$, then the aim of this work is to study the dynamics of $\bar{f}$ when f is turbulent (erratic, respectively) and its relationships.*

***Keywords*—** Chaos, dynamical systems, turbulent functions, erratic functions.

1 Introduction

Let (X,d) be a compact metric space and $f:X\to X$ a continuous function. If we consider the extension $\bar{f}:\mathcal{K}(X)\to\mathcal{K}(X)$, then an interesting problem is to analyze the chaotic relations between f and $\bar{f}$.

In this direction, Román-Flores in [1] proved that transitivity of $\bar{f}$ implies transitivity of f, whereas in [2] the author proved that sensitivity of $\bar{f}$ implies sensitivity of f and periodic density of f implies periodic density of $\bar{f}$. Also, Banks in [3] shows that transitivity of $\bar{f}$ is equivalent to f weakly mixing. Several other chaotic relations between f and $\bar{f}$ have also been investigated by many authors, for example Cánovas et al. [4] investigated some connections in relation to topological entropy, Gu in [5] studied Kato's chaos, and Ma et al. [6] investigated some aspects of topological entropy, Li-Yorke chaos and distributional chaos.

The aim of this work is to explore some turbulent and erratic connections between f and $\bar{f}$.

2 Preliminaries

Let (X,d) be a compact metric space and let $\mathcal{K}(X)$ be the class of all non-empty and compact subsets of X. If $A\in\mathcal{K}(X)$ we define the "ϵ-neighbourhood of A" as the set

$$N(A,\epsilon)=\{x\in X/\ d(x,A)<\epsilon\},$$

where $d(x,A)=\inf_{a\in A}d(x,a)$.

The Hausdorff metric on $\mathcal{K}(X)$ is defined as

$$H(A,B)=\inf\{\epsilon>0/\ A\subseteq N(B,\epsilon)\ and\ B\subseteq N(A,\epsilon)\},$$

and it is well known that $(\mathcal{K}(X),H)$ is a compact metric space (see [7]).

If $A\in\mathcal{K}(X)$ we denote by $B(A,\epsilon)$ the ball centered in A and radius ϵ in H-metric.

Remark 1 *An equivalent formula for H is given by*

$$H(A,B)=\max\left\{\sup_{a\in A}d(a,B),\ \sup_{b\in B}d(b,A)\right\}.$$

Remark 2 *If K_n,K_0 are nonemty compact subsets of X such that $K_n\xrightarrow{H}K_0$, then (see [7])*

$$K_0=\bigcap_{p=1}^{\infty}\overline{\bigcup_{n\geq p}K_n}. \tag{1}$$

If X has a real linear structure then, by using the Minkowski sum between two sets, a linear structure of convex cone is defined on $\mathcal{K}(X)$ by means

$$A+B=\{a+b,\ a\in A,b\in B\}\ \text{ and }\ \lambda A=\{\lambda a,\ a\in A\},$$

for all $A,B\in\mathcal{K}(X),\ \lambda\in\mathbb{R}$.

Let A be a subset of X. Then we define the extension of A to $\mathcal{K}(X)$ as

$$e(A)=\{K\in\mathcal{K}(X)/\ K\subseteq A\}.$$

The following results were proved by the author in [1].

Lemma 1 *Let A,B be two subsets of X and let $f:X\to X$ be a continuous function. Then,*

i) $e(A)=\emptyset$ *if and only if* $A=\emptyset$.

ii) $e(A\cap B)=e(A)\cap e(B)$

iii) If A is a nonempty open subset of X, then $e(A)$ is a nonempty open subset of $\mathcal{K}(X)$.

iv) $\bar{f}(e(A))\subseteq e(f(A))$

v) $\bar{f}^p=\overline{f^p}$ *, for every* $p\in\mathbb{N}$.

We remark that if f is a bijective map then

$$\bar{f}(e(A))=e(f(A)). \tag{2}$$

Also, as a direct consequence of (1) we have

Lemma 2 *If A is a nonempty closed subset of X, then $e(A)$ is a nonempty closed subset of $\mathcal{K}(X)$.*

ISBN: 978-989-95079-6-8

Proof. If K_0 is an acumulation point of $e(A)$ in $\mathcal{K}(X)$ then there exists a sequence (K_n) contained in $e(A)$ such that $K_n \xrightarrow{H} K_0$. Thus, $K_n \subseteq A$ for all $n \in \mathbb{N}$ and, due to (1), we have

$$K_0 = \bigcap_{p=1}^{\infty} \overline{\bigcup_{n \geq p} K_n} \subseteq A$$

which implies that $K_0 \in e(A)$.□

Lemma 3 *If A is a nonempty convex subset of X, then $e(A)$ is a nonempty convex subset of $\mathcal{K}(X)$.*

Proof. Let J, K be two nonempty compact subsets of X such that $J, K \in e(A)$. Thus, since $J, K \subseteq A$ and A is a convex set, then

$$\lambda J + (1-\lambda)K = \{\lambda x + (1-\lambda)y : x \in J, y \in K\} \subseteq A,$$

for all $\lambda \in [0,1]$, which implies that
$\lambda J + (1-\lambda)K \in e(A)$ and, consequently, $e(A)$ is a convex set.□

3 Erraticness and turbulence

In this section we will analyze some connections between the chaos induced by turbulent functions and erratic functions.

Definition 1 *Let $f : X \to X$ be a continuous function. We say that f is a turbulent function if there exist two disjoint nonempty compact subsets J, K of X such that*

$$J \cup K \subseteq f(J) \cap f(K)$$

Definition 2 *Let $f : X \to X$ be a continuous function. We say that f is chaotic in the sense of Block and Coppel (in short: B-C-chaotic) if one of its iterates is turbulent, i.e., there exist $n \geq 1$ and two disjoint nonempty compact subsets J, K of X such that $J \cup K \subseteq f^n(J) \cap f^n(K)$.*

Definition 3 *Suppose that X is a compact convex subset of a linear space and let $f : X \to X$ be a continuous function. We say that f is an erratic function if there exists a nonempty convex compact subset A of X such that*

a) $A \cap f(A) = \emptyset$;

b) $A \cup f(A) \subseteq f^2(A)$.

Theorem 1 *If $f : X \to X$ is a bijective turbulent function, then $\bar{f} : \mathcal{K}(X) \to \mathcal{K}(X)$ is also a turbulent function.*

Proof. Suppose that f is a turbulent function. Then, if J, K are two disjoint nonempty compact subsets of X such that $J \cup K \subseteq f(J) \cap f(K)$, then by Lemma 1 i)-ii) and Lemma 2 it is clear that $e(J), e(K)$ are two disjoint nonempty compact subsets of $\mathcal{K}(X)$.
Moreover, because $J \cup K \subseteq f(J) \cap f(K)$, we have $J \subseteq f(J) \cap f(K)$ and $K \subseteq f(J) \cap f(K)$, which implies that $e(J) \subseteq e(f(J) \cap f(K))$ and $e(K) \subseteq e(f(J) \cap f(K))$.
Thus, due to equality (2), we obtain

$$\begin{array}{rcl} e(J) \cup e(K) & \subseteq & e(f(J) \cap f(K)) \\ & = & e(f(J)) \cap e(f(K)) \\ & = & \bar{f}(e(J)) \cap \bar{f}(e(K)) \end{array}$$

and, consequently, $\bar{f}$ is a turbulent function.□

Corollary 1 *If the continuous bijection f is B-C-chaotic, then $\bar{f}$ is B-C-chaotic.*

Theorem 2 *If $f : X \to X$ is a bijective erratic function, then $\bar{f} : \mathcal{K}(X) \to \mathcal{K}(X)$ is also an erratic function.*

Proof. Suppose that f is an erratic function, and let A be a nonempty compact convex subset of X satisfying conditions a) and b) in Definition 3. Then, due to Lemmas 2 and 3, we have that $e(A)$ is a nonempty compact convex subset of $\mathcal{K}(X)$. Now we will prove that $e(A)$ satisfies conditions a) and b) for $\bar{f}$. In fact, by 2 and equality (2) we have:

$$\begin{array}{rcl} i)\ A \cap f(A) = \emptyset & \Rightarrow & e(A) \cap e(f(A)) = \emptyset \\ & \Rightarrow & e(A) \cap \bar{f}(e(A)) = \emptyset \end{array}$$

and, since $A \cup f(A) \subseteq f^2(A)$, we have:

$$\begin{array}{rcl} ii)\ e(A) \cup \bar{f}(e(A)) & = & e(A) \cup e(f(A)) \\ & \subseteq & e(A \cup f(A)) \\ & \subseteq & e(f^2(A)) = \bar{f}^2(e(A)) \end{array}$$

which implies that $\bar{f}$ is an erratic function.□

The following result shows that erraticness is stronger than B-C-chaos.

Theorem 3 *Let $f : X \to X$ be a continuous function. If f is an erratic function, then f is a B-C-chaotic function.*

Proof. Suppose that f is an erratic function, and let A be a nonempty compact convex subset of X satisfying conditions a) and b) in Definition 3 Then, due to a), taking $J = A$ and $K = f(A)$ we have $J \cap K = A \cap f(A) = \emptyset$, and by b):

$$\begin{array}{rcl} J \cup K & = & A \cup f(A) \\ & \subseteq & f^2(A) \\ & = & f(K) \\ & \subseteq & f(J \cup K) \\ & \subseteq & f(f^2(J)) = f^3(J) \end{array}$$

and

$$\begin{array}{rl} J \cup K & = A \cup f(A) \subseteq f^2(J) \\ & \subseteq f^2(J \cup K) \subseteq f^2(f(K)) = f^3(K) \end{array}$$

which implies that $J \cup K \subseteq f^3(J) \cap f^3(K)$ and, consequently, f is a B-C-chaotic function.□

Example 1 *If we consider the tent function $f : [0,1] \to [0,1]$ defined by*

$$f(x) = \begin{cases} 2x & if \quad 0 \leq x \leq \frac{1}{2} \\ 2(1-x) & if \quad \frac{1}{2} \leq x \leq 1, \end{cases}$$

then f is an erratic function. In fact, taking $A = [\frac{1}{4}, \frac{4}{9}]$ we have $f(A) = [\frac{1}{2}, \frac{8}{9}]$, which implies that $A \cap f(A) = \emptyset$ and :

$$A \cup f(A) = \left[\frac{1}{4}, \frac{4}{9}\right] \cup \left[\frac{1}{2}, \frac{8}{9}\right] \subseteq \left[\frac{2}{9}, 1\right] = f^2(A)$$

and, consequently, f is an erratic function.

On the other hand, if we take $J = A$ and $K = f(A)$ then we have $J \cap K = \emptyset$ and, since $f^2(J) = \left[\frac{2}{9}, 1\right]$ and $f^2(K) = [0,1]$, we obtain that $J \cup K \subseteq f^2(J) \cap f^2(K)$ which implies that f is B-C-chaotic.

ISBN: 978-989-95079-6-8

Acknowledgment

This work was supported by Conicyt-Chile by projects Fondecyt 1080438 and 1061244.

References

[1] H. Román-Flores. A note on transitivity in set-valued discrete systems *Chaos, Solitons & Fractals* 17, 99–104, 2003.

[2] H. Román-Flores and Y. Chalco-Cano Robinson's chaos in set-valued discrete systems *Chaos, Solitons & Fractals* 25, 33–42, 2005.

[3] J. Banks Chaos for induced hyperspace maps *Chaos, Solitons & Fractals* 25, 681–685, 2005.

[4] J. Cánovas and G. Soler. Topological entropy for induced hyperspace maps *Chaos, Solitons & Fractals* 28, 979–982, 2006.

[5] R. Gu Kato's chaos in set-valued discrete systems *Chaos, Solitons & Fractals* 31, 765–771, 2007.

[6] X. Ma, B. Hou and G. Liao Chaos in hyperspace system *Chaos, Solitons & Fractals* 40, 653–660, 2009.

[7] E. Klein and A. Thompson *Theory of Correspondences* Wiley-Interscience, New York, 1984.

ISBN: 978-989-95079-6-8

Verbalizing Time-series Data: With an Example of Stock Price Trends

Ichiro KOBAYASHI[1] Naoko OKUMURA[2]

1.Graduate School of Humanities and Sciences, Advanced Sciences, Ochanomizu University
2-1-1 Ootsuka Bunkyo-ku Tokyo, 112-8610 Japan
2.NTT Data Corporation Ltd.
Email: koba@is.ocha.ac.jp

Abstract— *We propose a method to verbalize time-series data and to generate texts that explain the behavior of the data expressed in a 2D chart. We use Nikkei stock price trends data as an example of time-series data, and then aim to generate a text which summarizes the behavior of the trends in a day with words as the one we often see on a newspaper or Web sites. The trends of the stock price are observed by numerical data expressed in a 2D chart. When observing the trends of time-series data, we usually observe the shape of the chart, and then verbalize the shape with proper words to express the behavior of the chart. In order to recognize the behavior in qualitative and quantitative ways, we use the least squares method to mathematically recognize the shape of a 2D chart and expressed it with words often used in news articles reporting the trends of stock prices. Our proposed method can change non-verbal information into verbal information — this provide us with high accessibility and usability for various kinds of information.*

Keywords— 2D chart, Stock price trends, Time-series data, Verbalizing

1 Introduction

We propose a method to verbalize time-series data. In particular, we focus on explaining the behavior of time-series data expressed in 2D charts. We often come across many documents with multimodal information like texts, graphs, tables, etc, therefore, technologies to deal with multimodal information have been increasingly required. In particular, if non-verbal information can be expressed with natural language, the information will be easily retrieved with the current text-based information retrieval system. Thus, changing non-verbal information into verbal information will provide us with high accessibility to various kinds of information, and also usability of non-verbal information can also be expanded. In this context, we propose a method to change modality through expressing non-verbal information with verbal information and to expand communication among different kinds of modalities by that. The proposed method is verified by applying it to a problem to generate a stock price trends report with the actual numerical data of Nikkei stock average.

2 Overview of system

First of all, we show the architecture of a system that generates verbal reports on stock prices trends in Figure 1. As the data for stock prices dealt with by the system, we use Nikkei stock average.

The system consists of several modules: a data base to manage numerical data of Nikkei stock average, a module to recognize the shape of 2D charts, a dictionary to express the shape of 2D line chart with words, and a module to provide syntactic rules for generating a text.

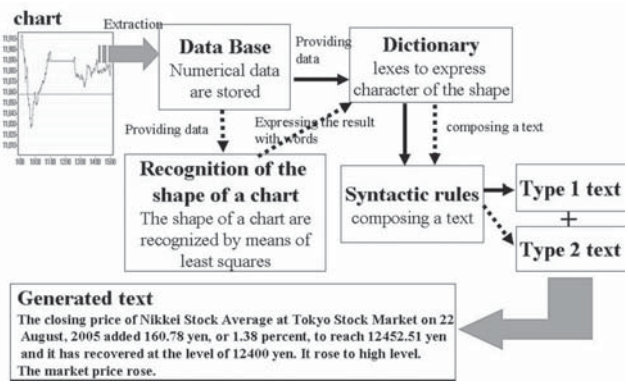


Figure 1: System architecture

The texts generated by the system are categorized into two types as follows:

type 1: The text which does not refer to the shape of 2D charts. The texts of this type are mainly confined to explaining the particular states or levels of numerical data.

type 2: The text which refers to the shape of 2D charts. The texts of this type explain the behavior of the 2D charts trends by recognizing the shape of the line charts.

2.1 Generation process

The algorithm of generating a type 2 text is as follows: (The algorithm of generating a text of type 1 merely omits *step 2* in the following steps.)

step 1. Extracting numerical data from a data base
Necessary numerical data to generate a text are extracted from a data base.

step 2. Recognition of the shape of a chart
The shape of chart is approximately recognized by means of the linear least squares.

step 3. Lexical selection for the partial shapes of a chart
To generate a type 1 text, proper lexes are selected to verbally express the numerical data obtained at step 1. And to generate a type 2 text, proper lexes are selected by recognizing the shape of a chart.

step 4. Providing templates and syntactic rules
As for type 1 text generation, the selected lexes are put

ISBN: 978-989-95079-6-8

into the slots of a prepared text template and short sentences to explain the trend of the price are added to output text. As for type 2 text generation, the short sentences to explain the behavior of a line chart are composed with syntactic rules, and then a text explaining the behavior of a line chart is generated.

Each component of the system will be explained in the following.

2.2 Data base

We use numerical data of Nikkei stock average from July 25, 2005 till August 30, 2005 as input information to the system. There are several kinds of input information which are 10 minutes interval data through a day, opening and closing price, and highest and the lowest prices data for past three days.

2.3 Recognition of the shape of a line chart

We apply the linear least squares to make an approximate line from the original line chart, and then verbally recognize the behavior of the original line chart from that of the approximate line chart. Figure 2 shows an example of applying the least squares to the original line chart. As we see from the example, we can recognize the behavior of the original line chart from the approximate line chart as its trend goes down at first and then it rises up.

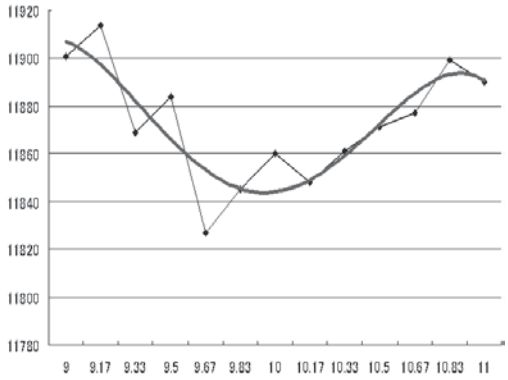


Figure 2: Example of the least squares

2.3.1 Least squares

The least squares method assumes that the best-fit curve of a given type is the curve that has the minimal sum of the deviations squared (least square error) from a given set of data. Suppose that the data points are,(x_1, y_1), (x_2, y_2),$\cdots$,(x_n, y_n) where x is the independent variable and y is the dependent variable. The fitting curve $f(x)$ has the deviation (error) d from each data point, i.e., $d_1 = y_1 - f(x_1)$, $d_2 = y_2 - f(x_2)$, $\cdots$, $d_n = y_n - f(x_n)$.

According to the least squares method, the best fitting curve has the property that:

$$F(x) = d_1^2 + d_2^2 + \cdots + d_n^2 = \sum_{i=1}^{n} d_i^2$$
$$= \sum_{i=1}^{n} [y_i - f(x_i)]^2 \rightarrow a\ minimum$$

Through analyzing corpus data — we used 27 news articles during the same period as numerical data — we adopted five dimensional polynomial function to properly approximate the shape of line charts which is because we found that the function is the most suitable polynomial function to be verbalized with the words often used in the news articles explaining the behavior of line charts.

2.3.2 Whole and partial shapes of charts

We have defined 11 types for the shape of a line chart as shown in Figure 3.

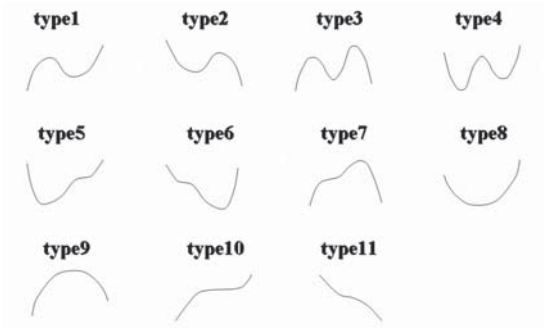


Figure 3: 11 types of whole shape

Furthermore, those 11 types are broken down into sub 13 types of partial shapes as shown in Figure 4.

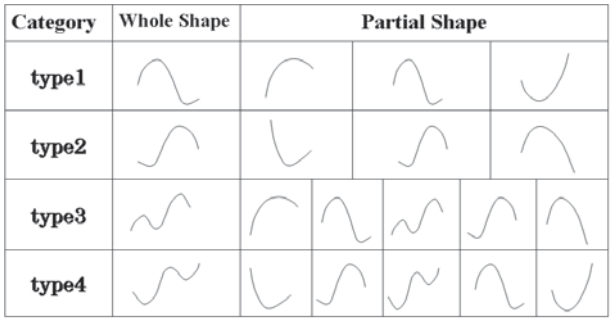


Figure 4: Example of partial shapes

This categorization for the shapes is decided based on the words often used to explain the behavior of a line chart in the corpus — we focus on the words used in explaining the behavior of a line chart, not on the characteristics of the shape of a chart, and extracted typical shapes expressed with words.

The algorithm we developed firstly selects a relevant type from the 11 types of whole shape and then the final shape of the line chart is decided through mathematical recognition of the partial shape of the selected whole type (see, Table 1 for example). Afterwards, the proper words to express the final shape are chosen to generate a report on stock price trends.

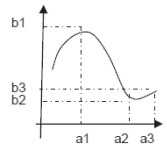
2.3.3 Target zones of recognizing the shape

There are two time zones in the trade of stocks, i.e., 'the morning session' and 'the afternoon session'(see, Fig. 5).

The texts explaining the trends of Nikkei stock price in a day on a news paper or Web sites are usually reported and summarized in each session. We follow this, and then recognize the shape of a chart with the approximate function in each session.

ISBN: 978-989-95079-6-8

Table 1: Partial shape and verbal expressions

Partial shape	Characteristics	Verbal expressions
	|b2-b1|/|MAX-MIN|>0.4 |a1-a2|/|max-min|<0.7	Sell order was dominant
	|a1-a2|/|max-min|>0.7	Sell order expanded
	|b2-b1|/|MAX-MIN|>0.4 |b2-b3|/|b2-b1|>0.5 |a1-a2|/|max-min|<0.7	Sell order was once dominant
	|a3-a2|/|max-min|<0.7 |a3 -a2|/|max-min|>0.5	Over the middle of the session
	|a3 -a1|/|max-min|<0.2 |a3 -a2|/|max-min|<0.2	Around the middle of the session
	|a3-a1|/|max-min|<0.6 |a3-a1|/|max-min|>0.45	After the middle of the session

MAX, MIN : the maximum and minimum prices at morning (afternoon) sessions.
max, min : the starting and closing time of morning (afternoon) sessions.

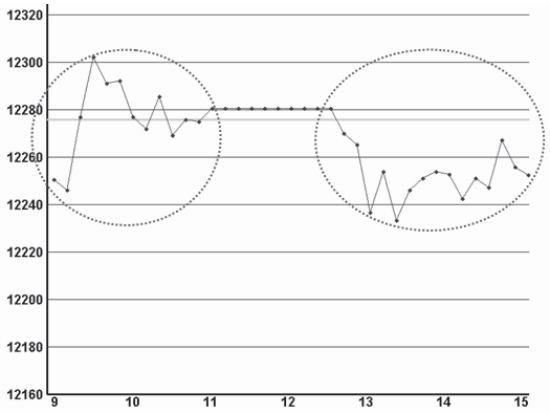


Figure 5: Target zones of recognizting the shape

2.4 *Dictionary*

In generating texts explaining the behavior of a chart, we start with the standing point where we use the words that are usually used to explain the behavior, therefore, we firstly analyze a news article corpus about the stock price trends and extract the words often used to explain the stock price trends. And then we observe the behavior of the trends by recognizing the shapes of the 2D chart of the trends from a viewpoint of explaining the behavior with the words. In concrete, we have analyzed 27 Japanese news articles about Nikkei stock average during the same period as the numerical data were obtained, i.e., from July 25, 2005 till August 30, 2005. We collected words and short sentences to explain the partial shapes of charts; words to explain particular conditions recognized from numerical data; words to express time and comparison between the current and past conditions; and conjunctions to bond those sentences.

Some examples registered in the dictionary are shown below.

- *Short sentences to explain the partial shapes of a chart*
  'Sell order was expanded', 'The price was edgingly rising', etc. 38 kinds of short sentences are registered in the dictionary.

- *Words to explain the states identified from a particular numerical data set*
  'rebound', 'continuously rise', 'The price was moving with high level through a day', etc. 19 kinds of words are registered in the dictionary.

- *Words to express time and comparison*
  'At morning (afternoon) session', 'At closing session', 'After the middle of the session', 'compared to the price of last weekend', etc. 10 kinds of words are registered in the dictionary.

- *Conjunction*
  4 kinds of conjunctions are registered in the dictionary. They are 'and then', 'therefore', 'afterwards', and 'however'.

2.5 *Syntactic rules for generated texts*

Syntactic rules to compose a text differ for each text type. Type 1 texts are generated basically by using templates. Type 2 texts are generated based on verbal expressions of the behavior of a line chart with some syntactic rules and short sentences.

2.5.1 *Type 1 text*

From the analysis of the 27 Japanese news articles, we have obtained typical three types of template to explain the Nikkei stock average trends. Type 1 text is normally generated based on one of the text templates shown below.

(a) The closing price of Nikkei stock average at Tokyo market on *DATE* added/lost *PRICE_VALUE*.

(b) The closing price of Nikkei stock average at Tokyo market on *DATE* added/lost *PRICE_VALUE*, and it got down to *PRICE_LEVEL*.

(c) The closing price of Nikkei stock average at Tokyo market on *DATE* added/lost *PRICE_VALUE*, and it has recovered at *PRICE_LEVEL*.

DATE, *PRICE_VALUE*, *PRICE_LEVEL* are the variables to be filled with numerical values obtained from the observed data stored in the data base.

And some phrases to explain the conditions of the price are usually added. For example, sentences as follows: '*The stock price was moving at high level throughout a day.*', '*The market closed with the high price of the day.*', etc.

2.5.2 *Type 2 text*

Type 2 texts are generated by composing words or short sentences to express the behavior of a chart with circumstantial information and conjunctions for the consistency among the sentences. Our system has 2 rules for adding circumstantial information and 4 rules about conjunctions. These rules reflect syntactic patterns used to explain the trends of the stock prices and are obtained through corpus analysis.

- Depending on time zone, the expression either 'in the morning session' or 'in the afternoon session' is put at the head of sentences.

- If the partial shapes of a chart can be expressed with the sentence either 'Buying is dominant' or 'Prices turn to be rising', one of the sentences expressing time interval: 'throughout the middle of trading', 'around the middle of

ISBN: 978-989-95079-6-8

trading', 'after the middle of trading' is put at the head of the sentence depending on the shape of an approximate function especially in terms of time interval.

- The partial shape expressed with the sentence 'The width of rising was small' comes with the conjunction either 'since' or 'but' at the head of the sentence, depending on the case whether or not there is a stock price trend expressed with the sentence 'It is likely to stay at a low level.' before the sentence.
- If there is not any short sentence expressing time or any conjunction before the sentence, furthermore, if there are sentences expressing the partial shape, e.g, 'Buying was ahead' before the sentence, the word 'afterwards' is inserted in between the two sentences.
- If there is a sentence describing 'Buying was happening' before the sentence 'It was hardly going down', a conjunction 'and then' is inserted in between them.
- The conjunction 'and then' is basically put at the head of the short sentences describing the width of rising or descending.
- Depending on the width of a shape expressed with a sentence 'Buying was increased', a short sentence describing 'after the middle of trading' is put at the head of the sentence.
- Depending on the width of a shape expressed with a sentence, 'the width of descending was gradually expanded', a sentence 'around the middle of trading' is put at the head of the sentence.

3 Generation Example

A screenshot of the developed system is shown in Fig. 6. The system generates type 1 and type 2 texts in Japanese, and displays the data of Nikkei stock average price on a particular day and also its 2D line chart.

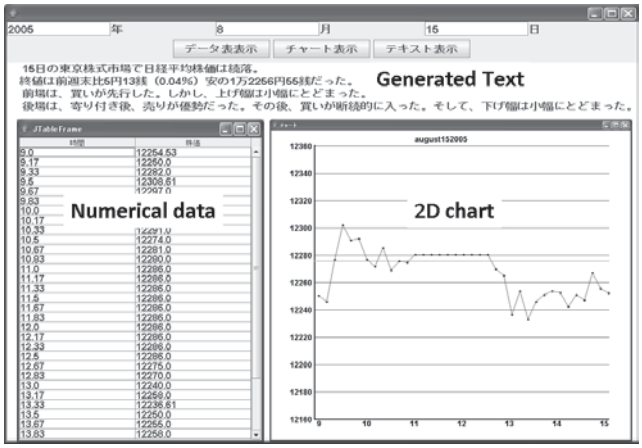


Figure 6: Screenshot of the system

3.1 Generation process of type 1 text

The generation process is explained by following generation steps explained in 2.1.

Here, let us assume that we want to get a report of the stock price trends on "August 15, 2005". This date information is the initial input information to the system.

step 1. Obtaining numerical data from a data base

Numerical data of the stock price on August 15, 2005 and the closing and the opening price, and the highest and the lowest prices data for the past three days are obtained from the data base.

step 2. Lexical selection for the shape of a chart

Proper lexes to express the obtained numerical data are selected from the dictionary.

- Date information is obtained from the input information, "August 15, 2005".
- From the obtained numerical data, the following events are found by comparing the states or levels of the points in the chart: The closing price of the stock is 'higher' compared to that of yesterday. The closing price of yesterday is 'cheaper' compared to that of the day before yesterday. The closing price of the day before yesterday is 'higher' compared to that of three days ago. Therefore, as a proper word to explain the trend of the stock price, '*rebound*' is selected.
- Compared the closing price to that of yesterday, the price of today is 160.78 yen higher (1.38 % higher), therefore, the actual values, in italic fonts, related to the price conditions were selected. 'Nikkei stock average added *160.78* yen, or *1.38* percent, to reach *12452.51* yen'.
- By the facts that the closing price of the day was over 12400 yen and the price of yesterday was under 12400 yen, a sentence, 'The closing prices has recovered at the level of *12400* yen' is generated.

step 3. Providing templates

- Compared the closing price of the day to that of the day before, the following template was selected — "The closing price of Nikkei stock average at Tokyo stock market on *DATE* added *PRICE_VALUE* and it has recovered at *PRICE_LEVEL*."
- The market closed with 25 yen higher of the closing price compared to the price of the day before, therefore, a sentence, '*It rose to a high level*' is added to a generated text.
- The closing price was higher than the opening price at both morning and afternoon sessions, and also, the closing price was higher than the opening price by 100 yen. Therefore, a sentence, '*The market price rose*' is added to a generated text.
- Compared to the closing price of the day before, the mean value between morning and afternoon sessions was over 25 yen, therefore, a sentence, '*The stock price was moving at a high level throughout the day*'.

The generated text of type 1 is as follows:
" The closing price of Nikkei stock average at Tokyo stock market on August 15, 2005 rebounded. It added 160.78 yen, or 1.38 percent, to reach 12452.51 yen and has recovered at the level of 12400 yen. It rose to high level. The market price rose. The stock price was moving at a high level throughout the day. "

ISBN: 978-989-95079-6-8

3.2 Generation process for type 2 text

step 1. Obtaining numerical data in a data base

The same process as shown in the generation process for type 1 text happens.

step 2. Recognition of the shape of a line chart

The shapes of a line chart at morning and afternoon sessions are recognized by using the least squares with the numerical data obtained at step 1.

In this example, the shapes of the chart at morning and afternoon sessions are recognized as type 1 and type 7 of whole the shape categories, respectively.

step 3. Lexical selection for the partial shapes of a chart

Proper sentences to express the behavior of a line chart are selected by recognizing the shape of the chart with verbal expressions (see, Figure 7).

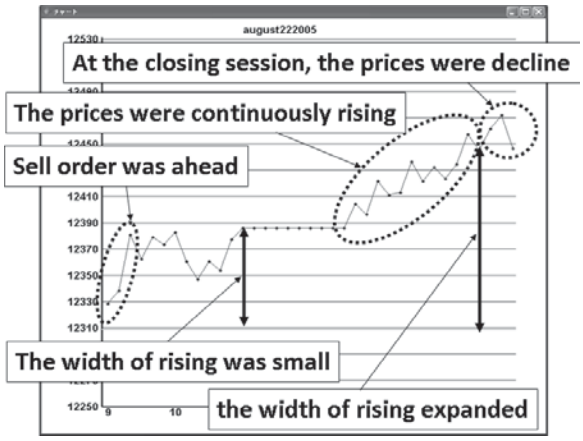


Figure 7: Shapes and their verbal expressions

- At morning session, '*Sell order was ahead.*', '*Trading was steady.*' '*The width of rising was small.*'
- At afternoon session, '*The prices were continuously rising*', '*The width of rising expanded*', '*The prices were decline*'.
- '*At the closing session*' is put ahead of '*the prices were decline*'.

step 4. Providing syntactic rules

Proper conjunctions and temporal information are added to the sentences generated at step 3 by following the syntactic rules for type 2 text generation.

- As for temporal information, '*At morning session*', and '*At afternoon session*' are added respectively at the head of the first sentence reporting an event of each session.
- A conjunction, '*afterwards*' is inserted between '*Sell order was ahead*' and '*Trading was steady.*'
- A conjunction, '*Therefore*' is put at the head of '*The width of rising was small.*'

The generated text of type 2 is as follows:
"At the morning session, sell order was ahead. Afterwards, trading was steady. therefore, the width of rising was small. At the afternoon session, trading was continuously rising. Therefore, The width of rising was expanded. At the closing session, the prices were decline. "

4 Related work

There are several studies on text generation for explaining the behavior of a line chart. Boyd [1] has applied Wavelet analysis to recognizing the characteristics of time-series data expressed with a line chart and proposed a method to express the characteristics of the chart with natural language based on the analysis result. Carberry et al. [2] is a work about summarizing graph information as well as textual information — here, summarizing graph information means expressing the graph information with natural language. Kobayashi [3] has investigated the relation between patterns of line chart behaviors and natural language to express the patterns, and has proposed an algorithm to generate a text explaining the behavior of a line chart. Furthermore, as studies to deal with time-series data, Goldberg et al. [4] and Coch [5] have developed systems that generate weather reports from weather data. As the studies that discover particular patterns from time-series data, Barndt et al. [6] dealt with economic data, Bakshi et al. [7] dealt with process control data, Himowitz et al. [8] dealt with medical data, etc.

5 Conclusions

As a method to transform non-verbal information into verbal information, we have proposed a method to generate a text by recognizing the shape of a 2D line chart. In particular, we used the linear least squares to make an approximate line chart against the original chart and then expressed the shape of the approximate line with words to generate a text reporting the stock price trends. This approach is, unlike the approach using Wavelet analysis, to verbalize the behavior of charts from the viewpoint of the words often used to verbalize time-series data, especially stock price trends, in the case of this paper. We verified our proposed method by comparing the generated reports with actual news articles. Although we used the same data for generating reports and for evaluation of the system, we have got 89.8 % precision rate between the generated report and news article for the same period of 27 days. This result shows us our system can properly verbalize the behavior of time-series data.

The proposed method is useful for reporting the trends of time-series data, and also for mining knowledge from the data with natural language. Furthermore, since the method changes visual information, i.e., 2D charts in the case of this paper, into verbal information, there is possibility to provide blind people with an interface which can report what is happening in time-series data via voice. In this context, the proposed method provides us with high accessibility and usability to various kinds of information. This provides high interaction among multimodal information.

References

[1] S. Boyd. Trend: A system for generating intelligent descriptions of time-series data. *Proc. IEEE-ICIPS.*, 1998.

[2] S. Carberry, S. Elzer, K. Green, N.and McCoy, and D. Chester. Extending document summarization to information graphics. *Proc. ACL Workshop on Text Summarization.*, 2004.

[3] I. Kobayashi. A study on text generation from non-verbal information on 2d charts. *In Computational Linguistics and Intelligent Text Processing*, pages 226–238, 2001.

ISBN: 978-989-95079-6-8

[4] E.Goldberg and N.Driedger. Using natural language processing to produce weather forecasts. *IEEE Expert.*, 1994.

[5] Jose Coch. Interactive generation and knowledge administration in multimeteo. *Proc. The 9th International Workshop on Natural Language Generation (INLG-98)*, pages 300–303., 1998.

[6] D.J.Berndt and J.Clifford. Finding patterns in time series: A dynamic programming approach. *Advances in Knowledge Discovery 1996*, 1996.

[7] B.R.Bakshi and G.Stephanopoulos. Reasoning in time. *Advances in Chemical Engineering*, 22:485–548, 1995.

[8] I.J.Himowitz, P. Phuc Le, and I.S.Kohane. Clinical monitoring using regression-based trend templates. *Artificial Intelligence in Medicine*, 7:473–496., 1995.

ISBN: 978-989-95079-6-8

Differential Equations based on Fuzzy Rules

Marina R. B. Dias, Laécio C. Barros[1]

1.Department of Applied Mathematics, IMECC, State University of Campinas
Campinas, SP - Brazil
Email: {madias, laeciocb}@ime.unicamp.br

***Abstract*— *The purpose of this paper is to use the Fuzzy Set Theory, specially fuzzy rule-based systems, as a mathematical tool for modeling dynamical systems given by differential equation, whose direction field is f (IVP-f). As fuzzy systems methodology produces input-output systems, that is, it produces m outputs from n inputs, one can see it as $f_r : \mathbb{R}^n \to \mathbb{R}^m$, where r is the number of rules of the rule base. Under certain assumptions, one can prove that the functions (f_r) have good analytical properties. Beyond it, they can approximate theoretical functions (f). The main result of this paper states that the solutions x of an ODE whose direction field is f (IVP-f) can be approximated by the solutions x_r of an ODE whose direction field is f_r (IVP-f_r). By utilizing important theorems as the Lebesgue's Dominated Convergence Theorem it was possible to prove the proposed theorem. Finally we use a numerical method (Runge-Kutta) to simulate solutions x_r^n that approximates x_r when n increases. The solutions x_r^n approximates the solution x of IVP-f also provided $\{x_r^n\} \overset{n\to\infty}{\longrightarrow} x_r \overset{r\to\infty}{\longrightarrow} x$.*

***Keywords*—** approximation theory, dynamical systems, fuzzy systems, initial value problem, numerical solutions.

1 Introduction

The investigation of a physical (biological) phenomena using differential equations needs knowledgement of the direction field. That is, it is necessary to adopt a theoretical function to represent the variation rates (derivatives) as function of the state variables. For deterministic differential equations, this direction field is a function of real variables that models its variation in time. Traditionally, aleatory uncertainties have been studied from Differential Equations in two ways: first, considering the state variables as aleatory variables. In this case it is necessary to define a variation rate (or derivative) of an aleatory variable. So, it is called stochastic differential equation. Second, considering only the aleatory parameters (therefore, it is not necessary a new definition of variation rate). These are called aleatory differential equations [1, 2]. Recently, the use of Fuzzy Set Theory to incorporate uncertainties in differential equations has been increased, so we have the fuzzy differential equations. Like the stochastic case, there are many proposals for studying these equations. The first one, using the notion of fuzzy derivative, is originated of the Hukuhara's derivative used to multifunctions and introduced in the fuzzy area by Puri and Ralescu [3]. Another notions do not require derivative to fuzzy variables. The variation rate is the same as the classical one, but some proposals have already been suggested: use differential inclusions [4, 5], use some fuzzification procedure for a deterministic solution [6], etc. In each one of these proposals there is an equation. That is, the derivatives are given by direction fields that depend on the state variables. Our research suggests a new methodology to study fuzzy differential equations using the fuzzy set theory. It will not be required to have the variation rates explicitly as a function of the state variables. We propose a fuzzy relation to model the association between the state variables and the direction field. We consider the variation rate and the state variables as linguistic variables. If no defuzzification procedure is used or if the used method produces an interval [7, 8], then the system can be studied with the aid of differential inclusion theory. The methodology we propose here uses some defuzzification method, which produces a crisp solution (real number or real vector) for each instant of time t. Regarding fuzzy differential equations, the methodology we proposed differs from previous methodologies in its formulation. While other methodologies use the concept of derivative (classical or fuzzy), our methodology uses a fuzzy formulation given by a rule base. More specifically, we adopt a fuzzy rule-based system (FRBS) to represent the differential equation. The FRBS methodology has been very important to solve modeling problems, since it is a universal approximator. For universal approximator we mean something that can approximate a continuous functions with the desired accuracy in compact sets. Thus, it is expected that fuzzy rule-based systems can model a direction field when the direction field is only partially known. Works such as Nguyen [9, 10] were extremely useful in our research. The main purpose of this work is to use fuzzy systems to obtain functions (f_r) in accordance with certain properties of the phenomenon being modeled. These families of functions will be constructed in such a way that they will be universal approximators of the function f that represents a theoretical field. Our greatest interest is to study the solutions of these new dynamical systems, given by the direction field (f_r) and establish conditions that ensure convergence of its solutions to the solution of the theoretical problem. The structure of this paper is as follows: in Section 2 we give all basic concepts and definitions the reader needs for a good understanding of our work. In Section 3 we present the problem we must solve, that is, the new IVP given by a fuzzy rule-based system. In Section 4 we present our main result as a theorem and then we prove it. Finally, in Section 5 we show an example to illustrate our result.

2 Basic Concepts and Definitions

2.1 Fuzzy Set Theory

An ordinary subset A of a set U is determined by its characteristic function χ_A defined by

$$\chi_A(x) = \begin{cases} 1, & if\ x \in A \\ 0, & if\ x \notin A \end{cases}.$$

ISBN: 978-989-95079-6-8

The characteristic function of a subset A of a set U specifies whether or not an element is in A. Zadeh [11] generalized this notion by allowing images of elements to be in the interval $[0, 1]$ rather than being restricted to the two element set $\{0, 1\}$.

Definition 1 *A fuzzy subset A of a set U is determined by a function $\varphi_A : U \to [0, 1]$, where the number $\varphi_A(x)$ represents the degree of membership of the element x to the fuzzy subset A.*

Definition 2 *A binary operation $\Delta : [0, 1] \times [0, 1] \to [0, 1]$ is a* t-norm *if it satisfies the following:*

1. $\Delta(1, x) = 1\Delta x = x$;
2. $\Delta(x, y) = x\Delta y = y\Delta x = \Delta(y, x)$;
3. $x\Delta(y\Delta z) = (x\Delta y)\Delta z$;
4. *If $x \leq u$ and $y \leq v$ then $x\Delta y \leq u\Delta v$.*

The operator t-norm models the connective "and".

Definition 3 *A binary operation $\nabla : [0, 1] \times [0, 1] \to [0, 1]$ is a* t-conorm *if and only if*

1. $\nabla(0, x) = 0\nabla x = x$;
2. $\nabla(x, y) = x\nabla y = y\nabla x = \nabla(y, x)$;
3. $x\nabla(y\nabla z) = (x\nabla y)\nabla z$;
4. *If $x \leq u$ and $y \leq v$ then $x\nabla y \leq u\nabla v$.*

The operator t-conorm models the connective "or".

Definition 4 *By a* defuzzification procedure $\mathcal{D}$ *we mean a mapping that transforms a membership function $\varphi(x)$ into a number and satisfies the following properties:*

1. *if $\varphi(x) = 0$ for all $x \in (-\infty, a)$, then $\mathcal{D}(\varphi) \geq a$;*
2. *if $\varphi(x) = 0$ for all $x \in (-\infty, a]$, then $\mathcal{D}(\varphi) > a$;*
3. *if $\varphi(x) = 0$ for all $x \in (a, +\infty)$, then $\mathcal{D}(\varphi) \leq a$;*
4. *if $\varphi(x) = 0$ for all $x \in [a, +\infty)$, then $\mathcal{D}(\varphi) < a$.*

2.2 Fuzzy Rule-Based System

A Fuzzy Rule-Based System (FRBS) (Fig. 1) is an input-output system based on fuzzy rules. It is composed of four components: an input processor; a collection of linguistic rules, called rule base; a fuzzy inference method and an output processor, which generates a real number as output.

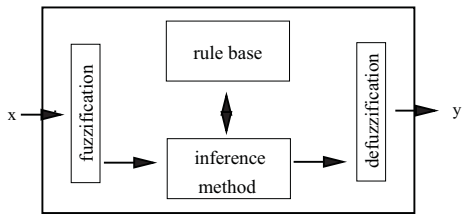


Figure 1: A FRBS scheme.

The fuzzification is the stage where the system's entries are modeled by fuzzy sets. The membership functions are constructed for each fuzzy set involved in the process and, even if the entry is crisp, it will be fuzzified by its characteristic function. The rule base is composed of a set of "if-then" rules describing an input-output relationship. In this component we can find the fuzzy propositions, which are provided in accordance with an expert. The defuzzification translates the result of fuzzy inference into a single numeric value of control (a real number) [12].

A rule base consisting of r rules has the form:

$$\begin{array}{c} R_1 : \text{"Fuzzy Proposition 1"} \\ \text{or} \\ R_2 : \text{"Fuzzy Proposition 2"} \\ \text{or} \\ \vdots \\ \text{or} \\ R_r : \text{"Fuzzy Proposition r"}. \end{array}$$

In general, each fuzzy proposition has the form

$$\text{If "}state\text{" then "}action\text{"},$$

where every "state" and every "action" are assumed as linguistic variables modeled by fuzzy sets.

In mathematical terms, each one of the r propositions is described as

$$\begin{array}{c} \text{"If } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ and} \ldots x_n \text{ is } A_n \text{ then} \\ y_1 \text{ is } B_1 \text{ and } y_2 \text{ is } B_2 \text{ and} \ldots u_m \text{ is } B_m\text{"}, \end{array}$$

where $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_m)$ are the input and output variables, respectively. A_i, $1 \leq i \leq n$ and B_j, $1 \leq j \leq m$ are fuzzy sets which model each one of the terms assumed by the input and output variables, respectively.

It is important to note that a FRBS maps $\mathbb{R}^n$ in $\mathbb{R}^m$ of a specific manner. Given an n-tuples $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, the FRBS produces an output $y = y(y_1, y_2, \ldots, y_m) \in \mathbb{R}^m$. For a set of rules of the form "If x_i is A_{ij} then y is B_j, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, r$", where $x_i,\ y_i \in \mathbb{R}$, A_{ij} and B_j are fuzzy sets, n is the number of input variables and r is the number of rules, we obtain $y_r = f_r(x_1, x_2, \ldots, x_n)$. This map always depends on: a) the membership functions A_{ij} and B_j; b) the t-norm and t-conorm chosen and c) the defuzzification method.

Many authors proved that, for particular methodologies, a fuzzy rule-based system is a universal approximator [13, 14, 15, 16], but Nguyen [9, 10] generalized and extended these results. Due to importance of his work we will report some aspects of it in here.

Let a general class of functions $\mathcal{F}(\mathcal{M}, \mathcal{L}, \mathcal{D})$, where $\mathcal{M}$ consists of those membership functions φ such that $\varphi(x) = \varphi_0(ax + b)$ for some $a, b \in \mathbb{R}$, and $a \neq 0$, and $\varphi_0(x)$ is continuous, positive on some interval of $\mathbb{R}$, and 0 outside that interval. $\mathcal{L}$ consists of continuous t-norms and t-conorms. $\mathcal{D}$ is a defuzzification procedure transforming each membership function into a real number in such a way that if $\varphi(x) = 0$ outside an interval (α, β), then $\mathcal{D}(\varphi) \in [\alpha, \beta]$.

Theorem 1 *For any design methodology $(\mathcal{M}, \mathcal{L}, \mathcal{D})$ and any compact subset K of $\mathbb{R}^n$, $\mathcal{F}(\mathcal{M}, \mathcal{L}, \mathcal{D})|_K$ is dense in $C(K)$ with respect to the sup-norm.*

ISBN: 978-989-95079-6-8

Proof: See [10].

According to Nguyen [10], for a FRBS with a $(\mathcal{M}, \mathcal{L}, \mathcal{D})$ methodology, the output f_r is continuous (therefore integrable) and it can be written as

$$f_r(x) = \frac{\sum_{j=1}^{r} y_j \ \Delta(\varphi_{A_{1j}}(x_1) \ \Delta \ ... \ \Delta \ \varphi_{A_{nj}}(x_n))}{\sum_{j=1}^{r}(\varphi_{A_{1j}}(x_1) \ \Delta \ ... \ \Delta \ \varphi_{A_{nj}}(x_n))}, \quad (1)$$

where r is the number of rules of the rule base, $\varphi_{A_{ij}}(x_i)$ is the degree of membership of x_i in the fuzzy set A_{ij} and Δ is a t-norm (see Definition 2).

Besides, the function f_r is bounded, since it is continuous and defined in a compact set.

Note that $f_r(K)$ is a compact set of $\mathbb{R}$, that is, $f_r(K)$ is a closed interval of $\mathbb{R}$.

3 The Problem

3.1 Construction of an Initial Value Problem and its solution

Suppose that the Initial Value Problem (IVP) models a particular phenomenon. Moreover, suppose that it is an Autonomous Ordinary Differential Equation, that is, the rate of change does not depend explicitly of the time. The IVP is, therefore,

$$\begin{cases} \dfrac{dx}{dt} = f(x(t)) \\ x(t_0) = x_0 \end{cases}, \quad (2)$$

where f is a known function.

The solution of the IVP (2) is guaranteed by the following proposition:

Proposition 1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous. The function $x : [a, b] \to \mathbb{R}^n$ is a solution of the IVP (2) if, and only if, it is continuous and satisfies the integral equation*

$$x(t) = x_0 + \int_{t_0}^{t} f(x(s))ds,$$

$t \in [a, b]$.

In practice, depending on the complexity of the field f, the solution $x(t)$ may not have an analytical expression. However, there are several numerical methods that estimate the unknown solution with the desired accuracy, that is, we can use a numerical method to obtain a numerical solution to the IVP (2) so that $\{x^n\} \overset{n\to\infty}{\longrightarrow} x$. Note that knowing the direction field f (or at least knowing it in some points) is necessary to use a numerical method.

We propose an alternative way to obtain the solution of the IVP (2) or at least an approximation of it, without knowing the field f explicitly. Our goal is to take advantage of the qualitative information available to construct a rule base that will represent the properties that characterize the phenomenon. To obtain the output f_r we use the methodology presented in Section 2.2 and therefore we guarantee that f_r is a good approach to f.

Thus, we have a new IVP to represent the studied phenomenon, which is given by

$$\begin{cases} \dfrac{dx}{dt} = f_r(x(t)) \\ x_r(t_0) = x_0 \end{cases}, \quad (3)$$

whose solution will be denoted by x_r and it will be studied later.

In the next subsection we will make an analysis about the existence and uniqueness of the solution x_r of the IVP (3). Since we can guarantee its existence, we will describe the main result of this work: the theorem ensures that the solution of the IVP (3) is arbitrarily close to the solution of the IVP (2) as r increases.

3.2 Existence and uniqueness of the new IVP's solution, x_r

Based on Proposition 1, if $f_r : \mathbb{R}^n \to \mathbb{R}$ is continuous, then the solution of (3) is given by

$$x_r(t) = x_0 + \int_{t_0}^{t} f_r(x(s))ds,$$

$t \in [a, b]$.

To ensure the existence and uniqueness of the solution we have the following result:

Proposition 2
- *If $f \in C_0(\mathbb{R}^n, \mathbb{R})$ then, for all $x_0 \in \mathbb{R}^n$, there is an interval $I_{x_0} \equiv (\alpha_{x_0}, \beta_{x_0})$ containing t_0 and a solution $x(., x_0)$ of the initial value problem that satisfies the initial condition $x(t_0, x_0) = x_0$.*
- *Besides, if $f \in C^1(\mathbb{R}^n, \mathbb{R})$, then $x(., x_0)$ is unique in I_{x_0} and it is a function of C^1 class.*

As we have seen in section 2.2, f_r is continuous and integrable. Therefore, the function f_r satisfies both conditions of the Proposition 2 and so we can guarantee that the solution of the IVP (3), x_r, exists and it is unique.

4 Main Result

In this section we will state the theorem that ensures the convergence of solution of the IVP (3), x_r, to the solution of the IVP (2), x, as r increases.

For a demonstration of this result we will make use of the following Lebesgue's Dominated Convergence Theorem:

Theorem 2 *Lebesgue's Dominated Convergence Theorem (LDCT): Let X be a measure space with a σ-algebra $\mathcal{A}$ and a measure μ on it. Let $f, f_1, f_2, \ldots : X \to C$ be measurable functions such that $f_n \overset{n\to\infty}{\longrightarrow} f$ for almost everywhere (that is, $\mu(\{x \in X; f_n(x) \nrightarrow f(x)\}) = 0$). Suppose there is an integrable function $g : X \to [0, +\infty]$ such that $|f_n(x)| \leq g(x)$ for almost every $x \in X$. Then $\int f_n d\mu \overset{n\to\infty}{\longrightarrow} \int f d\mu$.*

Proof: See [17].

In the following we state the main result of this work.

Theorem 3 *Let f be a continuous function in a compact set $K \subset \mathbb{R}^n$. Let $\{f_r\}$ be class of functions defined in K such that $f_r \overset{r\to\infty}{\longrightarrow} f$, for example, the fuzzy system output as in (1). Consider the IVPs (2) and (3) with x and x_r their solutions, respectively.*

Then $x_r \overset{r\to\infty}{\longrightarrow} x$.

ISBN: 978-989-95079-6-8

Proof: We know that the solutions x and x_r are given, respectively, by the equations:

$$x(t) = x_0 + \int_{t_0}^{t} f(x(s))ds,$$

and

$$x_r(t) = x_0 + \int_{t_0}^{t} f_r(x(s))ds.$$

Using the LDCT we can guarantee that the solution x_r converges to x as r increases if there is a positive and integrable function g such that $|f_r(x)| \leq g(x)$ in μ - for almost $x \in K$.

Let $g : K \to [0, +\infty]$ be such that $g(x) = max\{|\alpha|, |\beta|\}$, where $[\alpha, \beta] \subset \mathbb{R}$ is an interval such that $\mathcal{D}(\varphi(x)) \in [\alpha, \beta]$, for all $x \in K$ and $\mathcal{D}$ is a defuzzification procedure as we have seen in Section 2.2. Moreover g is a constant function in a compact set, then it is integrable. Therefore we have $|f_r(x)| \leq g(x)$ for every $x \in K$ and all r.

So, the hypothesis of LDCT are satisfied and we can conclude that

$$\int_{t_0}^{t} f_r(x(s))ds \overset{r\to\infty}{\longrightarrow} \int_{t_0}^{t} f(x(s))ds$$

and, therefore,

$$x_r \overset{r\to\infty}{\longrightarrow} x.$$

4.1 Numerical Solution

In practice, the solution, x, of a theoretical dynamical system cannot have an analytical expression due to the complexity of the direction field f. In this case it is necessary to use numerical methods, and then find solutions $\{x^n\}$ such that $\{x^n\} \overset{n\to\infty}{\longrightarrow} x$.

However $\{x^n\}$ will only be obtained if the direction field f is known, that is, knowing the direction field directions f is a restriction of the numerical method to produce the estimated $\{x^n\}$.

When there is no explicit knowledge of the field f, we can use the fuzzy sets theory and, as we have seen before, we can use fuzzy rule-based system to produce a function f_r that can replace the field f, getting the IVP (3).

Knowing f_r in a table form (for each x, we have $f_r(x)$), we can use stable and convergent numerical methods to obtain $\{x_r^n\}$ such that $\{x_r^n\} \overset{n\to\infty}{\longrightarrow} x_r$.

Therefore, as long as the fuzzy rule-based system used to produce f_r satisfies the assumptions of the main theorem, we have

$$\{x_r^n\} \overset{n\to\infty}{\longrightarrow} x_r \overset{r\to\infty}{\longrightarrow} x,$$

that is

$$x_r^n \overset{\substack{n\to\infty \\ r\to\infty}}{\longrightarrow} x.$$

Now we will show a practical example of the theory discussed here. More details can be found in [18].

5 Example

Suppose a population that has its dynamics following the qualitative characteristics of the Verhulst model (Logistic Equation). Thus, this population grows as the IVP

$$\begin{cases} \dfrac{1}{x}\dfrac{dx}{dt} = a(P_\infty - x) \\ x(t_0) = x_0 \end{cases}, \tag{4}$$

where a is the relative growth rate and P_∞ is the carrying capacity of the population (inhibition factor).

The classical solutions of (4), which represent the population $x(t)$ at every instant t, are given by

$$x(t) = \frac{P_\infty}{\left(\dfrac{P_\infty}{x_0} - 1\right)e^{-at} + 1}, \tag{5}$$

from where we conclude that:

- x increases if $x_0 < P_\infty$
- x is constant if $x_0 = P_\infty$
- x decreases if $x_0 > P_\infty$,

and is illustrated in Fig. 2.

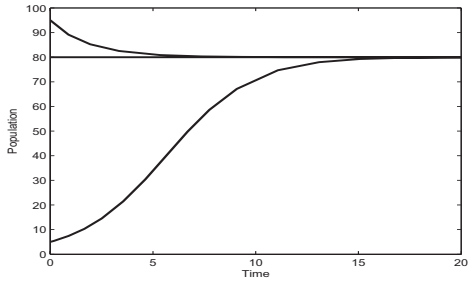


Figure 2: Possible solutions of the Verhulst model.

Considering the relative growth rate we have, rewriting (4) that

$$\frac{1}{x}\frac{dx}{dt} = a(P_\infty - x) = f(x), \tag{6}$$

where f is linear (decreasing).

Consider a FRBS with the Mamdani inference and the centroid as the defuzzification procedure. The input variable is *population* and the output variable is *relative growth rate* (here we will denote it by variation rate). The solution $x_r(t)$ is obtained using a Runge-Kutta method for each output $f_r(x)$.

Next, we have two simulations given by a fuzzy system with 3 and 8 rules in the rule base, respectively. Note that in all simulations we have a rule base that presents semantic opposition in the consequents, that is, there exists change of the consequents signals. That is what guarantees the "auto-inhibition" characteristic.

5.1 Simulation 1

In this first simulation we have three rules in the rule base. They are as follows:

R_1: If the population is *"small"* then the variation rate is *"positive medium"*
R_2: If the population is *"medium"* then the variation rate is *"positive small"*
R_3: If the population is *"large"* then the variation rate is *"negative small"*

The membership functions of the input and output variables are in Figs. 3 and 4, respectively.

The solution $x_r(t)$ is showed in Fig. 5.

ISBN: 978-989-95079-6-8

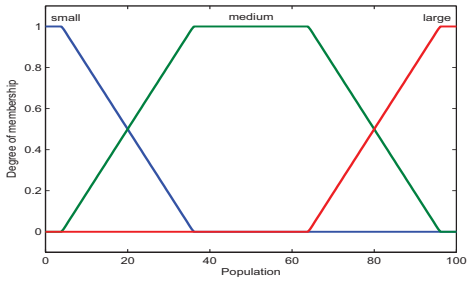


Figure 3: Input of variable *Population*.

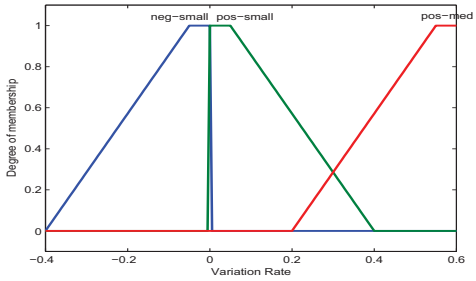


Figure 4: Input of variable *Variation Rate*.

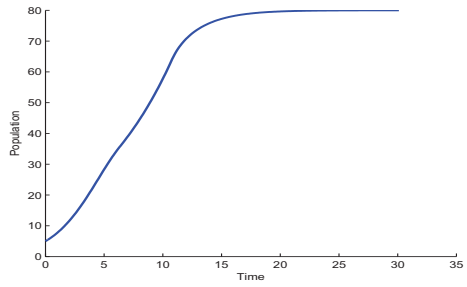


Figure 5: Solution $x_r(t)$ of simulation 1.

5.2 *Simulation 2*

In this second simulation we have eight rules in the rule base. They are as follows:

R_1: If the population is *"very small"* then the variation rate is *"positive very large"*
R_2: If the population is *"small A"* then the variation rate is *"positive large"*
R_3: If the population is *"small B"* then the variation rate is *"positive large"*
R_4: If the population is *"medium A"* then the variation rate is *"positive medium"*
R_5: If the population is *"medium B"* then the variation rate is *"positive medium"*
R_6: If the population is *"medium C"* then the variation rate is *"positive small"*
R_7: If the population is *"large"* then the variation rate is *"negative small"*
R_8: If the population is *"very large"* then the variation rate is *"negative medium"*

The membership functions of the input and output variables are in Figs. 6 and 7, respectively.

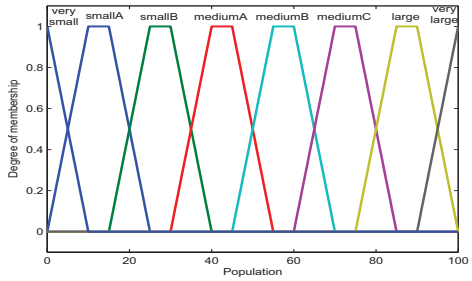


Figure 6: Input of variable *Population*.

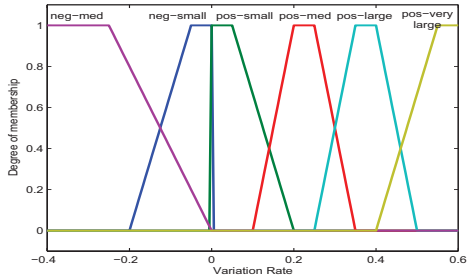


Figure 7: Input of variable *Variation Rate*.

The solution $x_r(t)$ is showed in Figure 8

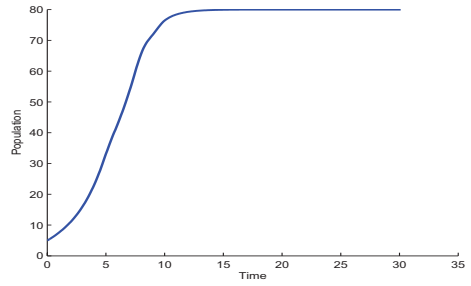


Figure 8: Solution $x_r(t)$ of simulation 2.

Fig. 9 shows the two solutions obtained in the simulations and the classical solution of the model, when adopted $K = 80$, $a = 0,006$ and $x_0 = 5$.

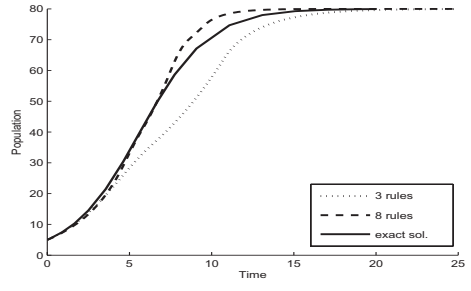


Figure 9: Comparison of the solutions obtained through simulations and the exact solution, using $x_0 = 5$.

Fig. 10 shows the two solutions obtained in the simulations

ISBN: 978-989-95079-6-8

and the classical solution of the model, when adopted $K = 80$, $a = 0,006$ and $x_0 = 90$.

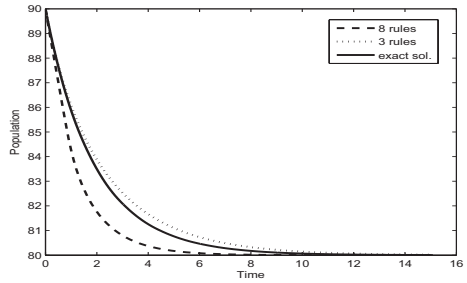


Figure 10: Comparison of the solutions obtained through simulations and the exact solution, using $x_0 = 90$.

Note that Figs. 9 and 10 show that the solutions of the IVP (3) converge to the exact solution of Verhulst Model. Moreover, the bigger the number of rules is, the better the approximation is. This is in accordance with Theorem 3, which ensures convergence as r increases. Table 1 shows the values obtained in the simulations and the value obtained by the solution of the Verhulst Model when $x_0 = 5$, for some values of t.

Table 1: Values obtained by simulation and by the equation.

Time	Theoretical Value (TV)	Sim. 1 (S1) (3 rules)	Sim. 2 (S2) (8 rules)
0	5	5	5
0,2	5,4152	5,4644	5,4460
0,9	7,4423	7,3920	7,2719
2,5	14,5304	13,7874	13,2925
4,6	30,4040	25,8578	28,7976
7,7	58,6421	42,7152	62,3510
11	74,6871	65,6782	78,3973
15,1	79,2707	77,3232	79,9304
17,1	79,7398	78,7905	79,9850
20	79,9407	79,6222	79,9984

The norm of the error of these approximations can be seen in Table 2. It is the difference between the exact value and the simulated values.

Table 2: Errors obtained comparing simulations 1 and 2 with the theoretical value.

Time	Error 1 ($\|TV - S1\|$)	Error 2 ($\|TV - S2\|$)
0	0	0
0,2	0,0491	0,0308
0,9	0,0504	0,1704
2,5	0,7430	2,2487
4,6	4,5416	1,6063
7,7	15,9269	3,7089
11	9,0059	3,7132
15,1	1,9475	0,6597
17,1	0,9493	0,2452
20	0,3185	0,0577

6 Conclusions

In this work we proposed a fuzzy rule-based system methodology to represent the direction field of differential equations. It is applicable when we have only partial knowledgement about the direction field in question. A typical case is when the informations about it are qualitative, as example, low, medium, high, conform the values of the state variable. As we have seen, this methodology was able to construct an input-output system from qualitative information by means of fuzzy sets. Thus, it was possible to adopt a direction field (f_r) (IVP (3)) obtained from theoretical direction field (f) (IVP (2)), that is given by Theorem 1, since f_r converges to f. Moreover, from theorem 3 and using the Lebesgue's Dominated Convergence Theorem we have proved that the solution of the IVP (3) converges to the solution of the IVP (2). Through a simple example we illustrated the methodology we proposed, confirming the obtained results. From the point of view of modeling this result is very important because in many cases, the available informations are qualitative. That is the case of Peixoto's work [19] that applies this methodology in a predator-prey model.

Acknowledgment

Authors thank CNPq (307890/2006-6 and 140297/2006-5) and FAPESP (06/05920-7).

References

[1] M. Turelli. Stochastic community theory: A partially guided tour. *Biomathematics*, 17:321 – 339, 1986.

[2] L. Arnold. *Random dynamical systems*. Springer, Verlag-Berlin-Heidelberg, Germany, 2nd edition.

[3] M. L. Puri and D. A. Ralescu. Differential of fuzzy functions. *Journ. Math. Analysis and Applications*, 91:552 – 558, 1983.

[4] H. Hullemeier. An approach to modelling and simulation of uncertain dynamical systems. *Intern. J. Uncertainty Fuzziness Knowledge-Based Systems*, 5:117–137, 1997.

[5] V. A. Baisodov. Fuzzy differential inclusions. *PMM USSR*, 54(1):8 – 13, 1990.

[6] M. T. Mizukoshi, L. C. Barros, Y. Chalco-Cano, Rman-Flores, and R. C. Bassanezi. Fuzzy differential equation and the extension principle. *Information Sciences*, 177, 2007.

[7] B. Bouchon-Meunier, D. Dubois, L. Godo, and H. Prade. *Fuzzy Sets in Approximate Reasoning and Information Systems*, chapter Fuzzy sets and possibility theory in approximate and plausible reasoning, pages 15–190. The Handbooks of Fuzzy Sets. Kluwer Academic Publisher, 1999.

[8] D. A. Ralescu, Y. Ogura, and S. Li. Set defuzzification and choquet integrals. *International J. Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:1 – 12, 2001.

[9] H. T. Nguyen, V. Kreinovich, and O. Sirisaengtaksin. Fuzzy control as a universal control tool. *Fuzzy Sets and Systems*, 80:71 – 86, 1996.

[10] H. T. Nguyen and E. A. Walker. *A First Course in Fuzzy Logic*. Chapman & Hall/CRC, 2000.

[11] L. A. Zadeh. Fuzzy sets. *Informat. Control*, 8:338 – 353, 1965.

[12] F. Gomide W. Pedrycz. *Fuzzy Systems Engineering - Toward Human-Centric Computing*. IEEE Press, John Wiley & Sons, Inc., 2007.

ISBN: 978-989-95079-6-8

[13] L. C. Barros and R. C. Bassanezi. *Tópicos de Lógica Fuzzy e Biomatemática*. Coleção IMECC Textos Didáticos, Campinas, 2006.

[14] J. J. Buckley and Y. Hayashi. Fuzzy input-output controllers are universal approximators. *Fuzzy Sets and Systems*, 58:273–278, 1993.

[15] B. Kosko. Fuzzy sets as universal approximators. *Proc. of IEEE Int. Conf. on Fuzzy Systems*, pages 1153–1162, Marc 8–12 1992.

[16] B. Kosko and J. A. Dickerson. *Theoretical Aspects of Fuzzy Control*, chapter Function Approximation with Additive Fuzzy Systems, pages 313–347. Wiley, 1995.

[17] A. A. Castro Jr. *Curso de teoria da Medida*. IMPA, Rio de Janeiro, 2004.

[18] M. R. B. Dias. Equações diferenciais ordinárias com campo de direções parcialmente conhecido. Master's thesis, IMECC - Unicamp, 2006.

[19] M. S. Peixoto, L. C. Barros, and R. C. Bassanezi. Predator-prey fuzzy model. *Ecological Modelling*, 214:39 – 44, 2008.

ISBN: 978-989-95079-6-8

Sparsely Connected Autoassociative Fuzzy Implicative Memories for the Reconstruction of Large Gray-Scale Images

Marcos Eduardo Valle
Department of Mathematics, University of Londrina
Londrina, Paraná, Brazil
Email: valle@uel.br

Abstract— *Autoassociative fuzzy implicative memories (AFIMs) are models that exhibit optimal absolute storage capacity and an excellent tolerance with respect to incomplete or eroded patterns. Thus, they can be effectively used for the reconstruction of gray-scale images. In practice, however, applications of AFIMs are confined to images of small size due to computational limitations. This paper introduces a class of sparsely connected AFIMs (SCAFIMs) that circumvent this computational overhead and, therefore, can be used for the reconstruction of large images. We show that SCAFIMs exhibit optimal absolute storage capacity and tolerance with respect to incomplete or eroded patterns. Furthermore, we compare the performance of SCAFIMs with their corresponding fully connected AFIM both theoretically and by means of computational experiments.*

Keywords— Fuzzy associative memory, reconstruction of large gray-scale images, sparse encoding, storage capacity, tolerance with respect to noise.

1 Introduction

Associative memories (AMs) are models inspirated in the human brain ability to recall by association [1, 2, 3, 4]. Here, a partial or approximate representation of a stored item is used to recall the full item. An example would be recalling a poem by knowing its first words as an initial clue. Applications of AM models include classification, prediction, control, and pattern recognition [5, 6, 7, 8, 9].

We speak of a *fuzzy associative memory* (FAM) if the AM model is used for the association of fuzzy sets [5, 8, 10, 11]. The *max-min* and *max-product FAMs* of Kosko, the *generalized FAMs* (GFAMs) of Chung and Lee, and the *implicative fuzzy associative memories* (IFAMs) are instances of FAMs [5, 8, 12, 13]. We would like to recall, however, that the FAMs of Kosko as well as the GFAMs usually fail to perfectly recall a memorized item due to crosstalk between the stored patterns [8, 12]. Therefore, despite their successful applications in a variety of problems [5, 12], the FAMs of Kosko and the GFAMs are not recommended for the reconstruction of gray-scale images.

In contrast, IFAMs exhibit *optimal absolute storage capacity* in the autoassociative case [13]. In other words, one can store has many patterns as desired in an *autoassociative fuzzy implicative memory* (AFIM). In addition, AFIMs exhibit one step-convergence and an excellent tolerance with respect incomplete or eroded patterns [8, 13]. As a consequence, AFIMs can be effectively used for the reconstruction of gray-scale images [8, 14].

Due to computational limitations, however, AFIMs can be used only for the storage and recall of images of small size. Specifically, such as several other AMs, including the *complex-valued neural network* [15, 16], the multi-state model of Costantini *et al.* [17, 18], and the *morphological associative memories* [14, 19], AFIMs require the storage of a large amount of synaptic junctions if used to process large images. For example, approximately 512 gigabytes of memory space would be consumed by an AFIM that is used for the storage of gray-scale images of size 512×512. In order to circumvent this computational overhead, we propose to remove a large amount of synaptic junctions of an AFIM. The resulting model is referred to as a *sparsely connected AFIM* (SCAFIM). We show that SCAFIMs inherit the optimal absolute storage capacity and the tolerance with respect to incomplete or eroded patterns of AFIMs. Therefore, they can be used for the reconstruction of large gray-scale images.

The paper is organized as follows. The next section reviews the basic concepts of AFIMs. SCAFIMs are introduced in Section 3. Some theoretical results concerning the storage capacity and noise tolerance of SCAFIMs, as well as two strategies for the construction of a SCAFIM from a (fully connected) AFIM, are also given in Section 3. Section 4 provides computational experiments. The paper finishes with the concluding remarks in Section 5.

2 A Brief Review on Autoassociative Fuzzy Implicative Memories

Associative memories (AMs) are input-output systems that store a set of associations $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1, \ldots, k\}$, called *fundamental memory set* [2, 3]. Formally, an AM corresponds to a mapping G such that $G(\mathbf{x}^\xi) = \mathbf{y}^\xi$ for every $\xi = 1, \ldots, k$. In addition, an AM model should be endowed with a certain tolerance with respect to noise. In other words, the mapping G should be such that $G(\tilde{\mathbf{x}}^\xi)$ equals $\mathbf{y}^\xi$ even for noisy or incomplete versions $\tilde{\mathbf{x}}^\xi$ of $\mathbf{x}^\xi$.

An associative memory can be classified as *autoassociative* or *heteroassociative* [3]. We have an *autoassociative memory* if $\mathbf{y}^\xi = \mathbf{x}^\xi$ for all $\xi = 1, \ldots, k$. We speak of an heteroassociative memory otherwise. This paper focus on the autoassociative case. The famous *Hopfield network* is an example of autoassociative memory for binary patterns [2, 4, 20].

We speak of a *fuzzy associative memory* (FAM) if the mapping G is given by a fuzzy neural network and the patterns $\mathbf{x}^\xi$ and $\mathbf{y}^\xi$ represent finite fuzzy sets for every $\xi = 1, \ldots, k$ [10, 11]. Applications of FAMs include backing up a truck and trailer, target tracking, and forecasting the average monthly streamflow of a large hydroelectric plant [5, 8]. This paper focus only on the class of *autoassociative fuzzy implicative memories* (AFIMs). The reader interested in a comprehensive survey on FAM models is invited to read [8].

ISBN: 978-989-95079-6-8

2.1 Autoassociative Fuzzy Implicative Memories

AFIMs are single-layer feedforward fuzzy neural networks equipped with neurons that compute the maximum of continuous triangular norms (t-norms) [13, 21]. Specifically, given a synaptic weight matrix $W \in [0,1]^{n\times n}$ and a threshold vector $\boldsymbol{\theta} \in [0,1]^n$, the output $\mathbf{y} \in [0,1]^m$ of an AFIM is given by

$$\mathbf{y} = \mathcal{W}(\mathbf{x}) = (W \circ \mathbf{x}) \vee \boldsymbol{\theta}\,, \tag{1}$$

where $\mathbf{x} \in [0,1]^n$ is the input pattern and the symbol "$\circ$" denotes a max-T product[1]. Recall that the max-T product of two matrices $A \in [0,1]^{n\times k}$ and $B \in [0,1]^{k\times n}$, denoted by $C = A \circ B \in [0,1]^{n\times m}$, is defined as follows [10]:

$$c_{ij} = \bigvee_{\xi=1}^{k} T(a_{i\xi}, b_{\xi j})\,. \tag{2}$$

The synaptic weight matrix W and the threshold vector $\boldsymbol{\theta}$ of an AFIM model are computed by means of a recording recipe called *implicative fuzzy learning* (IFL). Formally, given a fundamental memory set $\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$, where each $\mathbf{x}^\xi \in [0,1]^n$, IFL defines $W \in [0,1]^{n\times n}$ and $\boldsymbol{\theta} \in [0,1]^n$ as follows [13, 21, 11]:

$$[W \,|\, \boldsymbol{\theta}] = \bigwedge \left\{ [A \,|\, \boldsymbol{\beta}] : (A \circ \mathbf{x}^\xi) \vee \boldsymbol{\beta} \leq \mathbf{x}^\xi,\ \forall\, \xi \in \mathcal{K} \right\}. \tag{3}$$

Here, $[W \,|\, \boldsymbol{\theta}]$ represents the $n \times (n+1)$ fuzzy matrix obtained by concatenating W and $\boldsymbol{\theta}$, and $\mathcal{K} = \{1, \ldots, k\}$.

It is important to note that IFL makes optimal use of the synaptic weights and thresholds of the AM model given by (1). In fact, if there exist $A \in [0,1]^{m\times n}$ and $\boldsymbol{\beta} \in [0,1]^n$ such that $(A \circ \mathbf{x}^\xi) \vee \boldsymbol{\beta} = \mathbf{x}^\xi$ for all $\xi \in \mathcal{K}$, then W and $\boldsymbol{\theta}$ given by (3) also satisfies $(W \circ \mathbf{x}^\xi) \vee \boldsymbol{\theta} = \mathbf{x}^\xi$ for all $\xi \in \mathcal{K}$. Furthermore, the inequalities $A \leq W$ and $\boldsymbol{\beta} \leq \boldsymbol{\theta}$ hold true.

Observe, in particular, that the identity matrix $I \in [0,1]^{n\times n}$ and the vector of zeros $\mathbf{0} = [0, \ldots, 0]^T \in [0,1]^n$ are such that $(I \circ \mathbf{x}^\xi) \vee \mathbf{0} = \mathbf{x}^\xi$ for every original pattern $\mathbf{x}^\xi$. Thus, the synaptic weight matrix W and the threshold vector $\boldsymbol{\theta}$ given by (3) satisfy $I \leq W$ and $(W \circ \mathbf{x}^\xi) \vee \boldsymbol{\theta} = \mathbf{x}^\xi$ for all $\xi \in \mathcal{K}$. As a consequence, one can store as many patterns as desired in an AFIM model [13].

The following proposition establishes a relationship between the output of an AFIM and the fixed points of the synaptic weight matrix W [13, 8, 11]. Recall that $\mathbf{z} \in [0,1]^n$ is a fixed point of $W \in [0,1]^{n\times n}$ if and only if $W \circ \mathbf{z} = \mathbf{z}$. We denote the set of all fixed points of W by $F(W)$, i.e., $F(W) = \{\mathbf{z} \in [0,1]^n : W \circ \mathbf{z} = \mathbf{z}\}$.

Proposition 1. *Given an input pattern $\mathbf{x} \in [0,1]^n$, the output $\mathcal{W}(\mathbf{x}) = (W \circ \mathbf{x}) \vee \boldsymbol{\theta}$ of an AFIM is the smallest fixed point $\mathbf{z}$ of W such that $\mathbf{z} \geq \mathbf{x}$ and $\mathbf{z} \geq \boldsymbol{\theta}$, i.e.,*

$$\mathcal{W}(\mathbf{x}) = \bigwedge \{\mathbf{z} \in F(W) : \mathbf{z} \geq (\mathbf{x} \vee \boldsymbol{\theta})\}\,. \tag{4}$$

Proposition 1 shows that AFIMs exhibit one-step convergence. Moreover, it gives useful insights on the noise tolerance of AFIMs. For example, Proposition 1 says that an AFIM recalls an original pattern $\mathbf{x}^\xi$ only if the input $\mathbf{x}$ is smaller than $\mathbf{x}^\xi$. In other words, AFIMs are suited for the reconstruction of patterns corrupted by erosive noise, but are incapable of handling dilative noise. Recall that a distorted version $\tilde{\mathbf{x}}^\xi$ of the original pattern $\mathbf{x}^\xi$ has undergone an *erosive change* if $\tilde{\mathbf{x}}^\xi \leq \mathbf{x}^\xi$ and a *dilative change* if $\tilde{\mathbf{x}}^\xi \geq \mathbf{x}^\xi$ [19]. Further information on the noise tolerance of AFIMs can be obtained by investigating $F(W)$. A complete characterization of the set of fixed points of a general class of FAMs can be found in [22].

Let us conclude this section by recalling that the synaptic weight matrix $W = (w_{ij}) \in [0,1]^{n\times n}$ and the threshold vector $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]^T \in [0,1]^n$ given by IFL can be easily computed by means of the following equations for every $i, j = 1, \ldots, n$:

$$w_{ij} = \bigwedge_{\xi=1}^{p} I_T(x_j^\xi, x_i^\xi) \quad \text{and} \quad \theta_i = \bigwedge_{\xi=1}^{p} x_i^\xi\,. \tag{5}$$

Here, the operator $I_T : [0,1] \times [0,1] \to [0,1]$ denotes the R-implication associated with the t-norm that is used in the recall phase [23, 10]. Recall that the R-implication associated with a continuous t-norm T is defined as follows:

$$I_T(x,y) = \bigvee \{z \in [0,1] : T(x,z) \leq y\}\,. \tag{6}$$

The following equations present the R-implications associated with the *minimum*, *product*, and *Lukasiewicz t-norm*, respectively [23, 10, 13]:

$$I_M(x,y) = \begin{cases} 1, & x \leq y \\ y, & x > y \end{cases} \qquad \text{(Gödel)} \tag{7}$$

$$I_P(x,y) = \begin{cases} 1, & x \leq y \\ y/x, & x > y \end{cases} \qquad \text{(Goguen)} \tag{8}$$

$$I_L(x,y) = 1 \wedge (1 - x + y) \qquad \text{(Lukasiewicz)} \tag{9}$$

3 Autoassociative Fuzzy Implicative Memories for Large Gray-Scale Images

First of all, recall that a gray-scale image $\mathbf{a}$ of size $M \times N$ can be identified with a finite fuzzy set $\mathbf{x} = [x_1, \ldots, x_n]^T$ of length $n = NM$. The fuzzy set $\mathbf{x}$ is obtained by confining the gray-scale values to the unit interval and by arranging the pixels in a column vector using the standard column-scan method. As a consequence, AFIMs can be used for the storage and retrieval of gray-scale images.

Note, however, that an AFIM needs a synaptic weight matrix $W \in [0,1]^{n\times n}$ and a threshold vector $\boldsymbol{\theta} \in [0,1]^n$. In other words, the model requires computation and storage of $n(n+1)$ values. The following example reveals that the required computational resources restrain the applications of AFIMs to small gray-scale images.

Example 1. Consider a gray-scale image of size 512×512 with 256 shades of gray. This image corresponds to a finite fuzzy set that can be represented by a vector of length $n = 512^2 = 262144$. As a consequence, an AFIM requires the computation and storage of approximately 6.9×10^9 values corresponding to the synaptic weights and thresholds. If these values are represented in a computer using 64-bit double precision floating points, then the model allocates approximately 512 gigabytes of memory space. Similar considerations reveal that AFIMs allocate approximately 2 and 32 gigabytes of memory spaces for the storage and recall of images of size 128×128 and 256×256, respectively.

[1] In this paper, the symbols "$\bigvee$" and "$\bigwedge$" denote the supremum (or maximum) and infimum (or minimum) operations, respectively.

ISBN: 978-989-95079-6-8

Concluding, due to computer memory limitations, applications of AFIMs are usually bounded to images of size less than or approximately equal to 128×128. The following subsections provide strategies that allow for the storage and recall of large gray-scale images in AFIM based models.

3.1 Sparsely Connected AFIMs

Several studies suggest that neurons in the parts of the human brain that exhibit functional properties of associative memory are connected to few other neurons [24, 25, 26]. Motivated by this biological remark, we will remove a considerable amount of synaptic junctions of an AFIM model, i.e., we will introduce zeros in the synaptic weight matrix given by (5). However, in view of the following theorem[2], we will not delete synaptic weights in the main diagonal of W.

Theorem 1. *Consider an arbitrary fundamental memory set $\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$, where each $\mathbf{x}^\xi \in [0,1]^n$. If $V \in [0,1]^{n\times n}$ and $\boldsymbol{\vartheta} \in [0,1]^n$ satisfy the inequalities $I \leq V \leq W$ and $\boldsymbol{\vartheta} \leq \boldsymbol{\theta}$, where $W \in [0,1]^{n\times n}$ and $\boldsymbol{\theta} \in [0,1]^n$ are given by IFL, then the following equation holds true for every $\xi = 1, \ldots, k$:*

$$\mathbf{x}^\xi = (V \circ \mathbf{x}^\xi) \vee \boldsymbol{\vartheta}\,. \tag{10}$$

In view of Theorem 1, we define a *sparsely connected AFIM* (SCAFIM) as an autoassociative max-T FAM with threshold $\boldsymbol{\vartheta} \leq \boldsymbol{\theta}$ and a synaptic weight matrix $I \leq V \leq W$ that has very few nonzero elements. Since V has a sparse structure, SCAFIMs usually do not require large computational resources. Thus, they are suited for the storage and recall of large gray-scale images.

By definition, SCAFIMs exhibit optimal absolute storage capacity such as the fully connected AFIMs. However, in contrast to the latter, SCAFIMs do not necessarily exhibit one-step convergence. Therefore, in analogy to the famous *Hopfield network*, we may employ SCAFIMs with feed-back [2, 20]. In this case, given an input pattern $\mathbf{x} \in [0,1]^n$, we define the following sequence where $\mathbf{x}(0) = \mathbf{x}$:

$$\mathcal{V}_t(\mathbf{x}) = \mathbf{x}(t) = \big[V \circ \mathbf{x}(t-1)\big] \vee \boldsymbol{\vartheta}\,, \quad \forall\, t = 1, 2, \ldots \tag{11}$$

The resulting model is referred to as *dynamic or recursive SCAFIM*. Note that $\mathcal{V}_t(\mathbf{x})$ represents the output of a dynamic SCAFIM after t steps. In particular, $\mathcal{V}_1(\mathbf{x})$ corresponds to the output of a single-step model.

The following theorem shows that the sequence $\mathcal{V}_t(\mathbf{x})$ given by (11) converges for every input pattern $\mathbf{x} \in [0,1]^n$. Furthermore, Theorem 2 below relates the limit of the sequence $\mathcal{V}_t(\mathbf{x})$, denoted by $\mathcal{V}_*(\mathbf{x})$, with the fixed points of the synaptic weight matrix V. Recall that the set of fixed points of V is defined as $F(V) = \{\mathbf{z} \in [0,1]^n : V \circ \mathbf{z} = \mathbf{z}\}$.

Theorem 2. *Given a fundamental memory set $\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$, define $\boldsymbol{\vartheta}$ and V such that $\boldsymbol{\vartheta} \leq \boldsymbol{\theta}$ and $I \leq V \leq W$, where W and $\boldsymbol{\theta}$ are given by IFL. Then, for every input pattern $\mathbf{x} \in [0,1]^n$, the sequence given by* (11) *is monotonically increasing and converges to the smallest fixed point $\mathbf{z}$ of V such that $\mathbf{z} \geq \mathbf{x}$ and $\mathbf{z} \geq \boldsymbol{\vartheta}$, i.e.,*

$$\mathcal{V}_*(\mathbf{x}) = \lim_{t\to\infty} \mathbf{x}(t) = \bigwedge \{\mathbf{z} \in F(V) : \mathbf{z} \geq (\mathbf{x} \vee \boldsymbol{\vartheta})\}. \tag{12}$$

[2]We would like to point out that we intend to publish the proofs of Theorems 1, 2, and 3 in an upcoming journal paper.

Note that Theorem 2 is analogous to Proposition 1. Thus, such as AFIMs, SCAFIMs cannot recall an original pattern $\mathbf{x}^\xi$ if the input $\mathbf{x}$ is greater than $\mathbf{x}^\xi$. More importantly, both Proposition 1 and Theorem 2 relate noise tolerance with the set of fixed points of the synaptic weight matrix. The following theorem establishes a relationship between $F(V)$ and $F(W)$. As a consequence, we are able to relate the noise tolerance of a certain SCAFIM and its corresponding fully connected AFIM. Theorem 3 below also relates the set of fixed points of V with the set of fixed points of a matrix U that has a smaller number of non-zeros entries than V.

Theorem 3. *Let $F(U), F(V)$ and $F(W)$ denote the set of fixed points of matrices $U, V, W \in [0,1]^{n\times n}$ with respect to a certain max-T product. If $I \leq U \leq V \leq W$, then $F(W) \subseteq F(V) \subseteq F(U)$.*

Remark 1. Theorems 2 and 3 tell us that the noise tolerance of dynamic SCAFIMs with respect to erosive noise diminishes - because the number of fixed points increases - as we remove more and more synaptic junctions of a fully connected AFIM. In contrast, a dynamic SCAFIMs exhibit better noise tolerance with respect to dilative noise than its corresponding AFIM. Let us clarify this remark by means of an illustrative example.

Example 2. Let W and $\boldsymbol{\theta}$ denote the synaptic weight matrix and threshold vector of an AFIM. Suppose that U and V are sparsely connected synaptic weight matrices that satisfy the inequalities $I \leq U \leq V \leq W$, i.e., U has fewer non-zeros elements than V. Figure 1 depicts the sets of fixed points $F(W)$, $F(V)$, and $F(U)$. Here, we represented the set $[0,1]^n$ by a line from $\mathbf{0} = [0, 0, \ldots, 0]^T$ to $\mathbf{1} = [1, 1, \ldots, 1]^T$ although this set is not *totally ordered*[3]. The fixed points of W, V, and U are represented by short vertical lines. Note that, in agreement with Theorem 3, we have depicted $F(W) \subseteq F(V) \subseteq F(U)$.

Suppose that $\mathbf{x}^\xi$ represents an original pattern and $\check{\mathbf{x}}$ corresponds to an eroded version of $\mathbf{x}^\xi$. Let $\mathcal{W}$ denote the AFIM model whose synaptic weight matrix is W and threshold vector $\boldsymbol{\theta}$ is shown in Figure 1. By Proposition 1, $\mathcal{W}(\check{\mathbf{x}}) = \mathbf{x}^\xi$ because $\mathbf{x}^\xi$ is the smallest fixed point of W greather than $\check{\mathbf{x}}$ and $\boldsymbol{\theta}$. Similarly, let $\mathcal{V}_*$ and $\mathcal{U}_*$ denote dynamic SCAFIMs with threshold vector $\boldsymbol{\vartheta} = \boldsymbol{\theta}$ and synaptic weight matrices V and U, respectively. In contrast to the AFIM model, the dynamic SCAFIMs $\mathcal{V}_*$ and $\mathcal{U}_*$ fail to recall the original pattern $\mathbf{x}^\xi$ because there are fixed points in $F(V)$ and $F(U)$ between $\check{\mathbf{x}} \vee \boldsymbol{\vartheta}$ and $\mathbf{x}^\xi$. Furthermore, observe that $\mathcal{V}_*$ outperforms the SCAFIM $\mathcal{U}_*$ because U has a fixed point between $\check{\mathbf{x}} \vee \boldsymbol{\vartheta}$ and $\mathcal{V}_*(\check{\mathbf{x}})$.

Conversely, let us suppose that $\hat{\mathbf{x}}$ represents a corrupted version of $\mathbf{x}^\xi$ contaminated by a small dilative noise. In view of the relationship $F(W) \subseteq F(V) \subseteq F(U)$, the dynamic SCAFIM $\mathcal{U}_*$ and the AFIM $\mathcal{W}$ yield the best and the worst results, respectively.

3.2 Strategies for the Construction of SCAFIMs

This subsection briefly answers the question as to derive an SCAFIM model from an AFIM. First, note that the threshold

[3]Recall that $(X, \leq)$ is *totally ordered*, also called *linearly ordered*, if for any $x, y \in X$, either $x \leq y$ or $y \leq x$. Note, however, that there are patterns $\mathbf{x}, \mathbf{y} \in [0,1]^n$ such that neither inequalities $\mathbf{x} \leq \mathbf{y}$ nor $\mathbf{y} \leq \mathbf{x}$ hold true. Thus, Figure 1 is merely illustrative.

ISBN: 978-989-95079-6-8

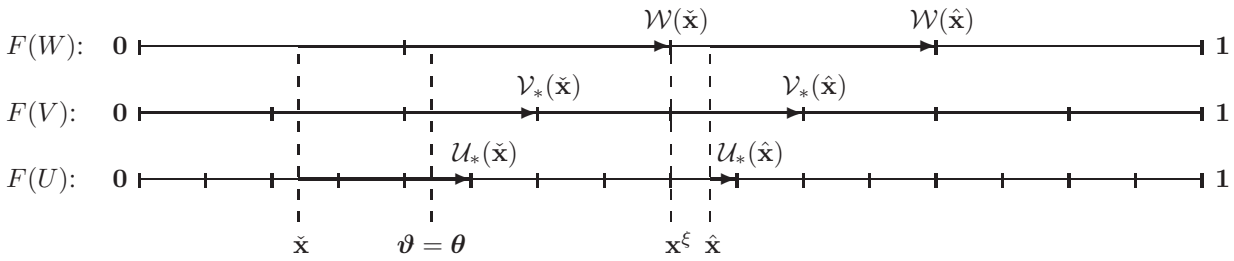


Figure 1: Illustrative example of the noise tolerance of an AFIM $\mathcal{W}$ and two dynamic SCAFIMs $\mathcal{V}_*$ and $\mathcal{U}_*$. Here, the synaptic weight matrices satisfy the inequalities $I \leq U \leq V \leq W$.

vector of an SCAFIM can be defined either as $\boldsymbol{\vartheta} = \boldsymbol{\theta}$ or $\boldsymbol{\vartheta} = \mathbf{0}$, depending on computational resources available. Let us consider $\boldsymbol{\vartheta} = \boldsymbol{\theta}$ in this paper. As far as we are concerned, the synaptic junctions of an AFIM can be removed in a strength-based manner or in a structured-based manner.

In the strength-based strategy, we delete synaptic weights by taking into account their strength. Specifically, we define the sparsely connected synaptic weight matrix $V \in [0,1]^{n \times n}$ as follows for some $\alpha \in (0,1]$:

$$v_{ij} = \begin{cases} w_{ij} & \text{if } w_{ij} \geq \alpha \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Note that the synaptic junctions with strength less than α are removed from the fully connected AFIM.

Alternatively, we say that we removed synaptic junctions in a structurated-based manner if we fix the non-zeros entries of the synaptic weight matrix before its computation. For example, we can set that the synaptic weight matrix V has a band or a block diagonal structure. We would like to point out that this strategy generalizes the widely used procedure of partitioning large gray-scale images in order to obtain several small-sized AM models [15, 16, 17].

4 Computational Experiments

This section provides some computational experiments concerning the reconstruction of noisy or incomplete gray-scale images. In order to compare the performance of SCAFIMs with the fully connected AFIM model, and due to page constrain, we only consider gray-scale images of small size. Specifically, let us consider the images of size 64×64 with 256 shades of gray displayed in the first row of Figure 2. For each of these image, we generated a vector $\mathbf{x}^\xi$ of length $n = 64 \times 64 = 4096$ using the standard column-scan method. Furthermore, we divided all entries of $\mathbf{x}^\xi$ by 255 in order to obtain patterns in the hypercube $[0,1]^{4096}$.

We stored the eight original patterns $\mathbf{x}^\xi$ in the Lukasiewicz AFIM, i.e., the AFIM based on the Lukasiewicz t-norm. We would like to recall that the Lukasiewicz AFIM outperformed several AM models, including the FAMs of Kosko and the GFAM based on the Lukasiewicz t-norm, in an experiment using faces images from the database of AT&T Laboratories, Cambridge [8, 13].

Next, we constructed six SCAFIM models, denoted by $\mathcal{V}^a, \mathcal{V}^b, \ldots, \mathcal{V}^f$, all of them with threshold vector $\boldsymbol{\vartheta} = \boldsymbol{\theta}$. The corresponding sparsely connected synaptic weight matrices $V^a, \ldots, V^f$ were derived as follows from W given by IFL: V^a has 16 blocks of size 256×256, V^b has 64 blocks of size 64×64, and V^c is a band matrix with bandwidth equals to 19. The synaptic weight matrices V^d, V^e, and V^f were obtained by means of (13) with $\alpha = 0.893, 0.961$, and 1, respectively. We would like to point out that we have chosen α and the bandwith of V^c in order to obtain pairs V^a and V^d, V^b and V^e, V^c and V^f with approximately the same number of non-zeros entries. Furthermore, the output of the SCAFIM $\mathcal{V}^b$ corresponds to the gray-scale image that is obtained by adopting the following procedure: Store each column of an original image in a separate Lukasiewicz AFIM. This results in 64 AFIMs, each of which has a 64×64 synaptic weight matrix that corresponds to a block of V^b. The recalled image is obtained by feeding each AFIM with its corresponding column of a given image and by concatenating the outputs.

The first row of Table 1 provides the density (i.e., percentage of non-zeros entries) of the synaptic weight matrices W, $V^a, \ldots, V^f$. Table 1 also provides the memory space required by the seven AM models if double precision numbers are used to represent the synaptic weights and threshold values. We would like to point out that we used the standart row-column-value representation to store the sparsely connected synaptic weight matrices in the computer.

Afterward, we confirmed that the eight original patterns represent fixed points of $W, V^a, \ldots, V^f$, i.e., the Lukasiewicz AFIM as well as the six SCAFIMs exhibit optimal absolute storage capacity.

In order to compare the noise tolerance of the seven AM models, we first introduced as input an image corrupted by *pepper noise* with probability 0.3. The second row of Figure 2 shows the corrupted image and the corresponding recalled patterns. Table 1 contains the *base-10 logarithm* applied to the *arithmetic mean* of the *relative normalized mean squared error* (RNMSE) in 100 experiments. Recall that the RNMSE is given by the following equation where x_i, $\tilde{x}_i$, and y_i denote the i-th entry of the original, noisy, and recalled patterns, respectively:

$$RNMSE = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \tilde{x}_i)^2}\,. \tag{14}$$

We would like to point out that we have corrupted an original pattern that was randomly selected in each trial. Moreover, the six dynamic SCAFIMs were iterated until convergence. The arithmetic means of the number of iterations required for convergence are shown between parentheses in Table 1.

We also conducted similar experiments using *salt noise* with probability 0.01, *salt and pepper noise* with probability 0.05, *additive Gaussian noise* with mean 0 and variance 0.01,

ISBN: 978-989-95079-6-8

	$\mathcal{W}$	$\mathcal{V}^a$	$\mathcal{V}^b$	$\mathcal{V}^c$	$\mathcal{V}^d$	$\mathcal{V}^e$	$\mathcal{V}^f$
Density of the AM model	100%	6.25%	1.56%	0.46%	6.3%	1.42%	0.48%
Required memory space	134.2 MB	16.8 MB	4.2 MB	1.3 MB	17.0 MB	4.2 MB	1.3 MB
Tolerance w.r.t. pepper noise	-3.05 (1)	-2.47 (1)	-1.93 (1)	-1.78 (2.3)	-2.37 (2.6)	-1.27 (2.7)	-0.62 (1)
Tolerance w.r.t. salt noise	1.85 (1)	1.55 (1)	1.13 (1)	0.97 (5.3)	1.80 (8.9)	1.51 (9.7)	0.96 (1)
Tol. w.r.t. salt & pepper noise	13.84 (1)	9.53 (1)	5.00 (1)	3.76 (3.7)	13.08 (7.6)	8.01 (6.7)	3.17 (1)
Tol. w.r.t. add. gaussian noise	6.40 (1)	3.50 (1)	2.18 (1)	1.64 (4.1)	4.97 (7.5)	2.78 (7.3)	1.35 (1)
Tolerance w.r.t. speckle noise	7.64 (1)	4.27 (1)	2.67 (1)	1.80 (4.8)	5.47 (8.7)	2.17 (8.1)	1.09 (1)

Table 1: Comparison of the noise tolerance, density of synaptic junctions, and required computational resources of Lukasiewicz AFIM and six dynamic SCAFIMs.

and *multiplicative gaussian noise (speckle)* with mean 1 and variance 0.04. Table 1 provides the base-10 logarithm of the mean of the RNMSE in 100 experiments. Again, the mean of the number of iterations required for convergence is shown between parentheses. Figure 2 shows instances of the noisy images and corresponding recalled patterns.

Note that, in accordance with Remark 1, the noise tolerance with respect to erosive (pepper) noise diminishes - whereas the tolerance with respect to dilative (salt) and mixed (gaussian and speckle) noise increase - as we remove more and more synaptic junctions from a fully connected AFIM.

The following compares the two strategies for the construction of dynamic SCAFIMs. Let us begin by considering $\mathcal{V}^a$ and $\mathcal{V}^d$. Both SCAFIMs have approximately the same number of synaptic junctions. Thus, they required almost the same amount of memory space. However, the SCAFIM $\mathcal{V}^a$ outperformed $\mathcal{V}^d$ with respect to erosive, dilative, and mixed noise. Furthermore, $\mathcal{V}^a$ is preferable than $\mathcal{V}^d$ in computational terms. First, because V^a converged to the fixed point with only one iteration whereas V^b required in mean 7 iterations. Secondly, in the structurated-based strategy, we only compute the synaptic junctions that are really stored in the computer. In contrast, in the strength-based strategy, we compute all synaptic junctions but we keep only those with strength greater than or equal to α. In addition, we can determine exactly the amount of bytes that the SCAFIM $\mathcal{V}^a$ will require before the computation of V^a, but we can only estimate the memory space that V^d will consume. We reach a similar conclusion by comparing the SCAFIMs $\mathcal{V}^b$ and $\mathcal{V}^e$.

Finally, let us compare $\mathcal{V}^c$ and $\mathcal{V}^f$. Note that both models consumed almost the same amount of computational resources. Moreover, they exhibited similar tolerance with respect to dilative or mixed noise. However, the SCAFIM $\mathcal{V}^c$ outperformed $\mathcal{V}^f$ with respect to erosive noise, but the former usually required more than one iteration to converge[4].

5 Concluding Remarks

This paper introduces the class of SCAFIMs. In few words, a SCAFIM is obtained by removing a considerable amount of synaptic junctions of an AFIM model. The novel models inherit the optimal storage capacity of AFIMs but do not necessarily exhibit one step convergence. Nevertheless, we showed that dynamic SCAFIMs, i.e., SCAFIMs with feedback, yield a monotonically increasing sequence of patterns that converges to a fixed point of the underlying (sparse) synaptic weight matrix. Thus, such as the AFIMs, SCAFIMs exhibit tolerance with respect to incomplete or eroded patterns. Furthermore, we pointed out that the noise tolerance of SCAFIMs decreases as the density of synaptic junctions. In contrast, the tolerance with respect to dilative and mixed noise increases as the density of synaptic junctions decreases.

In addition, we briefly investigated two strategies for the construction of SCAFIMs: the structurated-based and the strength-based strategies. In computational terms, the former is usually preferable than the latter due to the following reasons. First, because we compute only the synaptic junctions that will be really stored in the computer. Secondly, because we can determine exactly the amount of bytes that the sparse matrix will require. Finally, computational experiments revealed that the structurated-based strategy usually yields better noise tolerance than the strength-based strategy. However, further research is needed on how to choose the best strategy for the construction of SCAFIMs.

[4] We would like to point out that the SCAFIM $\mathcal{V}^f$ corresponds to the *Gaines fuzzy morphological associative memory* (FMAM) that was introduced and investigated in [22]. This model always converges to a fixed point with only one step.

References

[1] J. Anderson. *An Introduction to Neural Networks*. MIT Press, MA, 1995.

[2] M. Hassoun, editor. *Associative Neural Memories: Theory and Implementation*. Oxford University Press, Oxford, U.K., 1993.

[3] T. Kohonen. *Self-Organization and Associative Memory*. Springer Verlag, 1984.

[4] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ, 1999.

[5] B. Kosko. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[6] B. Zhang, H. Zhang, and S. Ge. Face recognition by applying wavelet subband representation and kernel associative memory. *IEEE Transactions on Neural Networks*, 15(1):166–177, January 2004.

[7] D. Zhang and W. Zuo. Computational intelligence-based biometric technologies. *IEEE Computational Intelligence Magazine*, 2(2):26–36, May 2007.

[8] P. Sussner and M. Valle. Fuzzy associative memories and their relationship to mathematical morphology. In W. Pedrycz, A. Skowron, and V. Kreinovich, editors, *Handbook of Granular Computing*, chapter 33, pages 733–754. John Wiley and Sons, Inc., New York, 2008.

[9] S. Teddy, C. Quek, and E. Lai. PSECMAC: A novel self-organizing multiresolution associative memory architecture.

ISBN: 978-989-95079-6-8



Figure 2: First row displays the original gray-scale images of size 64×64 with 256 shades of gray. The first column of the following three rows shows patterns corrupted by *pepper noise* with probability 0.3, *salt noise* with probability 0.01, and *salt and pepper noise* with probability 0.05, respectively. The other seven columns (from left to right) exhibit the corresponding patterns that were recalled by the Lukasiewicz AFIM $\mathcal{W}$ and the SCAFIMs $\mathcal{V}^a$, $\mathcal{V}^b$, $\mathcal{V}^c$, $\mathcal{V}^d$, $\mathcal{V}^e$, and $\mathcal{V}^f$, respectively.

IEEE Transactions on Neural Networks, 19(4):689–712, April 2008.

[10] W. Pedrycz and F. Gomide. *Fuzzy Systems Engineering: Toward Human-Centric Computing*. Wiley-IEEE Press, New York, 2007.

[11] M. Valle and P. Sussner. A general framework for fuzzy morphological associative memories. *Fuzzy Sets and Systems*, 159(7):747–768, 2008.

[12] F. Chung and T. Lee. On fuzzy associative memory with multiple-rule storage capacity. *IEEE Transactions on Fuzzy Systems*, 4(3):375–384, 1996.

[13] P. Sussner and M. Valle. Implicative fuzzy associative memories. *IEEE Transactions on Fuzzy Systems*, 14(6):793–807, 2006.

[14] P. Sussner and M. Valle. Grayscale morphological associative memories. *IEEE Transactions on Neural Networks*, 17(3):559–570, May 2006.

[15] D. Lee. Improvements of complex-valued hopfield associative memory by using generalized projection rules. *IEEE Transactions on Neural Networks*, 17(5):1341–1347, September 2006.

[16] M. Müezzinoğlu, C. Güzeliş, and J. Zurada. A new design method for the complex-valued multistate Hopfield associative memory. *IEEE Transactions on Neural Networks*, 14(4):891–899, July 2003.

[17] G. Costantini, D. Casali, and R. Perfetti. Neural associative memory storing gray-coded gray-scale images. *IEEE Transactions on Neural Networks*, 14(3):703–707, May 2003.

[18] G. Costantini, D. Casali, and R. Perfetti. Associative memory design for 256 gray-level images using a multilayer neural network. *IEEE Transactions on Neural Networks*, 17(2):519–522, march 2006.

[19] G. Ritter, P. Sussner, and J. Diaz de Leon. Morphological associative memories. *IEEE Transactions on Neural Networks*, 9(2):281–293, 1998.

[20] J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554–2558, April 1982.

[21] M. Valle, P., and F. Gomide. Introduction to implicative fuzzy associative memories. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages 925 – 931, Budapest, Hungary, 2004.

[22] M. Valle and P. Sussner. Storage and recall capabilities of fuzzy morphological associative memories with adjunction-based learning. 2007. Submitted for publication.

[23] H. Nguyen and E. Walker. *A First Course in Fuzzy Logic*. Chapman & Hall/CRC, Boca Raton, 2 edition, 2000.

[24] H. Bosch and F. Kurfess. Information storage capacity of incompletely connected associative memories. *Neural Networks*, 11(5):869–876, 1998.

[25] B. Olshausen and D. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, August 2004.

[26] R. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107, June 2005.

ISBN: 978-989-95079-6-8

A Restriction Level Approach for the Representation and Evaluation of Fuzzy Association Rules

Miguel Delgado[1] M. Dolores Ruiz[1] Daniel Sánchez[1]

1.Dept. Computer Science and Artificial Intelligence, University of Granada
ETSIIT, C/Periodista Daniel Saucedo Aranda s/n, 18071, Granada
Email: mdelgado@ugr.es, {mdruiz, daniel}@decsai.ugr.es

Abstract— *The goal of this paper is to present a logic model with very good properties for the representation and evaluation of fuzzy association rules. For that we are going to combine two approaches. The first one is a model for the representation and evaluation of crisp association rules. The second one is a new proposal for the representation of imprecise properties (in particular for fuzzy sets) by using restriction levels, which verifies all the crisp logic equivalences [1]. Combining both approaches we will achieve a solid model, with very good properties for the representation of fuzzy rules and a very simple framework for extending the crisp quality measures for the evaluation of fuzzy rules.*

Keywords— Logic model, restriction level, RL-probability, RL-numbers, fuzzy association rules.

1 Introduction

Fuzzy association rules have been very developed and have been applied in numerous situations since their appearance in the last nineties. The first works used the Fuzzy Subsets Theory proposed by L.A. Zadeh [2] for putting into groups imprecise values with a clear semantic and later they were used for representing the different types of imprecision found in a stored database. The advantages of using fuzzy sets are on the one hand making smother the bounds and on the other hand give a formal representation for the knowledge semantically significative and meaningful for the user. In [3], [4] we can find a description of some of the most important works in the field of extracting fuzzy association rules.

There are some approaches which offer a framework for generalizing the quality measures for fuzzy association rules (see [5] and references in it). Moreover, although the evaluation measures can be generalized by distinct ways, it should be necessary to search a logic support or an axiomatic base which reflects the choice of a specific measurement. This work tries to follow this line of research and it also tries to establish a logic model that generalizes and preserves in a natural way all the properties of crisp association rules, as well as the different interest measures used in their evaluation. For that we make use of two approaches: the logic model proposed in [6],[7] and the RL-representation theory proposed in [1], [8], [9].

The logic model we are dealing with can manage with different kinds of association rules which might be useful for the user. There are several examples in [10] where the model offers a good formalization for mining exception rules and double rules. Furthermore, we have implemented an algorithm for extracting any kind of rule having the quality measures used for assessing the validity of the rule represented by this model. The algorithm uses the items representation by means of bitsets [11] which reduces de time consuming when computing the conjunction between items and their associated support (bitset cardinality).

The rest of the paper follows with the approach for representing imprecise properties by means of restriction levels presented in [1]. Then we do a brief summary of a model proposed in [7] for the representation of crisp association rules which was widely developed in [6]. The last part of the paper contains our proposal for the representation and the evaluation of fuzzy association rules using a combination of the approaches presented in the previous sections. We finish with the conclusions and possible lines for future research.

2 Representation by Restriction Levels

An imprecise property in an universe X can be represented by a collection of crisp realizations. The approach of [1] for representing imprecise properties by means of restriction levels extends the usual operations from the crisp to the fuzzy case satisfying the logic equivalences between that operations.

Definition 1. [1] A *RL-set* Λ is a finite set of restriction levels $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$, $m \geq 1$.

In general, the *RL*-set of an atomic property represented by means of a fuzzy set A is defined as follows.

Definition 2. [1] Let be A a fuzzy set defined on the referencial X. Then the *RL*-set associated to A is given by:

$$\Lambda_A = \{A(x) \,|\, x \in X\} \cup \{1\} \tag{1}$$

where $A(x)$ is the grade of membership of x to the fuzzy set A.

The *RL*-set employed for representing an imprecise property is obtained by the union of the *RL*-sets associated to the atomic properties which define that property.

For representing an imprecise property in X by means of restriction levels we are going to use a *RL-representation* defined like a pair (Λ, ρ) where Λ is a *RL*-set and $\rho : \Lambda \to \mathcal{P}(X)$ is a function which applies each restriction level into a crisp realization in this level. For example, the *RL*-representation of an imprecise atomic property defined by a fuzzy set A will be the pair (Λ_A, ρ_A), where Λ_A is given by the equation (1) and $\rho_A(\alpha) = A_\alpha = \{x \in X \,|\, A(x) \geq \alpha\}$ for all $\alpha \in \Lambda_A$.

ISBN: 978-989-95079-6-8

Given an imprecise property P represented by (Λ_P, ρ_P), the set of crisp representatives of P, Ω_P is defined as

$$\Omega_P = \{\rho_P(\alpha) \mid \alpha \in \Lambda_P\}. \quad (2)$$

Definition 3. [1] Let be (Λ, ρ) a *RL*-representation with $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$. Let $\alpha \in (0,1]$ and $\alpha_i, \alpha_{i+1} \in \Lambda$ satisfying that $\alpha_i > \alpha > \alpha_{i+1}$. Then we define

$$\rho(\alpha) = \rho(\alpha_i). \quad (3)$$

If we look to this definition, this extension for values that there are not in the *RL*-set of the function ρ, is the natural extension if we think in a fuzzy set A and its α-cuts. Using this definition the concept of equivalence between two *RL*-representations is defined.

Definition 4. [1] Let (Λ, ρ) and (Λ', ρ') be two *RL*-representations on X. We will say that both representations are *equivalent* and will be noted by $(\Lambda, \rho) \equiv (\Lambda', \rho')$, if and only if, $\forall \alpha \in (0,1]$

$$\rho(\alpha) = \rho(\alpha'). \quad (4)$$

2.1 Logic Operations

In this section we present a brief overview of the logic operations necessary for the understanding of the rest of the paper. In particular we present the logic operations of disjunction, conjunction and negation that we will need for the generalization of the logic model to the case of fuzzy association rules. The basic ideas of how they are defined can be found in [1].

Definition 5. Let P, Q be two imprecise properties with *RL*-representations (Λ_P, ρ_P), (Λ_Q, ρ_Q). Then, $P \wedge Q$, $P \vee Q$ and $\neg P$ are imprecise properties represented by $(\Lambda_{P\wedge Q}, \rho_{P\wedge Q})$, $(\Lambda_{P\vee Q}, \rho_{P\vee Q})$ and $(\Lambda_{\neg P}, \rho_{\neg P})$ respectively, where

$$\begin{aligned} \Lambda_{P\wedge Q} &= \Lambda_{P\vee Q} = \Lambda_P \cup \Lambda_Q \\ \Lambda_{\neg P} &= \Lambda_P \end{aligned} \quad (5)$$

and, for all $\alpha \in (0,1]$,

$$\begin{aligned} \rho_{P\wedge Q}(\alpha) &= \rho_P(\alpha) \cap \rho_Q(\alpha), \\ \rho_{P\vee Q}(\alpha) &= \rho_P(\alpha) \cup \rho_Q(\alpha), \\ \rho_{\neg P}(\alpha) &= \overline{\rho_P(\alpha)}, \end{aligned} \quad (6)$$

where $\overline{Y}$ is the usual complement of a crisp set Y.

Proposition 1. [1] The operations $\wedge, \vee, \neg$ between *RL*-representations verify the ordinary properties of logic equivalence as for example $\neg\neg A \equiv A$, the Morgan's laws $(\neg(A \wedge B) \equiv (\neg A \vee \neg B)$, one of them) and the law of excluded middle that can be expressed as $A \wedge \neg A \equiv \bot$ or $A \vee \neg A \equiv \top$, where $\top$ and $\bot$ are the atomic properties which represent the tautology (whose *RL*-representation is obtained from the referencial X) and the contradiction (obtained from $\emptyset$).

2.2 RL-numbers

On the basis of *RL*-representations and operations, we introduced in [8] the *RL*-numbers as a representation of fuzzy quantities. This approach offers two main advantages: (1) *RL*-numbers are representations of fuzzy quantities that can be easily obtained by extending usual crisp measurements to fuzzy sets. (2) Arithmetic and logical operations on *RL*-numbers are straightforward and unique extensions of the operations on crisp numbers, verifying the usual properties of crisp arithmetic and logical operations. In addition, the imprecision does not necessarily increase through operations, and it can even diminish. The following definitions and properties are from [8]:

Definition 6. A *RL*-real number is a pair $(\Lambda, \mathcal{R})$ where Λ is a *RL*-set and $\mathcal{R} : (0,1] \to \mathbb{R}$.

We shall note $\mathbb{R}_{RL}$ the set of *RL*-real numbers. The *RL*-real number R_x is the representation of a (precise) real number x iff $\forall \alpha \in \Lambda_{R_x}$, $\mathcal{R}_{R_x}(\alpha) = x$. We shall denote such *RL*-real number as R_x or, equivalently, x, since in the crisp case, the set Λ_{R_x} is not important. Operations are extended as follows:

Definition 7. Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $R_1, \ldots, R_n$ be *RL*-real numbers. Then $f(R_1, \ldots, R_n)$ is a *RL*-real number with

$$\Lambda_{f(R_1,\ldots,R_n)} = \bigcup_{1 \le i \le n} \Lambda_{R_i} \quad (7)$$

and, $\forall \alpha \in \Lambda_{f(R_1,\ldots,R_n)}$

$$\mathcal{R}_{f(R_1,\ldots,R_n)}(\alpha) = f(\mathcal{R}_{R_1}(\alpha), \ldots, \mathcal{R}_{R_n}(\alpha)) \quad (8)$$

It is obvious that operations defined in this way are consistent extensions of crisp operations. We want to remark that operations not defined for certain combinations of real values are not defined for *RL*-numbers that verify that combination in at least one restriction level. This is the case of division by 0 for example, i.e., R/R' is defined iff $0 \notin \Omega_{R'}$.

2.3 Probabilities based on RL-numbers

Let us note $[0,1]_{RL} \subseteq \mathbb{R}_{RL}$ the set of *RL*-real numbers verifying $0 \le R \le 1$ $\forall R \in [0,1]_{RL}$ (see definition of *RL*-ranking in [9]). We introduce the notions of *RL*-probability space and *RL*-probability as in [9]:

Definition 8. A *RL*-probability space is a triple (X, Σ, P) where

1. X is a crisp set.
2. Σ is a collection of imprecise events defined by *RL*-representations on X, closed under complement and countable unions, and verifying $X \in \Sigma$.
3. $P : \Sigma \to [0,1]_{RL}$ verifying Kolmogorov's Axioms, i.e., $P(X) = 1$, $P(E) \ge 0$ $\forall E \in \Sigma$, and for any finite collection of disjoint representations $E_1, \ldots, E_n$, $P(E_1 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i)$.

In the previous definition, P is a *RL*-probability measure. In particular, *RL*-probability measures can be obtained from ordinary probability measures in an easy way as follows: let (U, F, p) be a (crisp) probability space and let U_F^{RL} be the set of all the *RL*-representations on U such that for any $A \in U_F^{RL}$, $\Omega_A \subseteq F$. By the properties of F, U_F^{RL} is closed under complement and countable unions, and $U \in U_F^{RL}$. For any $A \in U_F^{RL}$, let $P(A) \in [0,1]_{RL}$ defined by $\rho_{P(A)}(\alpha) = p(\rho_A(\alpha))\forall \alpha \in \Lambda_A$. It is easy to show that (U, U_F^{RL}, P) is a *RL*-probability space and P is a *RL*-probability measure.

ISBN: 978-989-95079-6-8

Also easy to show, *RL*-probabilities verify the addition law $P(A\cup B)=P(A)+P(B)-P(A\cap B)$ and the property $P(\neg A)=1-P(A)$. Finally, by definition 8, it is obvious that crisp probabilities are a particular case of *RL*-probabilities since crisp events are particular cases of imprecise events as described by *RL*-representations.

It is easy to check that the usual arithmetic relations between probabilities are preserved in each restriction level and, hence, they are preserved by the arithmetic of *RL*-numbers. For example, $p(A)+p(\neg A)=1$.

The concept of conditional probability is straightforwardly extended as follows:

Definition 9. Let A and B be two imprecise events defined on U by *RL*-representations $A=(\Lambda_A,\rho_A)$ and $B=(\Lambda_B,\rho_B)$ with $\rho_{p(B)}>0\ \forall\alpha\in(0,1]$. The *RL*-probability of A given B in U is

$$p(A|B)=\frac{p(A\wedge B)}{p(B)} \tag{9}$$

Proposition 2. [9] Let A and B be two imprecise events defined on U by *RL*-representations $A=(\Lambda_A,\rho_A)$ and $B=(\Lambda_B,\rho_B)$ with $\rho_{p(B)}>0\ \forall\,\alpha\in(0,1]$. Then

$$\rho_{p(A|B)}(\alpha)=\frac{\rho_{p(A\wedge B)}(\alpha)}{\rho_{p(B)}(\alpha)}. \tag{10}$$

Example 1. Let X and Y be atomic imprecise events defined by the following fuzzy sets on $U=\{u_1,\ldots,u_6\}$:

$$\begin{aligned} X&=1/u_1+0.8/u_2+0.5/u_3+0.4/u_5\\ Y&=0.9/u_1+0.6/u_3+0.5/u_4 \end{aligned} \tag{11}$$

Then we have $\Lambda_X=\Lambda_{\neg X}=\{1,0.8,0.5,0.4\}$ and $\Lambda_Y=\Lambda_{\neg Y}=\{1,0.9,0.6,0.5\}$. The *RL*-set for any operation between X and Y is $\Lambda_X\cup\Lambda_Y$. Table 1 shows the *RL*-representation of X, $\neg X$, Y, $X\wedge Y$, $X\vee Y$ and $X\wedge\neg Y$. Assuming that every u_i is equally probable, table 2 shows the corresponding *RL*-probabilities of imprecise events $p(X)$, $p(\neg X)$, $p(Y)$, $p(X\wedge Y)$, $p(X\vee Y)$ and $p(X\wedge\neg Y)$ and $p(Y|X)$. Remark that the *RL*-probability $p(X|Y)$ is undefined since $\rho_{p(Y)}(1)=0$.

3 A Logic Model for Association Rules

The logic model we are going to use is based in a method developed in the sixties by Háyek et al. [6]. This method calls GUHA (General Unary Hypotheses Automaton) and it has a good logic and statistical base which contributes to a better understanding of two important aspects of association rules: their nature and the properties of the interest measures used for their evaluation. Recently, several authors have implemented a good and fast algorithm [11] based on a data representation using bit strings. We also use a similar algorithm based on bitsets (an item representation by means of a set of bits) that is low time consuming for computing the items conjunction and their cardinality (support). Having the representation of every item into bitsets computing the contingency table which is based this model is straightforward and low time consuming [11].

Let M be a data matrix where its rows $O_1,\ldots,O_n$ are associated to a set of observed objects, and its columns $A_1,\ldots,A_K$ are the associated attributes which describe the objects. In this way, the entry (i,j) of M will be 1 when the object O_i satisfies the attribute A_j and 0 in other case. In the association rules framework, each matrix M will represent a transaction and the set of all matrixes (transactions) will be the database called D.

D	$<O_1,A_1>$	$\ldots$	$<O_n,A_K>$
t_1	1	$\ldots$	0
t_2	0	$\ldots$	1
$\vdots$	$\vdots$	$\ddots$	$\vdots$
t_n	1	$\ldots$	1

For the logic model we are presenting, an *item* will be a pair of the form $<O_1,A_1>$, and an *itemset* will be an aggregation of items using the usual logic connectors: $\wedge,\vee,\neg$. An association rule in the model proposed in [6] is an expression of the type $\varphi\approx\psi$ where φ and ψ represent itemsets (in the sense before) derived from a database D, and the symbol $\approx$ called *quantifier* is an evaluation or condition for the fulfilment of the association rule which will depend on the interest measure used and on the *four fold table*, $4ft$ associated to the itemsets φ and ψ. An example of association rule could be $<O_1,A_1>\wedge<O_3,A_2>\approx<O_2,A_5>\wedge\neg<O_3,A_7>$.

For any pair of attributes φ and ψ the so called *four fold table* may be constructed from the database D as follows:

$\mathcal{M}$	ψ	$\neg\psi$
φ	a	b
$\neg\varphi$	c	d

This table will be represented by $\mathcal{M}=4ft(\varphi,\psi,D)=\langle a,b,c,d\rangle$ where a,b,c and d will be non negative integers satisfying that a is the number of objects (i.e. the number of rows of D) which contain at the same time the itemsets φ and ψ, b the number of objects satisfying φ and not ψ, and analogously for c and d. It is obvious that the inequality $a+b+c+d>0$ is always satisfied.

The association rule $\varphi\approx\psi$ will be true in the database D (or in the matrix $\mathcal{M}$) if and only if the condition associated to the $4ft$-quantifier $\approx$ is satisfied for the four fold table $4ft(\varphi,\psi,D)$.

Depending on the type of $4ft$-quantifier we can express different kinds of associations between the itemsets φ and ψ. In [6] and [7] we can find some examples. The classical framework of support and confidence can be modeled by means of two quantifiers: the *support* and the *implication* $4ft$-quantifiers as follows:

$$\approx_{supp}(a,b,c,d)=\frac{a+b}{a+b+c+d},\quad\Rightarrow(a,b)=\frac{a}{a+b} \tag{12}$$

which must exceed the $0<minsupp,minconf<1$ thresholds respectively imposed by the user. In [12] we explain the existing relation between the $4ft$-quantifiers and the interestingness measures used in the evaluation and validation of association rules.

4 A Logic Model for Fuzzy Association Rules

In [13] the concepts of transaction and association rule are generalized to the fuzzy case.

Definition 10. [13] Let $I=\{i_1,\ldots,i_m\}$ be a finite set of items. A fuzzy transaction is a non empty fuzzy subset $\tilde{\tau}\subseteq I$.

ISBN: 978-989-95079-6-8

Table 1: *RL*-representations associated to several imprecise events derived from the atomic properties X and Y.

α_i	$\rho_X(\alpha)$	$\rho_{\neg X}(\alpha)$	$\rho_Y(\alpha)$	$\rho_{X\wedge Y}(\alpha)$	$\rho_{X\vee Y}(\alpha)$	$\rho_{X\wedge\neg Y}(\alpha)$
1	$\{u_1\}$	$\{u_2,u_3,u_4,u_5,u_6\}$	$\emptyset$	$\emptyset$	$\{u_1\}$	$\{u_1\}$
0.9	$\{u_1\}$	$\{u_2,u_3,u_4,u_5,u_6\}$	$\{u_1\}$	$\{u_1\}$	$\{u_1\}$	$\emptyset$
0.8	$\{u_1,u_2\}$	$\{u_3,u_4,u_5,u_6\}$	$\{u_1\}$	$\{u_1\}$	$\{u_1,u_2\}$	$\{u_2\}$
0.6	$\{u_1,u_2\}$	$\{u_3,u_4,u_5,u_6\}$	$\{u_1,u_3\}$	$\{u_1\}$	$\{u_1,u_2,u_3\}$	$\{u_2\}$
0.5	$\{u_1,u_2,u_3\}$	$\{u_4,u_5,u_6\}$	$\{u_1,u_3,u_4\}$	$\{u_1,u_3\}$	$\{u_1,u_2,u_3,u_4\}$	$\{u_2\}$
0.4	$\{u_1,u_2,u_3,u_5\}$	$\{u_4,u_6\}$	$\{u_1,u_3,u_4\}$	$\{u_1,u_3\}$	$\{u_1,u_2,u_3,u_4,u_5\}$	$\{u_2,u_5\}$

Table 2: *RL*-probabilities in U of the imprecise events in table 1 and the conditional *RL*-probability $p(Y|X)$.

α_i	$\rho_{p(X)}(\alpha)$	$\rho_{p(\neg X)}(\alpha)$	$\rho_{p(Y)}(\alpha)$	$\rho_{p(X\wedge Y)}(\alpha)$	$\rho_{p(X\vee Y)}(\alpha)$	$\rho_{p(X\wedge\neg Y)}(\alpha)$	$\rho_{p(Y\|X)}(\alpha)$
1	1/6	5/6	0	0	1/6	1/6	0
0.9	1/6	5/6	1/6	1/6	1/6	0	1
0.8	1/3	2/3	1/6	1/6	1/3	1/6	1/2
0.6	1/3	2/3	1/3	1/6	1/2	1/6	1/2
0.5	1/2	1/2	1/2	1/3	2/3	1/6	2/3
0.4	2/3	1/3	1/2	1/3	5/6	1/3	1/2

For every item $i \in I$ and every transaction $\tilde{\tau}$, an item i will belong to $\tilde{\tau}$ with grade[1] $\tilde{\tau}(i)$ where $\tilde{\tau}(i)$ is a real number in the interval $[0,1]$.

Let $A \subseteq I$ be an itemset. The membership grade of A to the fuzzy transaction $\tilde{\tau}$ is defined as

$$\tilde{\tau}(A) = \min_{i\in A} \tilde{\tau}(i).$$

Using the definition 10 a crisp transaction will be a special case of fuzzy transaction where every item in the transaction will have membership grade equal to 1 or 0 depending on if they are in the transaction or not.

Example 2. We consider the set of items $I = \{i_1, i_2, i_3, i_4, i_5\}$ and the set of fuzzy transactions given by the table 3.

Table 3: Set of fuzzy transactions $\tilde{D}_1$

	i_1	i_2	i_3	i_4	i_5
$\tilde{\tau}_1$	1	0.2	1	0.9	0.9
$\tilde{\tau}_2$	1	1	0.8	0	0
$\tilde{\tau}_3$	0.5	0.1	0.7	0.6	0
$\tilde{\tau}_4$	0.6	0	0	0.5	0.5
$\tilde{\tau}_5$	0.4	0.1	0.6	0	0
$\tilde{\tau}_6$	0	1	0	0	0

In particular, we can see that $\tilde{\tau}_6$ is a crisp transaction. Some membership grade could be: $\tilde{\tau}_1(\{i_3,i_4\})$ =0.9, $\tilde{\tau}_1(\{i_2,i_3,i_4\})$ =0.2 and $\tilde{\tau}_2(\{i_1,i_2\})$ =1.

Definition 11. [13] Let I be a set of itmes, $\tilde{D}$ a set of fuzzy transactions and $A,B \in I$ two disjoint itemsets, i.e. $A \cap B = \emptyset$. A fuzzy association rule is satisfied in $\tilde{D}$ if and only if, $\tilde{\tau}(A) \leq \tilde{\tau}(B)$ for all $\tilde{\tau} \in \tilde{D}$, that is, the membership grade of B is higher than the membership grade of A for all fuzzy transactions $\tilde{\tau}$ in $\tilde{D}$.

[1]Note that $\tilde{\tau}(i)$ is $\mu_{\tilde{\tau}}(i)$ where $\mu_{\tilde{\tau}} : I \longrightarrow [0,1]$ is the membership function associated to the fuzzy set $\tilde{\tau}$ on the referencial $I = \{$ set of items $\}$.

This definition maintains the meaning of crisp association rules because if we need that $A \subseteq \tilde{\tau}$ is satisfied, we also need that $B \subseteq \tilde{\tau}$ be satisfied, in our case this can be translated to $\tilde{\tau}(A) \leq \tilde{\tau}(B)$. In this way, as a crisp transaction is a special case of fuzzy transaction, a crisp association rule will be a special case of fuzzy association rule.

Let $\tilde{\Gamma}_A$ and $\tilde{\Gamma}_B$ two fuzzy sets defined on $\tilde{D}$ as $\tilde{\Gamma}_A(\tilde{\tau}) = \tilde{\tau}(A)$ and $\tilde{\Gamma}_B(\tilde{\tau}) = \tilde{\tau}(B)$ respectively. As both are fuzzy sets, we can obtain their own *RL*-representations that we note[2] as $(\Lambda_{\tilde{A}}, \rho_{\tilde{A}})$, $(\Lambda_{\tilde{B}}, \rho_{\tilde{B}})$, defined as the equations (1) and (3) show.

For the crisp case, some of the suggested measures for the validation of association rules are defined using the concept of probability. This is the case of support and confidence. Using the *RL*-probabilities we can easily extend the classical framework of support and confidence.

Definition 12. (Support of an itemset) Let $A \subseteq I$ be an itemset and $(\Lambda_{p(\tilde{A})}, \rho_{p(\tilde{A})})$ the *RL*-representation of the *RL*-probability associated to the fuzzy set $\tilde{\Gamma}_A$ in $\tilde{D}$. Then, the support of A in the fuzzy database $\tilde{D}$ is defined as

$$supp(A) = \sum_{\alpha_i \in \Lambda_{p(\tilde{A})}} (\alpha_i - \alpha_{i+1}) \left(\rho_{p(\tilde{A})}(\alpha_i)\right). \qquad (13)$$

Following a similar reasoning we will define the support and the confidence for a fuzzy association rule $A \rightarrow B$.

Definition 13. Let be $A,B \subseteq I$ two disjoint itemsets and $(\Lambda_{p(\tilde{A}\wedge\tilde{B})}, \rho_{p(\tilde{A}\wedge\tilde{B})})$, $(\Lambda_{p(\tilde{B}|\tilde{A})}, \rho_{p(\tilde{B}|\tilde{A})})$ the *RL*-representations of the *RL*-probabilities $p(\tilde{\Gamma}_A \wedge \tilde{\Gamma}_B)$ and $p(\tilde{\Gamma}_B|\tilde{\Gamma}_A)$ in $\tilde{D}$. Then, the support and the confidence of a fuzzy association rule $A \rightarrow B$ on $\tilde{D}$ are defined as

$$Supp(A \rightarrow B) = \sum_{\alpha_i \in \Lambda_{p(\tilde{A}\wedge\tilde{B})}} (\alpha_i - \alpha_{i+1}) \left(\rho_{p(\tilde{A}\wedge\tilde{B})}(\alpha_i)\right) \qquad (14)$$

$$Conf(A \rightarrow B) = \sum_{\alpha_i \in \Lambda_{p(\tilde{B}|\tilde{A})}} (\alpha_i - \alpha_{i+1}) \left(\rho_{p(\tilde{B}|\tilde{A})}(\alpha_i)\right) \qquad (15)$$

[2]A and B are crisp sets of disjoint items and $\tilde{\Gamma}_A$, $\tilde{\Gamma}_B$ are fuzzy sets defined on $\tilde{D}$

ISBN: 978-989-95079-6-8

We want to remark that in the previous definitions $\Lambda_{p(\tilde{A}\wedge\tilde{B})}$ and $\Lambda_{p(\tilde{B}|\tilde{A})}$ coincides and are equal to $\Lambda_{\tilde{A}}\cup\Lambda_{\tilde{B}}$.

But some other interesting measures are not in terms of probability. For those, the model is necessary for the generalization of those crisp measures of interest.

The following itemsets: $A\wedge B, A\wedge\neg B, \neg A\wedge B, \neg A\wedge\neg B$ are a partition of the database $\tilde{D}$ (in the same way as $\varphi\wedge\psi, \varphi\wedge\neg\psi, \neg\varphi\wedge\psi$ and $\neg\varphi\wedge\neg\psi$ are a partition for the crisp case). We take their associated fuzzy sets defined on $\tilde{D}$ with their own *RL*-representations that we are going to call $(\Lambda_{\tilde{A}\wedge\tilde{B}}, \rho_{\tilde{A}\wedge\tilde{B}})$, $(\Lambda_{\tilde{A}\wedge\neg\tilde{B}}, \rho_{\tilde{A}\wedge\neg\tilde{B}})$, etc. Note that the *RL*-sets obtained contain the same restriction levels, that is, the set of restriction levels is common and equal to $\Lambda_{\tilde{A}}\cup\Lambda_{\tilde{B}}$.

As for every $\alpha\in\Lambda_Y$, $\rho_Y(\alpha)$ is a crisp set, we can compute its cardinality as usual, and we note it by $|\rho_Y(\alpha)|$.

In this way, for every restriction level $\alpha_i\in\Lambda_{\tilde{A}}\cup\Lambda_{\tilde{B}}$ we define the associated four fold table $\mathcal{M}_{\alpha_i} = 4ft(\tilde{\Gamma}_A, \tilde{\Gamma}_B, \tilde{D}, \alpha_i)$ as follows:

$\mathcal{M}_{\alpha_i}$	$\tilde{\Gamma}_B$	$\neg\tilde{\Gamma}_B$
$\tilde{\Gamma}_A$	a_i	b_i
$\neg\tilde{\Gamma}_A$	c_i	d_i

where a_i, b_i, c_i and d_i are non negative integers such that

$$a_i = |\rho_{\tilde{A}\wedge\tilde{B}}(\alpha_i)|$$
$$b_i = |\rho_{\tilde{A}\wedge\neg\tilde{B}}(\alpha_i)|$$

and analogously with c_i and d_i.

Note that $\forall\alpha_i\in\Lambda_{\tilde{A}}\cup\Lambda_{\tilde{B}}$ the following equality is satisfied

$$a_i+b_i+c_i+d_i = n = |\tilde{D}|. \tag{16}$$

We are going to prove that this definition is a good generalization to the crisp case. Suppose that A and B are two itemsets in a crisp database D, in this case, $\tilde{\Gamma}_A(\tilde{\tau}) = \tilde{\tau}(A)\in[0,1]$ is equivalent to $t(A)\in\{0,1\}$ where $t(A)$ is given by the following indicator function:

$$t(A) = \begin{cases} 1 & \text{si } A\in t \\ 0 & \text{si } A\notin t \end{cases} \tag{17}$$

where t is a crisp transaction of D. The associated *RL*-sets to $\tilde{\Gamma}_A$ and to $\tilde{\Gamma}_B$ will be $\Lambda_{\tilde{A}} = \Lambda_{\tilde{B}} = \{1\}$. And the associated *RL*-representations will be: $(\Lambda_{\tilde{A}}, \rho_{\tilde{A}})$ y $(\Lambda_{\tilde{B}}, \rho_{\tilde{B}})$ where

$$\begin{aligned} \rho_{\tilde{A}}(1) &= (\tilde{\Gamma}_A)_1 = \{t\in D\,|\,t(A)\geq 1\} = \{t\in D\,|\,A\in t\} \\ \rho_{\tilde{B}}(1) &= (\tilde{\Gamma}_B)_1 = \{t\in D\,|\,t(B)\geq 1\} = \{t\in D\,|\,B\in t\} \end{aligned} \tag{18}$$

Following the same process, we can compute the associated *RL*-representations for the sets $\tilde{\Gamma}_A\wedge\tilde{\Gamma}_B$, $\tilde{\Gamma}_A\wedge\neg\tilde{\Gamma}_B$, $\neg\tilde{\Gamma}_A\wedge\tilde{\Gamma}_B$, $\neg\tilde{\Gamma}_A\wedge\neg\tilde{\Gamma}_B$. Therefore, the $4ft$ table for the restriction level $\alpha=1$ is the same as the $4ft$ table for the itemsets A and B seen in section 3:

$\mathcal{M}_1\equiv\mathcal{M}$	B	$\neg B$
A	a_1	b_1
$\neg A$	c_1	d_1

Using this representation it is easy to generalize every kind of interest measure used in the crisp case, in particular, we can generalize every $4ft$-quantifier.

If we look again to the classical framework of support and confidence, we can extend it as we show from now on.

Let be $A, B\subseteq I$ two disjoint itemsets and $(\Lambda_{\tilde{A}}, \rho_{\tilde{A}}), (\Lambda_{\tilde{B}}, \rho_{\tilde{B}})$ the *RL*-representations associated to the fuzzy sets $\tilde{\Gamma}_A$ and $\tilde{\Gamma}_B$ in $\tilde{D}$. Then, the support and the confidence of a fuzzy association rule $A\rightarrow B$ on $\tilde{D}$ are defined as

$$Supp(A\rightarrow B) = \sum_{\alpha_i\in\Lambda_{\tilde{A}}} (\alpha_i-\alpha_{i+1})\,(\approx_{supp}(a_i,b_i,c_i,d_i)), \tag{19}$$

$$Conf(A\rightarrow B) = \sum_{\alpha_i\in\Lambda_{\tilde{A}}} (\alpha_i-\alpha_{i+1})\,(\Rightarrow(a_i,b_i)). \tag{20}$$

The support and the confidence defined above coincide with those defined by means of probabilities based on RL-numbers. In this last equation (20) we can find an indetermination of the type "$\frac{0}{0}$", in this case, we will take the value 1 for preserving the definition 11 of fuzzy association rule, because this indetermination happens when there is no transactions satisfying the antecedent and the consequent of the rule among zero transactions satisfying only the antecedent. Again, it is easy to see that for every restriction level α_i the last part of the equations (19) and (20) are the same as the defined support and confidence for a crisp association rule (see quantifiers in (12)).

This process can be done for every measure of interest used in the validation and evaluation of crisp association rules. In particular we use the certainty factor instead of confidence because of its proven good properties: (1) it reduces the number of obtained rules and (2) it provides an useful meaning to the measurement of the validity of association rules. In [12], we prove by means of the logic model that the certainty factor can be seen as an equivalence $4ft$-quantifier as follows:

$$\equiv_{CF}(a,b,c,d) = \begin{cases} \dfrac{ad-bc}{(a+b)(b+d)} & \text{if } ad>bc \\ 0 & \text{if } ad=bc \\ \dfrac{ad-bc}{(a+b)(a+c)} & \text{if } ad<bc. \end{cases} \tag{21}$$

Its generalization for fuzzy rules is straightforward:

$$CF(A\rightarrow B) = \sum_{\alpha_i\in\Lambda_{\tilde{A}}} (\alpha_i-\alpha_{i+1})\,(\equiv_{CF}(a_i,b_i,c_i,d_i)). \tag{22}$$

We want to remark that the previous concepts of support, confidence and the certainty factor perfectly extends the support, confidence and certainty factor measures used for crisp association rules. This can be immediately proved because the measures obtained in equations from 19 to 21 coincide with the ones proposed in [13] where it is used a semantic approach based on the evaluation of quantified sentences (see also [14]) using the *GD* method and the fuzzy relative quantifier the majority $Q_M(x)=x$ for evaluating the sentences. In these papers it is proved that using the *GD* method and the quantifier Q_M, the obtained measures in equations (19)-(22) are the ordinary support, confidence and certainty factor for the rule $A\rightarrow B$ respectively.

In addition, using the proposed logic model which uses the representation by means of the $4ft$ tables for each restriction

ISBN: 978-989-95079-6-8

level, is immediate to prove that if A and B are two itemsets in a set of crisp transactions, the concepts of support, confidence and certainty factor for instance (19)-(22), are equal to the usual crisp concepts of support, confidence and certainty factor (see again equations (12) and (21)).

5 Concluding Remarks

We have presented a simple model with very good properties for the representation and evaluation of fuzzy association rules which extends in a natural way the evaluation of crisp association rules.

We have combined two existing approaches: a logic model for crisp association rules [7] and the recently developed RL-representation theory for imprecise properties [1]. In particular, using the philosophy of RL-representations we define the RL-numbers for the representation of fuzzy quantities [8] and the associated probabilities which are useful for extending crisp quality measures like support and confidence. RL-numbers and probabilities have been also introduced in [15],[16] named gradual numbers. Both approaches are similar with slight differences that mainly affect how is the negation considered [15]. The main similarities and differences between both approaches is treated in [17]. But this is beyond the scope of this paper.

We used RL-numbers and their associated probabilities for the special cases of computing support and confidence of fuzzy association rules but there are many cases that the rule's assessment is made by other measures which are not in terms of probabilities. For these cases we also provide a very simple and systematic approach for extending the crisp measures for mining fuzzy association rules, for instance the certainty factor is extended by equation (22).

In this line of research we have found works like the one in [5]. They develop a systematic approach to the assessment of fuzzy association rules. To this end, they proceed from the idea of partitioning the data stored in a database into examples of the given rules, counterexamples and irrelevant data. Maybe what they called irrelevant data is not so "irrelevant" for measuring the accomplishment of an association rule. See for example the case of the certainty factor in equation (21) which uses the so called irrelevant data (divided into two different frequencies, c and d) for measuring the strength of the rule, and in addition, its good properties are experimentally proved in [13] and formally proved in [12] using the introduced logic model.

Our approach has some similarities with the approach in [5] as we also divide the data, but we use a conjunction-based model, taking all the possible conjunctions between rule's antecedent and consequent. The main difference between the former and our proposal remains in the way of generalizing the quality measures for fuzzy associations, as we use a four fold table's RL-representation.

Since algorithmic aspects of fuzzy rule mining have not been addressed in this paper, let us mention that the method used in [11] for mining crisp rules using the bitset approach can be easily adapted to the fuzzy case and the complexity will increase depending on the number of restrictions levels taken into account.

Acknowledgment

This work has been partially supported by the projects TIN2006-15041-C04-01 and TIN2006-07262.

References

[1] D. Sánchez, M. Delgado, and M.A. Vila. A restriction level approach to the representation of imprecise properties. In *IPMU'08*, pages 153–159, Málaga, Spain, 2008.

[2] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[3] T-P Hong and Y-C Lee. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, chapter An Overview of Mining Fuzzy Association Rules, pages 397–410. Springer, 2008.

[4] E. Hüllermeier. Fuzzy sets in machine learning and data mining. *Applied Soft Computing, to appear*, 2008.

[5] D. Dubois, E. Hüllermeier, and H. Prade. A systematic approach to the assessment of fuzzy association rules. *Data Min. Knowl. Discov.*, 13(2):167–192, 2006.

[6] P. Hájek and T. Havránek. *Mechanising Hypothesis Formation-Mathematical Foundations for a General Theory*. Springer-Verlag: Berlin, Heidelberg and New York, 1978.

[7] J. Rauch. Logic of association rules. *Applied Intelligence*, 22:9–28, 2005.

[8] D. Sánchez, M. Delgado, and M.A. Vila. Rl-numbers: An alternative to fuzzy numbers for the representation of imprecise quantities. In *WCCI 2008*, Hong Kong, China, To appear.

[9] D. Sánchez, M. Delgado, and M.A. Vila. A restriction level approach to probability and statistics. In *ESTYLF'08*, pages 107–112, Asturias, Spain, 2008.

[10] M. Delgado, M.D. Ruiz, and D. Sánchez. Analyzing exception rules. In *IPMU'08*, Málaga, Spain.

[11] J. Rauch and M. Šimunek. An alternative approach to mining association rules. *Studies in Computational Intelligence (SCI)*, 6:211–231, 2005.

[12] M. Delgado, M.D. Ruiz, and D. Sánchez. Studying interest measures for association rules through a logical model. *IJUFKS*, Submitted, 2007.

[13] M. Delgado, N. Marín, D. Sánchez, and M.A. Vila. Fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–225, 2003.

[14] M. Delgado, M.D. Ruiz, and D. Sánchez. Fuzzy association rules: new challenges (in spanish). In *ESTYLF'08*, Langreo-Mieres, Spain.

[15] D. Dubois and H. Prade. Gradual elements in a fuzzy set. *Soft Computing*, 12:165–175, 2008.

[16] D. Dubois and H. Prade. Fuzzy elements in a fuzzy set. In *Proc. 10th Int. Fuzzy Systems Assoc. (IFSA)*, pages 55–60, Beijing, 2005.

[17] D. Sánchez, M. Delgado, and M.A. Vila. A discussion about representation models for imprecise quantities and probabilities. In *30th Linz Seminar on Fuzzy Set Theory*, Linz, Austria, 2009.

ISBN: 978-989-95079-6-8

Semantic Proximity Between Queries and the Empty Answer Problem

Patrick Bosc Carmen Brando Allel Hadjali Hélène Jaudoin Olivier Pivert

IRISA/ENSSAT, Université de Rennes 1
6, Rue de Kerampont, BP 80518, 22305 Lannion Cedex, France
Email: {bosc, brando, hadjali, jaudoin, pivert}@enssat.fr

Abstract— This paper proposes an approach aimed at obviating empty answers for a family of conjunctive queries involving Boolean or fuzzy value constraints. Contrary to the approaches based on a relaxation of the predicates involved in the query, the principle suggested here consists in replacing the query by another one similar which has been processed previously and whose answer is known to be non-empty. This technique thus avoids the combinatory explosion induced by classical relaxation-based approaches.

Keywords— Databases, fuzzy queries, empty answers, proximity.

1 Introduction

Since the late 80's, there is an increasing interest in designing intelligent information systems endowed with some cooperative behavior [12]. The most well-known issue approached in this field is the empty answer problem, that is, the problem of providing the user with some alternative data when there is no data fitting his/her query. Several approaches have been proposed to deal with this issue. The relaxation paradigm [13] is one of the basic cooperative techniques used in most of such approaches. Query relaxation aims at expanding the scope of a query searching for answers which are in the neighborhood of the user's query and consists in replacing some query conditions by more general conditions or in just eliminating some of them.

Let us note that manually relaxing failing queries is a tedious and time-consuming task because, in the worst case, one must consider an exponential number of possible relaxations [21]. Hence, several automated approaches to query relaxation have been proposed (see, e.g., [8, 11, 15, 17, 20, 21]). The main objective of those approaches is to modify a failing user query into a relaxed query whose answer is non-empty, or at least to identify the cause of the failure. Some of those works rely on the key concept of false presuppositions (a presupposition of a statement is any statement entailed by the original, for instance, the statement "the king of France is bald" has as presupposition "there is a king of France" which is a false presupposition). Motro [20] has addressed the issue of empty answers, i.e. when a query fails to produce any answers, by proposing a relaxation method which focuses on finding the false presuppositions of a failing query. A related approach has been proposed by Godfrey [15], who considers any subquery as a presupposition of the query itself. The focus of this work is the search for Minimal Failing Subqueries (MFSs) of a failing query.

In this paper, we propose an approach for dealing with failing conjunctive queries (Boolean or fuzzy), according to another approach than relaxation as described above. The idea that we advocate is not to modify/suppress some predicates from a failing query Q using this sole query, but rather to find a "good" global substitute to Q among queries previously submitted to the system whose answer is known to be non-empty. We thus consider a context where the system stores the non-failing queries in a repository D^+. One also assumes available a resemblance measure over every attribute domain involved in the database considered. The approach raises the question of defining the notion of semantic proximity between queries. With respect to related works, which will be briefly presented farther, the main originality of the approach introduced here is to take into account queries involving Boolean or fuzzy value constraints in an explicit form.

An important gain brought by this method, relatively to a classical relaxation-based approach, lies in the fact that it avoids the combinatory explosion induced by the relaxation of the predicates from a conjunctive query. Indeed, there exists in general a high number of relaxed queries and one cannot know whether these queries provide a non-empty answer before processing them. With the approach proposed here, one has the guarantee to obtain a non-empty answer in one step because only one query needs to be processed. In this paper, we limit the scope to conjunctive selection queries involving value constraints which can be represented either by crisp sets, intervals or fuzzy sets. The remainder of the paper is organized as follows. Section 2 presents a query substitution approach in the case where the predicates expressing value constraints are Boolean. Section 3 generalizes this approach to the case of conjunctive fuzzy queries. In Section 4, we describe the principle of a mechanism aimed at providing the user with some (at least partial) explanations about the causes of the original empty answer. The technique proposed uses both the repository D^+ and a second one denoted by D^- which contains the failing queries previously submitted to the system. Section 5 is devoted to a comparison of the approach with some related works. Finally, the conclusion recalls the main contributions of the paper and outlines some perspectives for future work.

2 Boolean Queries

2.1 Single-Predicate Queries

Let $Q = (A \text{ in } E)$ denote the user query where A is an attribute and E a set or an interval, and Q' a query from the repository D^+. We denote by r the database relation concerned by Q and by res the proximity relation defined over the domain of A.

2.1.1 Case where E is a finite set

It is assumed that the user query Q returns an empty answer. One thus has to search through D^+ so as to retrieve the queries

ISBN: 978-989-95079-6-8

involving a predicate of the form (A in E').

Remark 1. Any set E' present in the queries from D^+ is such that $E' \not\subseteq E$. Otherwise the associated query Q' would have returned an empty answer and Q' would not be in D^+.

The emptiness of the answer to Q means that none of the elements from E is present as an A-value in relation r. In order to obtain a substitute to Q which returns a non-empty answer, it is thus necessary to find a query Q' bearing on r and involving a predicate (A in E') such that E' contains at least one value absent from E. However, so as not to drift too far away from the initial user need, it is desirable that the elements from E' absent from E be sufficiently close to at least one element from E. Consequently, one looks for the set E' which is as close as possible to E, so as to replace E by $(E' - E)$ in Q. In order to find this "best" E', a measure is needed. It is not strictly speaking a proximity measure, since the symmetry property is not desired here. Indeed one wants to know whether E' is a good substitute to E, but not necessarily the reciprocal. Several possible substitutivity measures (denoted by sbs_i later on) are discussed hereafter.

1st idea: one assesses the extent to which every element from $(E' - E)$ resembles at least one element from E:

$$sbs_1(E,\ E') = inf_{x' \in (E'-E)}\ sup_{x \in E}\ res(x,\ x'). \quad (1)$$

The problem with this measure is that the worst element "masks" the others, as illustrated in the next example. In the following, we assume available the following subset of a resemblance relation on animals:

res(rooster, hen) = 0.9, res(rooster, duck) = 0.6, res(rooster, turkey) = 0.7, res(hen, duck) = 0.6, res(hen, turkey) = 0.7, res(cow, hen) = res(cow, duck) = res(cow, turkey) = 0.

Example 1. E = {hen, duck, turkey}, E'_1 = {hen, turkey, cow}, E'_2 = {cow, rooster}. We get $sbs_1(E,\ E'_1) = sbs_1(E,\ E'_2) = 0$ but since there are neither hens nor turkeys in the database — otherwise Q would not be failing —, it seems reasonable to claim that E'_2 should be a better substitute than E'_1. However, in the computation of $sbs_1(E,\ E'_2)$, the element cow "masks" rooster.◇

2nd idea: one assesses the extent to which there is an element from $(E' - E)$ which resembles at least an element from E:

$$sbs_2(E,\ E') = sup_{x' \in (E'-E)}\ sup_{x \in E}\ res(x,\ x'). \quad (2)$$

Here, the difficulty is that the "winning set" may include elements which are very distant from those desired by the user.

Example 2. E = {hen}, E'_1 = {hen, cow, turkey}, E'_2 = {duck}. Here, E'_2 should win, since it includes only elements close to the desired ones, contrary to E'_1, which includes "cow". However, it is E'_1 which wins since $sbs_2(E,\ E'_1) = 0.7$ while $sbs_2(E,\ E'_2) = 0.6$.◇

3rd idea: one mixes the quantitative and the qualitative aspects by measuring the average resemblance degree between an element from $(E' - E)$ and an element from E. For each element x' from $(E' - E)$, the corresponding measure looks for the maximal proximity between x' and an element x from E, computes the sum of these maximal proximities, and divides this sum by the number of elements present in $(E' - E)$:

$$sbs_3(E,\ E') = \frac{\sum_{x' \in (E'-E)}\ sup_{x \in E}\ res(x,\ x')}{|E' - E|}. \quad (3)$$

Example 3. E = {hen, duck, turkey}, E'_1 = {hen, turkey, cow}, E'_2 = {cow, rooster}. We get: $sbs_3(E,\ E'_1) = 0$ and $sbs_3(E,\ E'_2) = 0.45$.◇

Since measure sbs_3 appears the most satisfactory, it will be used in the following.

2.1.2 *Case where E is an interval*

It is quite straightforward to extend measure sbs_3 defined previously so as to make it work with intervals instead of sets: one just has to replace the sum by an integral. The calculus is rather simple since it boils down to computing areas of rectangles or trapezoids.

Let us first consider the simple case where resemblance is defined in a Boolean manner:

$$res(x,\ y) = 1 \text{ if } |x - y| \leq \delta, 0 \text{ otherwise.} \quad (4)$$

Let us consider two intervals: $I = [m,\ M]$, that from the user query, and $I' = [m',\ M']$, that from the candidate substitute. Let us assume that $m \leq m'$, the dual case can be obtained straightforwardly from this one. One gets:

$$sbs_3(I, I') = \begin{cases} 0 \text{ if } M + \delta \leq m', \\ 1 \text{ if } M' \leq M + \delta, \\ (M + \delta - p)/(M' - p) \text{ otherwise,} \end{cases} \quad (5)$$

where $p = m'$ if $m' \geq M$, M otherwise. A slightly more complex case is that where resemblance is defined by means of a fuzzy tolerance indicator Z with a trapezoidal membership function of support $[-\alpha,\ \alpha]$ and of core $[-\beta,\ \beta]$ [4]. In other words:

$$res(x,\ y) = \mu_Z(|x - y|). \quad (6)$$

The principle is the same as for Boolean resemblance, except that one has to compute areas of trapezoids instead of rectangles. The case where the predicate from Q involves an interval I and that from Q' involves a set E can be managed by rewriting the definition of sbs_3 the following way:

$$sbs_3(E,\ E') = \frac{\sum_{x' \in E \wedge x' \notin I}\ sup_{x \in I}\ res(x,\ x')}{|\{x \in E \mid x \notin I\}|}. \quad (7)$$

On the other hand, the dual case (set in Q and interval in Q') is more tricky and cannot be captured by the formula defining sbs_3 when the attribute domain is continuous. Consequently, we introduce the constraint that a set can only be replaced by another set.

2.2 *Conjunctive Queries*

Let Q be the user query and Q' a query from the repository D^+.

ISBN: 978-989-95079-6-8

Remark 2. Even if query Q' involves a predicate which is subsumed by the corresponding one in Q, query Q' can be an interesting substitute to Q since the other predicates must also be taken into account. For instance, if query Q = (A in {rabbit, hen} and B in {wheat, cabbage}) returns an empty answer, it is still possible that query Q' = (A in {rabbit} and B in {wheat, oats}) returns a non-empty one whereas the predicate on A in Q' is subsumed by that in Q. For a query Q' to be a possible substitute, it is necessary that Q' involves at least one predicate which is not subsumed by the corresponding one in Q (but notice that if it were not the case, the answer to Q' would be empty — since the answer to Q is — and Q' would therefore not be in D^+).

Remark 3. The predicates from Q' which are strictly subsumed by those from Q can be replaced by the latter ones.

The substitution process that we propose consists of the following three steps:

1. select the candidate queries (and adapt these queries, see algorithm below),
2. compute the proximity degrees between the queries retained and the user query Q through the measure sbs_3,
3. determine the closest substitute to Q and process it.

Remark 4. For every predicate P from Q which is not "covered" by Q', i.e., which concerns an attribute on which there is no constraint in Q', one may compute the proximity between P and the entire domain of the attribute considered.

The conjunctive combination of the proximities related to the atomic predicates can be performed by means of a triangular norm, so as to obtain the overall proximity between two queries. Notice that alternative solutions could also be possible, for instance one might use a mean operator. The substitution algorithm is outlined hereafter.

Algorithm:

Let Q' be a query from D^+ which shares at least one attribute from its "where" clause with that from Q. The five steps of the algorithm are:

1. replace the "select" clause from Q' by that from Q;
2. remove from Q' every predicate that concerns an attribute absent from the "where" clause from Q;
3. replace every predicate from Q' which is strictly subsumed by the corresponding one from Q by the latter;
4. for the other predicates, compute the proximity between the predicate from Q' and the corresponding one from Q, by means of measure sbs_3, and replace the predicate from Q' by its union with that from Q. As to the predicates from Q which are not covered by Q', one computes their substitutivity degree relatively to the entire domain of the attribute involved;
5. aggregate the local proximities by means of a triangular norm (the idea is to assess the extent to which every predicate of the substitute query is close to the corresponding predicate from the initial failing query).

Table 1: Resemblance relation over attribute veg

	co	ra	su	wh	ca	br	po	ru
co	1	0.4	0.3	0.8	0.1	0.1	0.6	0.4
ra	0.4	1	0.9	0.6	0.2	0.2	0.1	0.1
su	0.3	0.9	1	0.5	0.1	0.1	0.3	0.3
wh	0.8	0.6	0.5	1	0.2	0.1	0.5	0.4
ca	0.1	0.2	0.1	0.2	1	0.9	0.6	0.7
br	0.1	0.2	0.1	0.1	0.9	1	0.4	0.6
po	0.6	0.1	0.3	0.5	0.6	0.4	1	0.8
ru	0.4	0.1	0.3	0.4	0.7	0.6	0.8	1

Example 4. Let Q be the following failing user query:

select #id **from** F
where veg **in** {corn, rapeseed} **and**
city **in** {Lannion, Caouennec, Prat} **and**
area **in** [60, 100].

Let us assume that the domain of "veg" is: {corn, rapeseed, sunflower, wheat, cabbage, broccoli, potato, rutabaga} and that the associated resemblance relation is given in Table 1. Let Q'_1 be the following query from D^+:

select #name **from** F
where veg **in** {wheat, rapeseed, sunflower} **and**
city **in** {Lannion, Prat} **and** area = 125 **and**
animal **in** {cow, pig}.

The query Q''_1 obtained by adapting Q'_1 according to the algorithm above is:

select #id **from** F
where veg **in** {corn, rapeseed, wheat, sunflower} and
city **in** {Lannion, Caouennec, Prat} and
(area **in** [60, 100] **or** area = 125).

The degree computed by sbs_3 for the substitution of {corn, rapeseed} by {wheat, rapeseed, sunflower} equals:

$$\frac{max(0.8,\ 0.6) + max(0.3,\ 0.9)}{2} = 0.85.$$

Let us assume that the proximity over the areas is based on a fuzzy tolerance indicator Z with a triangular membership function of support $[-50,\ 50]$. The substitution of [60, 100] by 125 is assigned the degree 0.5 (i.e., the proximity degree between 100 and 125). Finally, the degree computed for Q''_1 using the t-norm minimum is:

$$min(0.85, 0.5) = 0.5.$$

Let us now consider another query, denoted by Q'_2, from D^+:

select #name **from** F
where city = Caouennec **and**
area **in** [80, 180] **and** animal **in** {sheep, goat}.

Altering Q'_2 according to the algorithm yields Q''_2:

ISBN: 978-989-95079-6-8

select #id **from** F
where city **in** {Lannion, Caouennec, Prat} **and**
area **in** [60, 180].

As to the condition on attribute *veg* we get:
sbs_3({corn, rapeseed}, *domain(veg)*) =
(0.9 + 0.8 + 0.2 + 0.2 + 0.6 + 0.4)/6 = 0.52.

As to the condition on attribute *area*, we get:
sbs3([60, 100], [80, 180]) =
$((150 - 100)/2)/(180 - 100) = 25/80 = 0.31$.

Thus, the degree attached to Q_2'' is: $min(0.52, 0.31) = 0.31$ and Q_1'' is a better substitute to Q than Q_2''.$\diamond$

Remark 5. In case of ties, one could take into account the cardinality of the result of each candidate query so as to break these ties, provided that these cardinalities are stored in D^+.

3 Fuzzy Queries

Let us now move to the case where value constraints are expressed by means of fuzzy predicates. Let us consider a conjunctive fuzzy query $Q = P_1$ and ... and P_n where any predicate P_i is of the form (A_i is T_i) and T_i is a fuzzy term. Here, the fact that Q returns an empty answer means that there does not exist any element x in the database such that $\top_{i=1..n}(\mu_{T_i}(x)) > 0$, where $\top$ denotes a triangular norm generalizing the conjunction. This state of fact can be expressed by saying that the support of the query relatively to the database is empty.

In order to deal with this kind of queries, one needs to generalize measure sbs_3 by replacing the arithmetic mean by a weighted mean, and by taking into account the resemblance between the degrees coming from the two fuzzy terms considered. The generalized measure obtained, which can also be seen as a variant of the interchangeability measure proposed in [5], is defined as:

$$sbs_3(E,\ E') = \frac{\sum_{x' \in sp(E'-E)} w(x') \times \Gamma(x')}{\sum_{x' \in sp(E'-E)} w(x')} \qquad (8)$$

where:

$$\Gamma(x') = sup_{x \in E}\, min(res(x,\ x'),\ \Psi(\mu_{E'}(x'),\ \mu_E(x))), \quad (9)$$

$sp(E)$ denotes the support of E, function Ψ assesses the resemblance between two degrees in the unit interval — it can be defined e.g. as $\Psi(a,\ b) = 1 - |a - b|$ — and:

$$w(x') = (E' - E)(x') = min(E'(x'),\ 1 - E(x')). \quad (10)$$

The weight $w(x')$ captures the fact that it is all the more important to find a good substitute to x' as x' strongly belongs to $E' - E$. It is straightforward to show that if the sets are crisp, this formula reduces to that given in Section 2.

The definition above can be directly extended to the case of continuous fuzzy sets by replacing the sum by an integral.

Example 5. Let us consider the fuzzy sets: E = {1/rapeseed, 0.8/cabbage, 0.3/wheat} and E' = {0.4/rapeseed, 0.3/cabbage, 0.4/corn, 0.7/broccoli}. With the most commonly used definition of the difference between fuzzy sets, i.e.,

$$\mu_{(A-B)}(x) = min(\mu_A(x),\ 1 - \mu_B(x)), \qquad (11)$$

one gets: $E' - E$ = {0.2/cabbage, 0.4/corn, 0.7/broccoli}.

For "cabbage", the supremum equals:
sup(min(0.2, 0.3), min(1, 0.5), min(0.2, 1) = 0.5,

for "corn", it equals:
sup(min(0.4, 0.4), min(0.1, 0.6), min(0.8, 0.9)) = 0.8,

and for "broccoli" we get the degree 0.9. Hence, the final substitutivity degree equals:

(0.2×0.5 + 0.4×0.8 + 0.7×0.9)/(0.2 + 0.4 + 0.7) = 0.81.$\diamond$

The algorithm given in Subsection 2.2 can be adapted straightforwardly. The notion of subsumption between fuzzy predicates can be based on the inclusion between fuzzy sets proposed by Zadeh, i.e., $E \subseteq F \Leftrightarrow \forall x,\ \mu_E(x) \leq \mu_F(x)$.

4 Explaining the Emptiness of the Answer

Besides providing the user with a non-empty answer, it is important to also give him/her some explanations about:

- the reasons why the answer was originally empty,
- the way his/her query has been modified.

This section deals with the first point and gives the principle of a mechanism for explaining (at least partly) the emptiness of the original answer, somewhat in the spirit of [20].

4.1 Boolean Queries

Let us assume that one has available not only D^+ but also a repository D^- containing the failing queries submitted previously to the system. We suggest using both D^+ and D^- to identify some of the failing subqueries (in the sense of [15], i.e., the subsets of atomic conditions from the original query) which cause the empty answer.

Example 6. Let us consider the following failing query:

Q: veg ∈ {corn, wheat} ∧ animal ∈ {cow, pig}.

Let Q_1 and Q_2 be two queries from D^+:

Q_1: veg ∈ {corn, rapeseed} ∧ animal ∈ {sheep, pig},

Q_2: animal = pig,

and Q_3 a query from D^-: veg ∈ {rapeseed, broccoli}.

From $Q_3 \in D^-$, one can deduce that there is no rapeseed in the database relation considered. From this result and $Q_1 \in D^+$, one infers that there are some farms growing corn in the relation. From $Q_2 \in D^+$, one deduces that there are farms breeding pigs. From these two results, one can conclude that none of the terms from query Q leads to an empty answer. Consequently, the "minimal failing

ISBN: 978-989-95079-6-8

subquery" is Q itself, which means that there is some sort of "incompatibility" between (corn, wheat) and (cow, pig).$\diamond$

In order to provide the user with explanations, we suggest to partition the subqueries of Q into three classes: L_1 (those which are known to produce non-empty answers), L_2 (those which are known to be failing), L_3 (the others). Let us first consider the case of set-based predicates. A query from D^- (or D^+) is expressed: P_1 **and** ... **and** P_n where every P_i is a condition of the type (A_i **in** E_i), which corresponds to a disjunction ($A_i = v_{1,1}$ **or** ... **or** $A_i = v_{1,p}$). The first step is to transform every query from D^- into a rule:

($A_i = v_{1,1}$ **or** ... **or** $A_i = v_{1,p}$) **and** ...
and ($A_n = v_{n,1}$ or ... or $A_n = v_{n,k}$) $\rightarrow false$.

Symmetrically, every query from D^+ gives birth to a set of rules:

($A_i = v_{1,1}$ **or** ... **or** $A_i = v_{1,p}$) $\rightarrow true$
...
($A_n = v_{n,1}$ **or** ... **or** $A_n = v_{n,k}$) $\rightarrow true$.

These two types of rules constitute the rule base of the reasoning system used further.

Let us now consider a failing query Q. For each subquery of Q, one checks whether it is possible to infer *true* or *false* using the rule base. If it is possible to infer *true*, the subquery belongs to L_1, if one can infer *false* it belongs to L_2, otherwise it belongs to L_3. The subqueries have to be examined in increasing order of their size, starting with the atomic predicates, considering that a subquery including a failing subquery is itself failing (it is then useless to examine it). The corresponding algorithm, which exploits the concept of Minimal Failing Subqueries (MFS), is thus somewhat analogous to that proposed in [15], but here, we do not have to process any subquery to know if it is failing or not, due to the existence of the repositories D^+ and D^-.

In the case of intervals, one cannot replace a predicate by an explicit disjunction. One thus needs an inference engine able to deal with interval constraints.

4.2 Fuzzy queries

For discrete fuzzy sets, one uses the same principle as for Boolean queries. In the rules, the terms of a disjunction correspond to the values from the support of a fuzzy predicate. For continuous fuzzy sets, one cannot express rules involving explicit disjunctions. A query $Q = P_1$ and ... and P_n from D^- produces a rule:

A_1 **in** *support*(P_1) **and** ... **and** A_n **in** *support*(P_n) $\rightarrow false$

and a query $Q = P_1$ **and** ... **and** P_n from D^+ gives birth to the set of rules:

A_1 **in** *support*(P_1) $\rightarrow true$
...
A_n **in** *support*(P_n) $\rightarrow true$.

Here again, the reasoning engine must be able to deal with interval constraints.

5 Related Work

Some related work can be found in both domains of databases and information retrieval, including web search.

The semantic caching approach proposes to keep in a cache some previously executed queries along with their results and checks whether the answers to a given user query can be retrieved from the cache for optimization purposes. It uses the notion of query containment (i.e. a query Q is contained in a query G, if all answers to Q are also answers to G) to find the answers in the cache (see, e.g., [1, 7, 16, 19]). In a similar spirit, Ghosh *et al.* [14] propose a query clustering approach aimed at optimizing queries by reusing execution plans computed for similar queries.

On the other hand, the approach of query rewriting using views aims at finding view-based queries which are equivalent to (or contained in) a given user query. View-based query rewriting was first introduced using materialized views for query optimization purposes [6]. Afterwards, it was brought to the domain of data integration systems [2, 18] where the objective is to find certain answers to a query in a decentralized database context.

However, none of these approaches deals with the empty answer problem: they are concerned either in optimizing the access to information or in computing the set of certain answers to a query from distributed data sources. In our case, given a failing query, we are not interested in finding neither contained nor equivalent queries, since those would produce empty answers too.

In the information retrieval domain, other techniques are based on similarity measures and make use of previous executed queries to improve web search (see, e.g., [22, 23]). The measures underlying these approaches are strongly based on relationships between keywords, whereas we deal with a more general type of conditions than those expressed by a set of keywords, namely value constraints.

There are also relevant work in the domain of case-based reasoning, which integrates past user experience to solve current queries. In particular, Fouqué *et al.*[10] propose to include additional information in the answer to a user query, using an approach that they call associative query answering. The basic idea is to extend the "select" clause of a user query with attributes which appeared in similar queries previously submitted by other users. As we do, the authors use a nearness measure between every pair of values of each attribute, which is used to evaluate similarity between the user query and those previously executed. However, that work does not deal with failing queries.

In contrast, Bidault *et al.* [3] tackle the empty answer issue and use a repository of predefined queries in the context of mediation systems. Although their approach shows some similarity with ours, their solution is still quite different. They build a set of predefined successful queries for each source to offer a substitute for a failing user query. Similarity degrees between the initial user query and predefined successful queries are computed on the basis of a hierarchy of concepts. Finally, some heuristics make it possible to select the "best" substitute to the initial user query. In this approach, the similarity degree concerns the extent to which two concepts share

ISBN: 978-989-95079-6-8

the same characteristics whereas the substitutivity measure we defined is based on the proximity between the domain values of the attributes involved in the queries.

Let us also emphasize that none of the aforementioned approaches deals with failing fuzzy queries.

6 Conclusion

In this paper, we have outlined an approach aimed at obviating empty answers to Boolean or fuzzy queries involving value constraints. The method proposed uses a query repository and is based on the adaptation of the past non-failing query which is the most similar to the user query considered. Moreover, a technique aimed at providing the user with some explanations about the emptiness of the result of his/her original query has been briefly outlined.

The perspectives for future work are manifold. First, it would be useful to tackle the implementation of the query repository and to devise access methods for retrieving the candidate queries as efficiently as possible. The logical formalism proposed in [16] to represent information in cache and the index implementation presented in [10] to access it could be of interest for that purpose. Secondly, it would be worth investigating the possibility of reusing the execution plans of past queries to optimize the evaluation of the selected substitute query, in a spirit similar to [14]. Another point worthy to study concerns the substitutivity measure which is at the heart of the approach. The measure advocated here should be compared with some others adapted from classical similarity measures, cf. [9]. Finally, we intend to generalize this approach to a broader class of queries, e.g., those involving fuzzy joins.

References

[1] Y. Arens and C.A. Knoblock. Intelligent caching: selecting, representing and reusing data in an information server. In *Proc. of CIKM 1994*, pages 433–438, 1994.

[2] C. Beeri, A. Halevy, and M.-C. Rousset. Rewriting queries using views in description logics. In *Proc. of PODS'97*, pages 99–108, 1997.

[3] A. Bidault, C. Froidevaux, and B. Safar. Similarity between queries in a mediator. In *Proc. of ECAI 2002*, pages 235–239, 2002.

[4] P. Bosc, A. Hadjali, and O. Pivert. Weakening of fuzzy relational queries: an absolute proximity relation-based approach. *Mathware & Soft Computing*, 14(1):35–55, 2007.

[5] P. Bosc and O. Pivert. On the comparison of imprecise values in fuzzy databases. In *Proc. of the 6th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'97)*, pages 707–712, 1997.

[6] S. Chaudhuri, R. Krishnamurthy, S. Potamianos, and K. Shim. Optimizing queries with materialized views. In *Proc. of ICDE 1995*, pages 190–200, 1995.

[7] B. Chidlovskii and U.M. Borghoff. Semantic caching of web queries. *VLDB Journal*, 9:2–17, 2000.

[8] W. Chu, H. Yang, K. Chiang, M. Minock, G. Chow, and C. Larson. Cobase: A scalable and extensible cooperative information system. *Journal of Intelligent Information Systems*, 6(2-3):223–259, 1996.

[9] V.V. Cross and T.A. Sudkamp. *Similarity and Compatibility in Fuzzy Set Theory – Assessment and Applications*. Physica-Verlag, Heidelberg, 2002.

[10] G. Fouqué, W.W. Chu, and H. Yau. A case-based reasoning approach for associative query answering. In *Proc. of ISMIS'94*, pages 183–192, 1994.

[11] T. Gaasterland. Cooperative answering through controlled query relaxation. *IEEE Expert*, 12:48–59, 1997.

[12] T. Gaasterland, P. Godfrey, and J. Minker. An overview of cooperative answering. *Journal of Intelligent Information Systems*, 1(2):123–157, 1992.

[13] T. Gaasterland, P. Godfrey, and J. Minker. Relaxation as a platform for cooperative answering. *Journal of Intelligent Information Systems*, 1:293–321, 1992.

[14] A. Ghosh, J. Parikh, V.S. Sengar, and J.R. Haritsa. Plan selection based on query clustering. In *Proc. of VLDB 2002*, pages 179–190, 2002.

[15] P. Godfrey. Minimization in cooperative response to failing database queries. *Int. Journal of Cooperative Information Systems*, 6(2):95–149, 1997.

[16] P. Godfrey and J. Gryz. Answering queries by semantic caches. In *Proc. of DEXA 1999*, pages 485–498, 1999.

[17] S.Y. Huh, K.H. Moon, and H. Lee. A data abstraction approach for query relaxation. *Information and Software Technology*, 42:407–418, 2000.

[18] H. Jaudoin, J.-M. Petit, C. Rey, M. Schneider, and F. Toumani. Query rewriting using views in presence of value constraints. In *Proc. of the Int. Workshop on Description Logics (DL'05)*, pages 250–262, 2005.

[19] D.P. Miranker, M.C. Taylor, and A. Padmanaban. A tractable query cache by approximation. In *Proc. of SARA 2002*, pages 140–151, 2002.

[20] A. Motro. SEAVE: A mechanism for verifying user presuppositions in query systems. *ACM Trans. on Off. Inf. Syst.*, 4(4):312–330, 1986.

[21] I. Muslea. Machine learning for online query relaxation. In *Proc. of the Int. Conf. of Knowledge and Discovery and Data mining (KDD'04)*, pages 246–255, 2004.

[22] V. Raghavan and H. Sever. On the reuse of past optimal queries. In *Proc. of ACM SIGIR*, pages 344–350, 1995.

[23] J.R. Wen, J.-Y. Nie, and H.-J. Zhang. Query clustering using user logs. *ACM Transactions on Information Systems*, 20(1):59–81, 2002.

ISBN: 978-989-95079-6-8

A Fuzzy Delay Differential Equation Model for HIV Dynamics

Rosana M. Jafelice[1] Laécio C. Barros[2] Rodney C. Bassanezi[3]

1.Faculty of Mathematics, Federal University of Uberlândia
Uberlândia, Brazil
2.Department of Applied Mathematics, State University of Campinas
Campinas, Brazil
3. Center of Mathematics, Computation and Cognition, Federal University of ABC
Santo André, Brazil
Email: rmotta@ufu.br, laeciocb@ime.unicamp.br, rodney.bassanezi@ufabc.edu.br

Abstract— *This paper presents a process to obtain the solution (or flux) of a fuzzy delay system and to determine the fuzzy expected curve for the HIV (human immunodeficiency virus) when HIV-positive individuals receive antiretroviral therapy. This delay is defined as the time between the infection of a CD4+ type T-lymphocyte cell by the virus and the production of new virus particles. The intracellular delay is represented by an uncertainty parameter that depends on the individual characteristics of HIV-positive patients. A fuzzy rule-based system is used to model this parameter. The solution of the system of delay differential equations, which is a fuzzy process, is obtained from Zadeh's Extension Principle. Lastly, for each instant t, we calculate the fuzzy expected value obtained by the Center of Gravity.*

Keywords— Delay Differential Equation; Epidemiological Modeling; Fuzzy Expected Value; Fuzzy Set; HIV Dynamics; Mamdani Controller.

1 Introduction

HIV is an uncommon type of virus called a retrovirus and drugs developed to disrupt the action of HIV are known as antiretrovirals, or ARVs. The AIDS (Acquired Immunodeficiency Syndrome) virus mutates rapidly, which makes it extremely skillful at developing resistance to drugs. To minimize this risk, people with HIV are generally treated with a combination of ARVs that attack the virus on several fronts at once. The advent of ARVs in 1996 transformed the treatment of HIV and AIDS, improving the quality and greatly prolonging the lives of many infected people in places where the drugs are available. Of the estimated 6.5 million people in need of antiretroviral treatment in June 2006, 1.65 million people were reported to have had access to ARV treatment in low- and middle-income countries (World Health Organization, Jun 2006) [1]. In HIV-1 infection, treatment with reverse transcriptase or protease inhibitors results in a decline of free virus in several distinct phases. Fig. 1 summarizes the results of clinical studies, representing the different phases of viral decline after *in vivo* [2] treatment. Initially, the plasma viral load remains at approximately pretreatment levels, which, in the asymptomatic stage of the infection, are almost stationary on a time scale of weeks. This phase, which lasts from 0 to 1 day, is called the quasi-stationary initial stage. This initial phase is followed by a transition phase of 1 or 2 days, which is explained by the combination of pharmacological and intracellular delay effects that cause the disappearance of the free particles of the virus and the decline of infected cells. Another phase observed is the rapid decline of the plasma virus, which lasts from 2 to 7 days. Finally, the decline flattens out and virus levels may even rise again [2].

In this paper, we propose a process to determine the fuzzy flux of HIV decline as a function of time and, for each t, we determine the fuzzy expected value when the HIV-positive individual is receives antiretroviral therapy, considering the intracellular delay of the viral life cycle. Intracellular delay is defined as the time between the infection of a T lymphocyte cell of the $CD4+$ type by the virus and the production of new virus particles. The main lymphocyte that HIV attaches to when it reaches the bloodstream is the $CD4+$ type T lymphocyte. Based on the solution of a delay differential equation system [2], considering that the antiretroviral therapy is 100% effective, we introduced the delay τ as a fuzzy triangular number, where the degree of pertinence 1 is assumed to be $\tau = 12$h. Starting from a fuzzy rule-based system (FRBS), we determined the mortality rate of the virus as a function of delay. The fuzzy rule base was obtained from information given in [2]. With this fuzzy parameter [3], and by means of Zadeh's extension principle, we obtained a solution for the system. With this methodology applied to each instant t, the solution was a fuzzy set. Hence, we found the degree of pertinence for each solution of the deterministic system as a function of time. We then calculated the fuzzy expected value for the solutions of the system at each instant t, using the center of gravity. The curve of the expected value shows a profile similar to the curves of the clinical data when antiretroviral therapy is applied, as illustrated in Fig. 1.

2 Classic Models

In [2] was introduced two microscopic models for HIV infection dynamics in the individual organism undergoing antiretroviral therapy. These models are used in this paper.

The first is a model for viral dynamics with three variables: the population of uninfected cells, $x(t)$; the population of infected cells that produce the virus, $y(t)$, and the plasma virus population, $v(t)$. In a first approximation of the dynamics, we assumed that the influx λ and death rate d of uninfected cells are constant. Uninfected cells and free virus produce infected cells at a rate of $\beta(t)x(t)v(t)$. Infected cells produce free virus particles at a rate of $k(t)$ and die at a rate of a. Free virus

ISBN: 978-989-95079-6-8

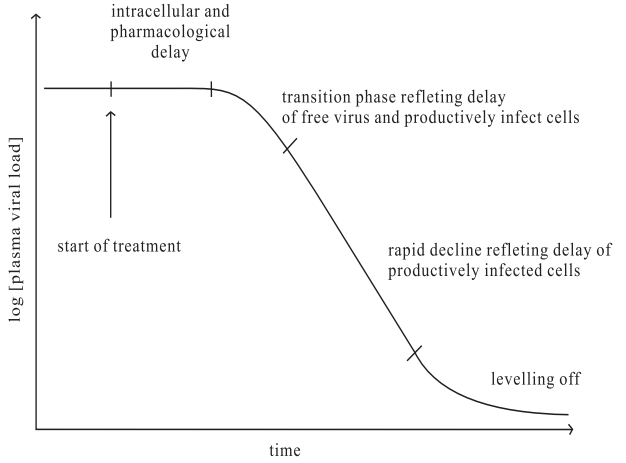


Figure 1: Schematic illustration of the different phases of plasma virus decline following antiretroviral therapy [2].

particles are cleared at a rate of u. To describe the effects of various drug therapies, the parameters $\beta(t)$ and $k(t)$ are time-dependent. Based on these assumptions, the differential equation system is given by:

$$\frac{dx(t)}{dt} = \lambda - dx(t) - \beta(t)x(t)v(t) \quad (1)$$

$$\frac{dy(t)}{dt} = \beta(t)x(t)v(t) - ay(t) \quad (2)$$

$$\frac{dv(t)}{dt} = k(t)y(t) - uv(t) \quad (3)$$

The model does not contain an intracellular time delay between infection of a cell and production of new virus particles. The model (4)-(6) incorporates the intracellular phase of the virus life cycle. In [2], it was assumed that virus production lags by a delay τ behind the infection of a cell. This implies that the recruitment of virus-producing cells, at time t, is given by the density of cells that were newly infected at time $t-\tau$ and are still alive at time t. Moreover, we assumed a constant death rate $\tilde{a}$ for infected but not yet virus-producing cells. The probability of survival from $t-\tau$ to time t is only $e^{-\tilde{a}\tau}$. More generally, the probability of survival is given by a nonincreasing function $f(\tau)$ with $0 \le f(\tau) \le 1$. Thus, (1), (2) and (3) can be written as:

$$\frac{dx(t)}{dt} = \lambda - dx(t) - \beta(t)x(t)v(t) \quad (4)$$

$$\frac{dy(t)}{dt} = \beta(t-\tau)x(t-\tau)v(t-\tau)e^{-\tilde{a}\tau} - ay(t) \quad (5)$$

$$\frac{dv(t)}{dt} = k(t)y(t) - uv(t) \quad (6)$$

The equation (5) has become a 'delay-differential equation τ'. Analytical solutions of this type of equation are generally difficult to find. However, for the problem in question, the populations of uninfected cells, infected virus-producing cells and free virus are at a steady-state level before treatment sets in [2]. This situation facilitates the mathematical analysis and enables us to derive simple analytical solutions. The nontrivial steady state of the system is given by:

$$x_0 = \frac{au}{\beta k}e^{\tilde{a}\tau}$$

$$y_0 = \frac{\lambda}{a}e^{\tilde{a}\tau} - \frac{du}{\beta k}$$

$$v_0 = \frac{ky_0}{u} \quad (7)$$

where β and k are constant pretreatment rates.

2.1 Protease Inhibitor Therapy

Reference [2] includes treatment with HIV protease inhibitors that block the production of new infectious virus v_I from already infected cells, allowing only noninfectious virus to be generated. As previously reported, the infectious virus declines but continues to infect cells [4] and [5].

According to [2], within the present framework, equation (6) also describes the dynamics of the total free virus. Infectious virus v_I, however, is not produced in $t > 0$ and declines according to $\frac{dv_I(t)}{dt} = -uv_I(t)$. Also, equations (4) and (5) remain valid if one replaces v with v_I.

In [2], it is assumed that the uninfected cell population remains constant, $x(t) = x_0$ within the time scale under consideration. For $x(t) = x_0$ and exponentially declining $v_I(t)$, equation (5) is solved by:

$$y(t) = \frac{y_0}{a-u}[ae^{-u(t-\tau)} - ue^{-a(t-\tau)}] \quad for \quad t > \tau. \quad (8)$$

From (6), the time evolution of free virus is then given by $v(t) = v_0$ for $0 < t \le \tau$ and

$$v(t) = v_0 e^{-u(t-\tau)} + \frac{uv_0}{a-u}\left\{\frac{u}{a-u}\left[e^{-a(t-\tau)} - e^{-u(t-\tau)}\right]\right\}$$

$$+\frac{uv_0}{a-u}\left\{a(t-\tau)e^{-u(t-\tau)}\right\} \quad for \quad t > \tau. \quad (9)$$

Combined reverse transcriptase inhibitor and protease inhibitor therapy may, in the long term, delay the evolution of drug-resistant virus strains [2].

After application of any antiviral drug, there is a short delay in the pharmacological effect due to the time required for drug absorption, distribution, and penetration into target cells. In the initial phase, the plasma virus load remains constant. The duration of this phase is the sum of the pharmacological delay [2], defined as the time span the drug requires to reach an effective concentration, and the intracellular delay, defined as the time between infection of a cell and production of new virus particles.

Thus, differences in plasma virus decline for a fixed half-life of infected cells ($a = 0.5$/day), with a delay of about 0.9 days, are shown in Fig. 2 for protease inhibitors. Intracellular delays were chosen to ensure that the long-term decline is identical in all cases. The curves differ mostly at the end of the shoulder phase, but even for strongly varying parameters, the differences are minor, see Fig. 3.

It is important to note that for different values of intracellular delay (τ) and rate of decline in the plasma virus concentration (u), the behavior of the plasma viral load (Fig. 3) is similar to that of the initial phases of antiretroviral treatment,

ISBN: 978-989-95079-6-8

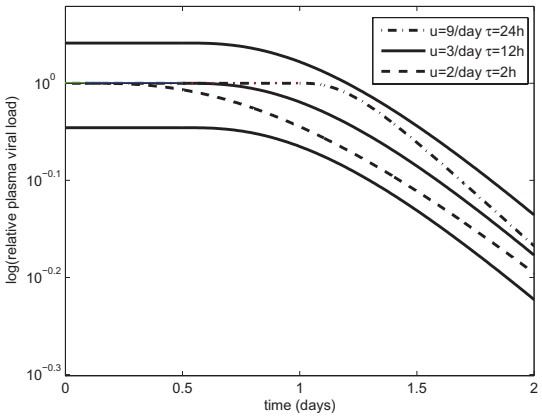


Figure 2: Sensitivity parameter of the protease inhibitor equation (9). Parameter a is fixed ($a = 0.5/day$), whereas u and τ vary. The two thin lines represent 10% deviations from the mean solution to illustrate expected measurement errors under ideal experimental conditions [2].

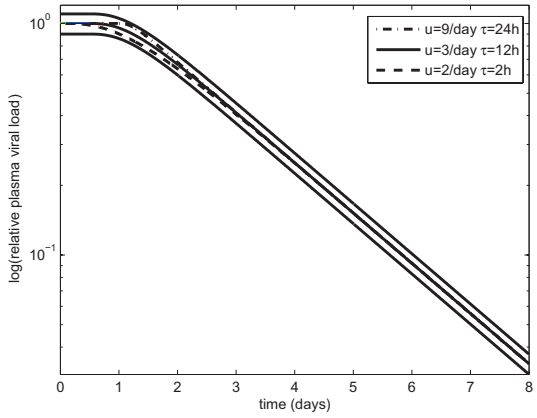


Figure 3: Range of HIV decline in response to protease inhibitor therapy [2].

as illustrated in Fig. 1, which summarizes the clinical studies, rendering the model credible (4)-(6). Reference [6] states that the length of delays is not directly measurable in vivo. Thus, additional in vitro data would prove valuable in expanding our knowledge of the characteristics of delay. Although these in vitro data are fraught with uncertainties, they clearly indicate that there is a delay between the initial infection and the production of new virus [7].

Reference [6] also states the need to explore the effects of different types of delay, such as distributed delays. Reference [7] assumes that the delay is given by a probability distribution. We propose a process to study the decline of the viral load after the beginning of protease inhibitor therapy. To this end, we consider that the intracellular and pharmacological delay is modeled by a fuzzy number, which takes into account a possibility distribution that can be given by a specialist.

The next section introduces the fuzzy model, which is the main object of this paper.

3 Fuzzy Model

Mathematical models for biological phenomena are normally fraught with uncertainties, in both the state variables and the parameters of the model's equations [8]. In our specific case, for reasons set forth in the introduction of this paper, we will consider that the delay time (τ) is uncertain and is modeled by the fuzzy set theory. Therefore, to study HIV decline over time, we developed a model which is a combination of classic delay differential equations and fuzzy logic.

A mathematical model in [4] was developed for protease inhibitor therapy, which fitted theoretically predicted curves of plasma virus decline to empirical data. The authors determined an average of about 0.9 days for the intracellular phase.

The next subsection presents a summary of the fuzzy set theory that was adopted in this work.

3.1 Fuzzy Sets

A *fuzzy subset* F of the universe set $\mathcal{U}$ is defined in terms of a function of *pertinence* u , which, at each element x of $\mathcal{U}$, associates a number $u(x)$ from *zero* and *one* called degree of pertinence of x to F. Thus, the fuzzy set F is symbolically indicated by its pertinence function

$$u_F : \mathcal{U} \rightarrow [0,1].$$

The values $u_F(x) = 1$ and $u_F(x) = 0$ indicate, respectively, the complete pertinence and the nonpertinence of element x to F.

A concept that will play a key role in this paper is fuzzy rule-based systems (FRBS) that have four components: an input processor, a collection of fuzzy rules called rule base, a fuzzy inference machine, and an output processor [9]. As the fuzzy inference machine we adopted the Mamdani Inference Method, while the output processor we used was the Center of Gravity.

Another concept used in this paper is Zadeh's Extension Principle.

Let X and Y be sets and f na application of X in Y. Let Abe a fuzzy set in X. The extension principle states that the image of A by the function f is a fuzzy set $B = f(A)$ in Y, whose pertinence function is given by

$$u_B(y) = \sup_x u_A(x) \tag{10}$$

for each $y \in Y$ with $x \in X$ and $y = f(x)$.

3.2 Methodology

Using an FBRS with the Mamdani Inference Method and center of gravity defuzzification, we obtained the defuzzified values of u as a function of τ. The least-square regression yielded $u(\tau)$. To represent the distributed delay, we adopted τ as a fuzzy triangular number. We then applied Zadeh's extension principle in two stages: first, to obtain the fuzzy u, and then to fuzzify the vital load. Finally, for each instant t, we calculated the fuzzy expected value by means of the Center of Gravity.

3.3 Fuzzy Rules

To determine the relationship between u and τ ($u = u(\tau)$), we adopted a FBRS, taking advantage of the fact that these variables are positively correlated [2].

ISBN: 978-989-95079-6-8

The intracellular delay (τ) and rate of decline in plasma virus concentration (u) are linguistic variables: the values of intracellular delay are expressed as $\{$*low*, *medium*, *high medium*, *high*, and *very high*$\}$, while the rate of decline in plasma virus concentration is expressed in the term set $\{$*low*, *medium*, *high medium*, *high*, *very high* $\}$.

Based on the values

- $u = 9/\text{day}$ and τ=0.08 day (approximately 2h);
- $u = 3/\text{day}$ and τ=0.5 day (12h);
- $u = 2/\text{day}$ and τ=1 day (24 h).

employed in the studies of [2], we adopted the domains of the pertinence functions of the intracellular delay (τ) and of the rate of decline in plasma virus concentration (u) as intervals that contain these values.

Following the positive correlation, the rule base that encodes the correlation between τ and u is given by:

- If τ is *low* then u is *low*.
- If τ is *medium* then u is *medium*.
- If τ is *high medium* then u is *high medium*.
- If τ is *high* then u is *high*.
- If τ is *very high* then u is *very high*.

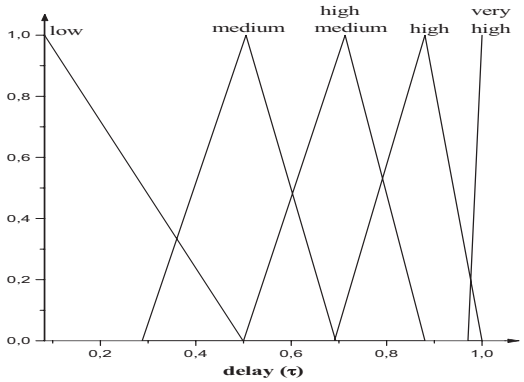


Figure 4: Membership functions for delay (τ).

The membership functions for each fuzzy set that specify the meaning of the linguistic variables for intracellular delay (τ) and rate of decline in plasma virus concentration (u) are shown in Fig. 4 and 5, respectively.

The data on virus mortality rate as a function of delay presented in Fig. 6 were obtained from the defuzzified values in the $[0.08, 1]$ interval of the FRBS. Using least-squares regression, we obtained the curve of the mortality rate of the virus as a function of delay, which is given by

$$u(\tau) = 1.67e^{1.63\tau} \tag{11}$$

with a determination coefficient (R^2) of 0.9873, as indicated in Fig. 6.

For biologic reasons, Mittler et al. [7] used a probabi-

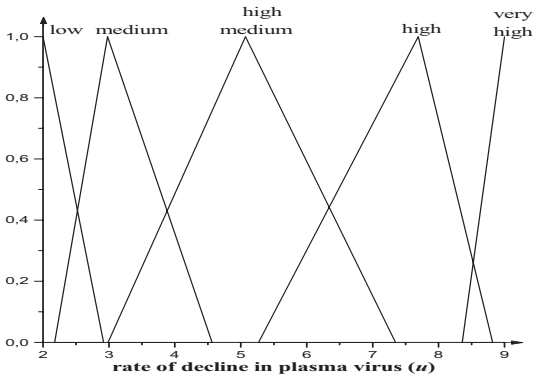


Figure 5: Membership functions for rate of decline in plasma virus concentration (u).

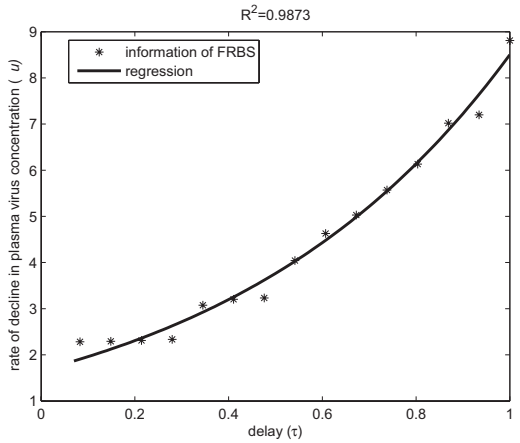


Figure 6: Regression of the rate of decline in plasma virus concentration

lity distribution function for τ. Here, in this paper, we will consider a possibility distribution instead a probability distribution function for τ. The rate of decline (u) will be obtained from a fuzzy rule-based system combined with the extension principle. Another method could be used for obtaining u. For example, the fuzzy systems given by gradual rules [10]. However, in this last case, we are not sure if these systems, in fact, approximate real functions.

3.4 Fuzzy Delay Distribution

We considered τ as a fuzzy parameter given by the fuzzy triangular number, whose support is [0.08,1] and maximum pertinence in 0.5, Fig.7.

The extension principle was used to obtain the image u of the fuzzy set τ, through function $u(\tau) = 1.67e^{1.63\tau}$. More specifically, from the formula:

$$\mu_{u(\tau)}(y) = \sup_{\{s:u(s)=y\}} \mu_{\tau}(s) \tag{12}$$

where τ is the fuzzy set whose membership function is u_τ, and $u(\tau)$ is the corresponding fuzzy set with membership function $\mu_{u(\tau)}$. Fig. 8 shows the extension principle for $u(\tau)$, assuming

ISBN: 978-989-95079-6-8

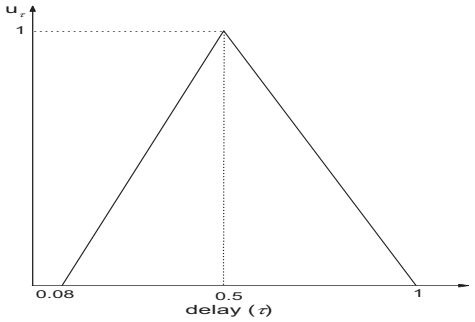


Figure 7: Fuzzy parameter (τ).

μ_τ, as in Fig. 7.

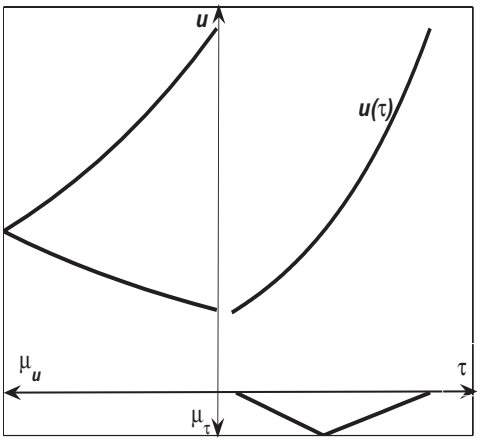


Figure 8: Computation μ_u.

3.5 *Fuzzy Solution*

The fuzzy solution for the viral load was obtained by fuzzification of the deterministic solution given by:
$v(t;\tau) = v_0$ for $0 < t \leq \tau$ and

$$v(t;\tau) = v_0 e^{-u(\tau)(t-\tau)} + \frac{u(\tau)v_0}{a-u(\tau)}\left\{\frac{u(\tau)}{a-u(\tau)}\left[e^{-a(t-\tau)} - e^{-u(\tau)(t-\tau)}\right]\right\}$$

$$+\frac{u(\tau)v_0}{a-u(\tau)}\left\{a(t-\tau)e^{-u(\tau)(t-\tau)}\right\} \quad for \quad t > \tau \qquad (13)$$

since we are admitting τ, and hence, u, as a fuzzy number.

At each instant t, Zadeh's Extension Principle is adopted for the function $v(t;\tau)$ applied in τ.

Fig. 9 presents the graphic solution of the fuzzy model. The light region shows the curves with the highest possibility of representing the phenomenon.

3.6 *Defuzzification of the Fuzzy Solution*

A common defuzzification scheme is the center of gravity method. For a fixed t, denote $v(t;\tau)$ by $v_t(\tau)$ to simplify the notation. Let $\mu_{v_t(\tau)}$ be the membership function of $v_t(\tau)$. Thus,

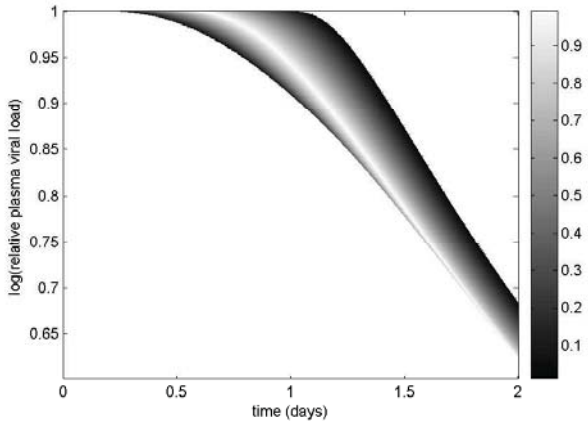


Figure 9: Solution of the fuzzy viral load for t varying from 0 to 2 days

a real-valued output $\bar{v}(t)$ is obtained at each time instant t, according to the formula:

$$\bar{v}(t) = \frac{\int\limits_{supp(v_t(\tau))} v_t(\tau)\mu_{v_t(\tau)}(v_t(\tau))dv_t(\tau)}{\int\limits_{supp(v_t(\tau))} \mu_{v_t(\tau)}(v_t(\tau))dv_t(\tau)} \qquad (14)$$

Given the fuzzy solution shown in Fig.9 and using the center of gravity, we obtained the defuzzified solution depicted in Fig.10.

The method generates a fuzzy expected curve which we have called a (defuzzified) solution $\bar{v}(t)$.

Reference [11] showed that this curve is the solution of the nonautonomous differential equation and may be viewed as the expected value of $v_t(\tau)$.

The solution $v(t;\tau)$ is obtained from a family of classic differential equations (Fig.11). However, it does not coincide with any solution for a fixed τ. For each t, $v(t;\tau)$ (14) is a value that belongs to the solution of the family of equations parameterized by $u(\tau)$. What differs here from the derived curve (14) suggested by the deterministic solutions is that, in deterministic solutions, all the uncertainties are excluded at the beginning (defuzzified at $t = 0$ and solved), whereas here the uncertainties evolve and defuzzification occurs at the time instant of interest (defuzzified when necessary) [11] and [12]. Hullermeier [13] suggests that the solution of a fuzzy differential equation is a fuzzy function obtained from a family of differential inclusions. Mizukoshi et al [14] have shown that when parameters (coefficients and/or initial conditions) are fuzzy, the Hullermeier solution is the same as that obtained from the extension principle. Therefore, we first obtained the solution of the classical differential equation and then used the extension principle to obtain fuzzy solutions. To obtain a real-valued trajectory, we adopted the center of gravity to defuzzify the fuzzy solution.

By means of equation (14), we obtained the fuzzy expected value $\bar{v}(t)$ at each instant t, which at first intercepts curves that present greater delays and, over time, intercepts curves with smaller delays, Fig. 11.

ISBN: 978-989-95079-6-8

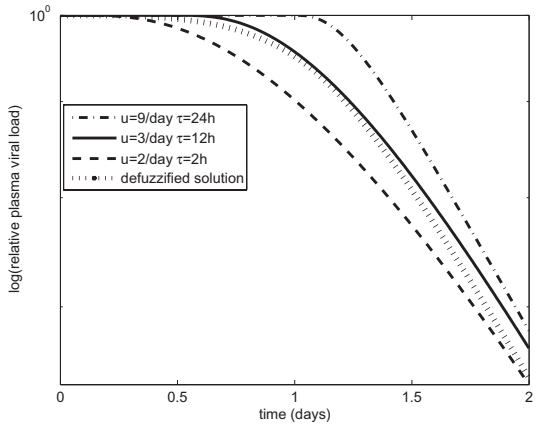


Figure 10: Defuzzified solution at each time instant t

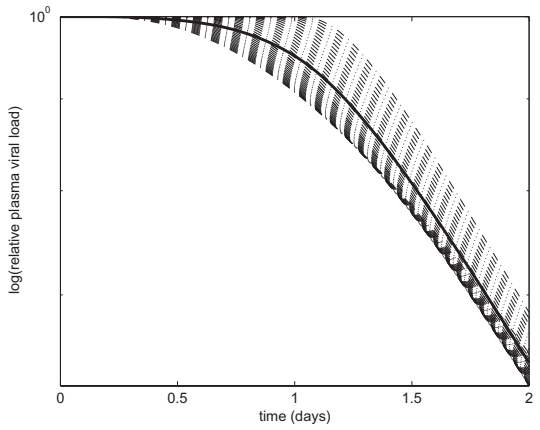


Figure 11: In the graph, the solid line represents the fuzzy expected value $\overline{v}(t)$.

4 Conclusions

In this work, emphasis was given to the study of the initial phase of treatment, since the plasma concentration of HIV declined by approximately 90% in the first two weeks of therapy due to the rapid elimination of free virus and the decline in infected cell productivity [15], see Fig. 10.

The work of [2] shows that intracellular delay may affect the magnitude of the range observed in Fig. 3, which represents variability or uncertainty. The adoption of distributed delay has yielded good results in the study of the dynamics of HIV under treatment [7]. In this paper, we adopted a fuzzy number to represent the distributed delay, which enabled us to obtain curves with different degrees of possibility to represent the phenomenon, Fig. 9. The degrees of pertinence of the curves come close to 1, as the curves approach the light region. This region is the one that best represents the phenomenon from the standpoint of credibility.

The defuzzified curve $\overline{v}(t)$ represents the amount of virus at each instant t. This curve is an average of the deterministic solutions given in (13), weighted by their respective degrees of possibility, which renders the curve more representative than each of the solutions of (13), adopting a value for τ.

Acknowledgements

The first author gratefully acknowledges CNPq, the Brazilian National Research Council, for granting a postdoctoral fellowship (152068/2007-4). The second author thanks FAPESP, the Research Foundation of the State of São Paulo, for grant no. 06/05920-7, and CNPq for grant no. 307890/2006-6.

References

[1] Joint United Nations Programme on HIV/AIDS. *www.unaids.org*, Acessed: June 18 2008.

[2] A.V.M. Herz, S. Bonhoeffer, R.M. Anderson, R.M. May, and M.A. Nowak. Viray dynamics *in vivo*: Limitations on estimates of intracellular delay and virus decay. In *Proc. Natl. Acad. Sci (Medical Sciences)*, pages 7247–7251, USA, 1996.

[3] R.M. Jafelice, L.C. Barros, R.C. Bassanezi, and F. Gomide. Fuzzy modeling in asymptomatic HIV virus infected population. *Bulletin of Mathematical Biology*, 66(6):1463–1942, 2004.

[4] A.S. Perelson, A.U. Neummann, M. Markowitz, J.M. Leonard, and D.D. Ho. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span and viral generation time. *Science*, 271:1582–1586, 1996.

[5] D.L. Winslow and M.J. Otto. HIV protease inhitors. *AIDS 9 (Suppl.A)*, pages S183–S192, 1995.

[6] P.W. Nelson, J.D. Murray, and A.S. Perelson. A model of HIV-1 pathogenesis that includes an intracellular delay. *Mathematical Biosciences*, 163:201–215, 2000.

[7] J.E. Mittler, B. Sulzer, A.U. Neumann, and A.S. Perelson. Influence of delayed viral production on viral dynamics in HIV-1 infected patients. *Mathematical Biosciences*, 152:143–163, 1998.

[8] E. Massad, N.R.S. Ortega, L.C. Barros, and C.J. Struchiner. *Fuzzy Logic in Action: Applications in Epidemiology and Beyond*. Springer-Verlag Berlin, 2008.

[9] W. Pedrycz and F. Gomide. *An Introduction to Fuzzy Sets: Analysis and Design*. MIT Press, Cambridge, EUA, 1998.

[10] B. Bouchon-Meunier, D. Dubois, L. Godo, and H. Prade. *Fuzzy Sets and Possibility Theory in Approximate and Plausible Reasoning. Fuzzy Sets in Approximate Reasoning and Information Systems*. The Handbooks of Fuzzy Sets. Kluwer Academic Publishers, Kluwer, Boston, Mass., 1999.

[11] R.M. Jafelice, L.C. Barros, R.C. Bassanezi, and F. Gomide. Methodology to determine the evolution of asymptomatic HIV population using fuzzy set theory.

[12] L.C. Barros, R.C. Bassanezi, and M.B. Leite. The epidemiological models SI with fuzzy parameter of transmission. *Computers and Mathematics with Applications*, 45:1619–1628, 2003.

[13] E. Hüllermeier. An approach to modelling and simulation of uncertain dynamical systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5(2):117–137, 1997.

[14] M.T. Mizukoshi, L.C. Barros, Y. Chalco-Cano, H. Román-Flores, and R.C. Bassanezi. Fuzzy differential equation and the extension principle. *Information Sciences*, 177:3627–3635, 2007.

[15] A.S. Perelson, P. Essunger, Y. Cao, M. Vesanen, A. Hurley, K. Saksela, M. Markowitz, and D.D. Ho. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*, 387:188–191, 1997.

ISBN: 978-989-95079-6-8

Dispersal of Leaf-Cutting Ants: Fuzzy Mathematical Modeling, Numerical Approximation and Simulations

Rosana M. Jafelice, César G. Almeida[1] J.F.C.A. Meyer[2] Heraldo L. Vasconcelos[3]

1.Faculty of Mathematics, Federal University of Uberlândia
Uberlândia, Brazil
2.Department of Applied Mathematics, State University of Campinas
Campinas, Brazil
3.Institute of Biology, Federal University of Uberlândia
Uberlândia , Brazil
Email: {rmotta,cesargui}@ufu.br, joni@ime.unicamp.br, heraldo@umuarama.ufu.br

Abstract— *The objective of this paper is to propose a model for the occupation, by leaf-cutting ants, of an attractive area using a form of the diffusion-advection partial differential equation, in which the population dispersal is modeled using a fuzzy parameter. In the studied domain there is an attractive region, which can represent areas with an elevated concentration of palatable host-plants with better nutritious qualities. The developed algorithm uses information about the foraging behavior of a leaf-cutting ant colony from the Amazon region, in northern Brazil. The solution of the chosen partial differential equation is approximated using first order finite triangular elements and determines population dispersal with the use of a fuzzy rule-based system, which depends upon the number of individuals in the population and terrain characteristics. The numerical algorithm was developed in a MATLAB environment. The calculated solution is compatible with ant behavior when there is an attraction region nearby, as reported in the literature.*

Keywords— Fuzzy Set; Leaf-Cutting Ants; Mathematical Modeling; Numerical Methods; Partial Differential Equation;

1 Introduction

Ants of the genera *Atta* and *Acromyrmex* (*Hymenoptera: Formicidae*), collectively known as leaf-cutting ants, have the unique habit of culturing fungus on fresh plant materials [1]. These ants are generalist foragers that exploit a large number of plant species, although they usually focus on a subset of these species [2], especially those with low levels of toxic secondary compounds and high nutrient content [3, 4]. Their system of foraging trails influence the spatial organization of the foraging activity. Spatial and temporal heterogeneity in plant resources within the vicinity of the ants' nest produces changes in trail direction and modifications in the geometry of the foraging territories [5, 6]. Trails often lead to sites where, at a given time, plant resources are more attractive and abundant. For instance, a study in Costa Rica showed that, when leaf-cutting ant colonies have access to different plant communities, they concentrate their foraging efforts on the community with the greatest density of their preferred plant species (a monoculture of cassava) [7].

This paper proposes a model for the occupation of an attractive area by leaf-cutting ants, using a form of the diffusion-advection partial differential equation to model population dispersal [8] using a fuzzy parameter. The mathematical model for population dispersal is based on the balance of mass in phenomena of fluid dynamics and their physical principles. In this case, Fick's law; a conservation equation - such as conservation of mass -; and the notion of advection and diffusive flux arising from directed motion (convection) and effective random motion, respectively, are fundamental for the development and understanding of the model. In this work, the dispersion depends on the number of individuals of the population and on the characteristics of the terrain over which ants travel. This dispersion is determined by means of a system based on fuzzy rules. The model is constructed using expert entomological knowledge on behavior of leaf-cutting ants. The domain of this study contains an attractive region, which can represent areas with a high concentration of palatable host plants with good nutritious qualities. In literature, this research usually involve the use of a deterministic or stochastic model. However, mathematical literature on uncertainty has grown considerably over the last decade, especially in system modeling, optimization, control, and pattern recognition areas, to mention just a few. Recently, several authors have proposed the use of fuzzy set theory in epidemiology problems [9, 10, 11, 12] and population dynamics [13]. In this paper, we suggest fuzzy set theory, introduced in the 1960s by Lotfi Zadeh [14], to deal with the uncertain nature of ant population dispersal. In [15], the authors propose fuzzy rules that have input variables given by the difference (D) in the pheromone concentrations of the ants on the left and right trail branches. Based on these rules, the authors are able to determine the probability of choosing the left branch. The behavior of ants has also inspired several optimization techniques, the most exhaustively studied of which is known as ant colony optimization [16], which is also based on ant pheromone. The present study is based on the fact that the availability of plant resources to leaf-cutting ant colonies varies considerably in space and time. Thus, colonies must continually gather information about the current availability or stock of resources within their territories. Once a patch of palatable vegetation (i.e., an attractive area) is found, a chemically and physically marked trail is established and the ant foragers follow this trail in order to retrieve the newly found resource.

The next section presents information about the area where the research on ants was conducted [17].

ISBN: 978-989-95079-6-8

2 Ant Study Area in Amazonas

This study was carried out in a forest reserve located about 70 kilometers north of Manaus, state of Amazonas, Brazil. The vegetation is that of a primary inland rain forest, which, unlike other types of forest in the Amazon basin, is not subject to annual flooding. The climate is tropical and the annual precipitation, which is about 2,100 mm, is seasonal in distribution [18]. The rainy season lasts from November to May, and the dry season from June to October.

Vasconcelos's research [17] involved a study of the foraging activity of an *Atta cephalotes* colony from July 1985 to January 1986 and from September 1986 to March 1987. This study determined the spatial distribution of ant foraging trails and of the plant species exploited by the ants (i.e., their foraging sites).

Fig. 1 shows that the foraging sites of the *Atta cephalotes* were scattered over the entire extent of the trails, though mainly at intermediate distances from the nest (40 - 60m).

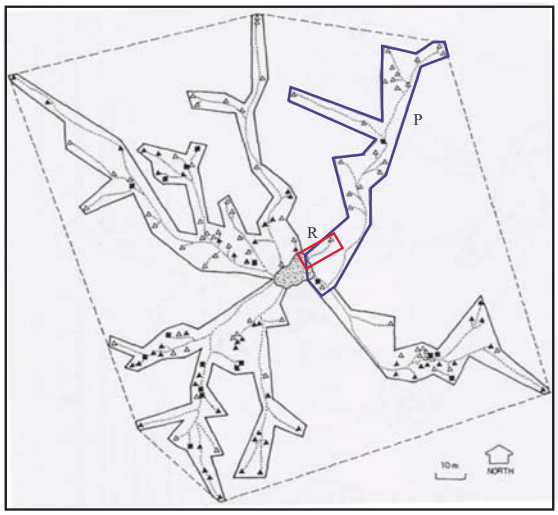


Figure 1: Foraging range (broken line) and foraging territory (continuous line) of the *Atta cephalotes* colony. Solid triangles represent the plants attacked between July 1985 and January 1986, and empty triangles those attacked between September 1986 and March 1987. Solid squares represent plants attacked during both periods. Dotted lines represent the foraging trails radiating from the nest (shaded area) [17].

The next section presents a partial differential equation with a fuzzy parameter for the occupation of leaf-cutting ants.

3 Fuzzy Model

The study of population dynamics involves several uncertain variables, such as the number of individuals in the population, the landscape, the speed at which the population travels, and population dispersal. Fuzzy set theory is a mathematical tool for modeling uncertain phenomena [19], for example, provides key notions for modeling epidemiological phenomena. Thus, the aim of this research is to find a numerical solution for modeling the presence of ant populations, treating dispersal as a fuzzy parameter [20].

The model proposed for the occupation by leaf-cutting ants will be studied by means of the diffusion-advection partial differential equation, given by:

$$\frac{\partial P}{\partial t} + v \cdot \nabla P - \nabla \cdot (\alpha(P, loc_{tot}) \nabla P) = 0. \qquad (1)$$

The functional variable $P = P(\mathbf{x}, t)$ indicates the population at the instant $t \in [0, T]$ and at the point $\mathbf{x} \in \Omega_1 \subset \mathbf{R}^2$ (see Fig. 2). The speed of occupation of the region v is a constant vector (v_1, v_2). Thus, we assume that the population dispersal is represented by parameter $\alpha(P, loc_{tot})$, where total locomotion (loc_{tot}) is determined by a fuzzy rule-based system that depends on the input variables loc_x (horizontal movement) and loc_y (vertical movement). This dispersal depends on the subregions of the domain, indicating the difficulty of locomotion of the ants over the terrain. The fuzzy parameters are modeled as follows:

- the domain is divided into subregions;
- the characteristics of the environment in which the individuals travel are represented by trapezoidal membership functions that indicate the degree of difficulty in locomotion through the domain;

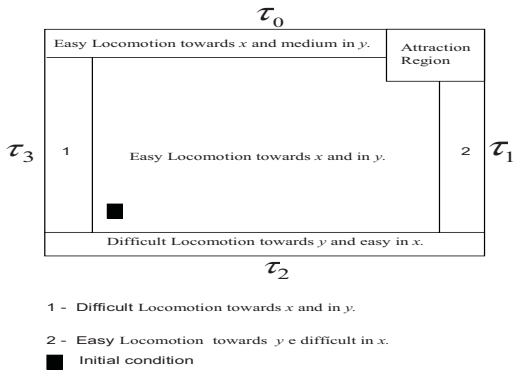


Figure 2: Domain Ω_1.

The adopted boundary conditions considered are:

$$\alpha(P, loc_{tot}) \frac{\partial P}{\partial y} = P v_2, \qquad \forall \mathbf{x} \in \tau_0, \qquad \forall t \in [0, T] \quad (2)$$

$$\alpha(P, loc_{tot}) \frac{\partial P}{\partial x} = P v_1, \qquad \forall \mathbf{x} \in \tau_1, \qquad \forall t \in [0, T] \quad (3)$$

$$\alpha(P, loc_{tot}) \frac{\partial P}{\partial y} = P v_2, \qquad \forall \mathbf{x} \in \tau_2, \qquad \forall t \in [0, T] \quad (4)$$

$$\alpha(P, loc_{tot}) \frac{\partial P}{\partial x} = P v_1, \qquad \forall \mathbf{x} \in \tau_3, \qquad \forall t \in [0, T] \quad (5)$$

The initial condition for this problem is given by:

$$P(\mathbf{x}, 0) = P_0(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega_1. \qquad (6)$$

To determine the numerical solution for (1), (2)-(5), the finite element method was used with a triangular uniform grid and linear functions [8]. The Crank-Nicolson Method was employed for the time discretization.

ISBN: 978-989-95079-6-8

3.1 Boundary Conditions

The spatial domain of the partial differential equation is rectangular: $\Omega_1 = [0, X] \times [0, Y]$, where X is the longitudinal length (axis x) and Y is the transversal length (axis y) of domain Ω_1. The domain boundary, $\partial\Omega_1$, is divided into parts which are indicated by: τ_0: the upper edge; τ_1: the right edge; τ_2: the lower edge and τ_3: the left edge. The unitary vector with an external direction normal to $\partial\Omega_1$ is given by: $n = (0, 1)$, in τ_0; in τ_1, $n = (1, 0)$, at the right edge; in τ_2, $n = (0, -1)$, at the lower edge; and $n = (-1, 0)$, in τ_3. In order to preserve the population mass as a function of time $M(t)$, the proper boundary conditions must be identified [**?**, 21]. Initially, the equation (1) is written as follows:

$$\frac{\partial P}{\partial t} + div(vP) - div(\alpha(P, loc_{tot})\nabla P) = 0. \quad (7)$$

Integrating equation (7) throughout the domain Ω_1, with $\int_{\Omega_1} Pd\mathbf{x} = \mathbf{M(t)}$ and using the Divergence Theorem, one has:

$$\frac{d}{dt}M(t) + \int_{\partial\Omega_1} (Pv \cdot n - \alpha(P, loc_{tot})\nabla P \cdot n)dS = 0 \quad (8)$$

To ensure that $M(t)$ is constant, one must have: $\frac{d}{dt}M(t) = 0$. Therefore:

$$(Pv \cdot n - \alpha(P, loc_{tot})\nabla P \cdot n) = 0. \quad (9)$$

Thus, one obtains the boundary conditions for the Ω_1, from (2) to (5).

3.2 Parameters for Computational Simulation

Fig. 2 represents domain Ω_1. To find the solution for equation (1) through first-order finite elements, this domain was subdivided into small triangles. The discrete solution will be obtained at the vertices (or nodes) of these triangles. The objective of computational simulations was to describe region R, given in Fig. 1. The solution for equation (1) was obtained from the finite triangular elements, using 3600 elements with 1860 nodes. The region of attraction has eight triangles and is located in the upper right-hand side of domain Ω_1, Fig. 2. The *A. cephalotes* colony is represented by four initial nodes close to boundaries τ_2 and τ_3, where an initial population of 400 ants was placed, Fig. 10. At each time step, two individuals are added to the calculated population, simulating ants present at the entrance of the ant nest at all times.

To calculate the dispersal several calculations were made based on Fig. 1. Initially, we measured the dotted trails in region P (large polygon) of Fig. 1. These trails were about 173 meters (m) in length, and the ants took approximately 40 days to occupy this area [17]. The dotted trail in region R (small rectangle) of Fig. 1 was $12m$ long. The rectangle of region R is approximately $13m$ by $5m$ and, to increase the performance of the computational program, let us consider that domain Ω_1 is $6m$ by $3m$. Because the computational program was implemented in a region whose sides measured half of the original region R, we considered that the trail was also half as long as the original one, i.e., $6m$. Therefore, we determined that the ants took 1.3873 days to occupy the $6m$ trail. According to the entomologist, ant trails have a width of approximately $0.15m$. Thus, the area occupied by the ants in the region for the implementation is approximately $4.7971m^2$. Since the dispersion (α) is determined by the area/time ratio, one has: $\alpha \cong 0.14m^2$/hour. We considered the speed in direction y to be equal to $0.07m$/hour. The speed in direction x was considered to be equal to $0.14m$/hour, i.e., double the speed in direction y.

3.3 Linguistic variables and rule base

Fuzzy sets are a way to represent imprecise information and knowledge. The values of locomotion along the x and y directions are expressed as {*constant*, *difficult*, *medium*, *easy*} while those of the population (P) are expressed in term sets {*very small*, *small*, *medium*, *large*, *very large*}. The membership functions that specify the meaning of the linguistic values are shown in Fig. 3, 4, 6, 7 and 8 for locomotion along the x direction, locomotion along the y direction, total locomotion, population and dispersal, respectively. Total locomotion is the output variable in the fuzzy rule-based system (FRBS 1) that depends on the variables of locomotion along the x and y directions. Dispersal is the output variable in the fuzzy rule-based system (FRBS 2) that depends on the variables of total locomotion and population number, Fig. 5. Population dispersal varies from 0 to $0.14m^2$/hours, according to the result of the calculations of this part of region R of Fig. 1. Total locomotion is expressed in a term set {*constant*, *very difficult*, *difficult*, *medium*, *easy*}. The membership functions that designate the meaning of the linguistic values are given in Figs. 6 and 8 for total locomotion and dispersal, respectively. The rule base that encodes the relationship of locomotion along the x and y directions; and total locomotion is summarized in Table 1. The rule base that encodes the relationship of total locomotion, population and dispersal is shown in Table 2.

All the rule bases are processed using Mamdani's Inference Method with center of gravity defuzzification [22]. There are other inference methods that could be adopted, such as Takagi-Sugeno's Inference Method [23]. Nevertheless, we could not use this method because the FRBS output variables are not precise and also they can not be written as input variables.

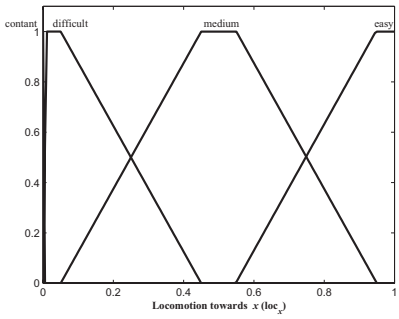


Figure 3: Membership functions for locomotion along the x direction (loc_x).

3.4 Fuzzy Computational Simulation

In order to define the type of ant locomotion as a function of the type of ground (e.g., rivers, hills, mountains and rocks)

ISBN: 978-989-95079-6-8

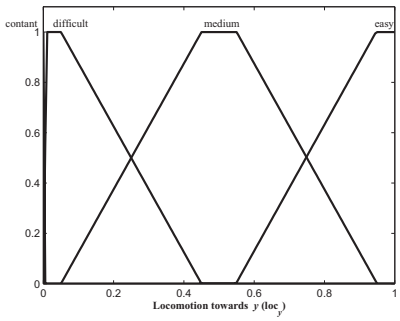


Figure 4: Membership functions for locomotion along the y direction (loc_y).

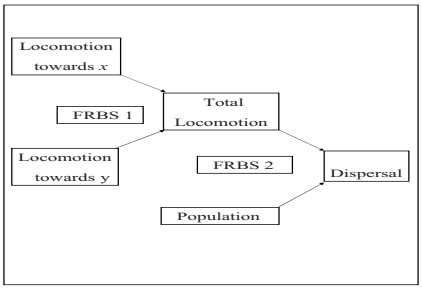


Figure 5: Fuzzy rule-based systems 1 and 2.

in the area of study to each node is allotted a value $[0, 1]$ for movement in x and y according to its location in the area, see Fig. 2.

The chosen trapezoidal membership function, which is associated to a fuzzy set A, $u_A(x)$, involving four parameters $[c_1\ c_2\ c_3\ c_4]$, see Fig. 9, is given by:

$$u_A(x) = \begin{cases} \dfrac{x - c_1}{c_2 - c_1} & \text{if } c_1 \leq x < c_2 \\ 1 & \text{if } c_2 \leq x \leq c_3 \\ \dfrac{-x + c_4}{c_4 - c_3} & \text{if } c_3 < x \leq c_4 \end{cases} \quad (10)$$

If the triangular elements are located in the area of difficult locomotion, the value for locomotion in those triangular elements is calculated by $c_1 + (c_4 - c_1) * rand$, where rand is equal to a random value in interval $(0, 1)$, $c_1 = 0$ e $c_4 = 0.45$. If the triangular elements are located in the area of medium locomotion, the value for locomotion in those triangular elements is calculated by $c_1 + (c_4 - c_1) * rand$, where rand is equal to a random value of $(0, 1)$, $c_1 = 0.05$ and $c_4 = 0.95$. If the triangular elements are located in the area of easy locomotion, the value for locomotion in those triangular elements is calculated by $c_1 + (1 - c_1) * rand$, since the values of x that represent the degree of membership between 0 and 1 for easy locomotion vary from $c_1 = 0.548$ to 1. Thus, locomotion in directions x and y in each triangle is determined randomly depending on the domain of the function of pertinence corresponding to the classification of the region of Fig.2. For horizontal and vertical locomotion, the triangular elements of the region of attraction are allotted the value of zero (0), which has a unitary fuzzy set. Therefore, each triangular element received a value for locomotion along x and y.

To obtain the fuzzy output for each triangular element, for FRBS 2, we use the average of population values on each of three nodes.

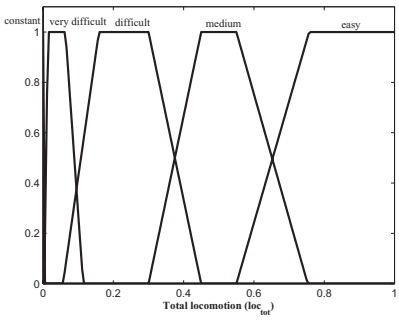


Figure 6: Membership functions for total locomotion (loc_{tot}).

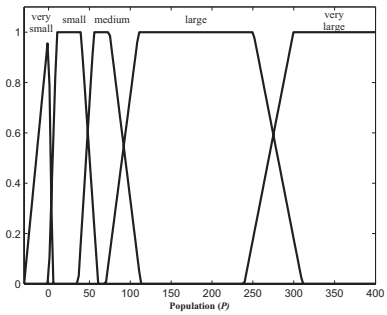


Figure 7: Membership functions for population (P).

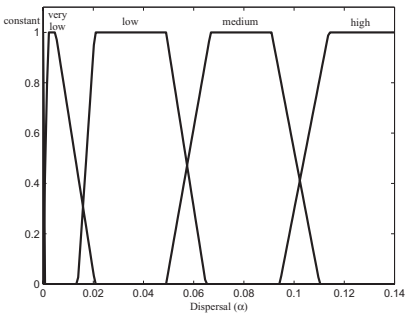


Figure 8: Membership functions for dispersal (α).

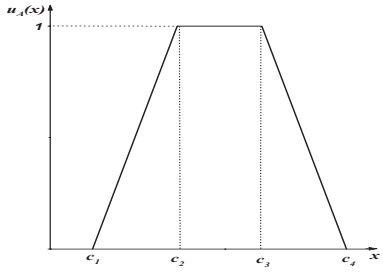


Figure 9: Parameters of trapezoidal membership functions.

ISBN: 978-989-95079-6-8

Table 1: Fuzzy rules for total locomotion (loc_{tot}).

(loc_x) \ (loc_y)	*constant*	*difficult*	*medium*	*easy*
constant	*constant*	*very difficult*	*difficult*	*medium*
difficult	*very difficult*	*difficult*	*difficult*	*medium*
medium	*difficult*	*difficult*	*medium*	*medium*
easy	*medium*	*medium*	*medium*	*medium*

Table 2: Fuzzy rules for dispersal (α).

(loc_{tot}) \ (P)	*very small*	*small*	*medium*	*large*	*very large*
constant	*constant*	*very low*	*very low*	*very low*	*very low*
very difficult	*constant*	*constant*	*very low*	*low*	*low*
difficult	*constant*	*very low*	*very low*	*medium*	*low*
medium	*constant*	*very low*	*very low*	*medium*	*medium*
easy	*constant*	*low*	*very low*	*medium*	*high*

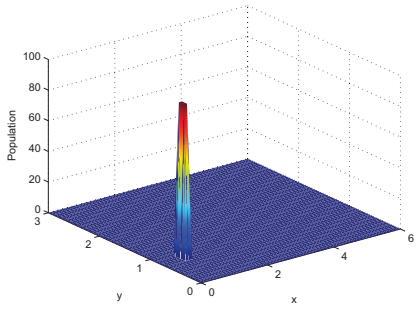


Figure 10: Initial population.

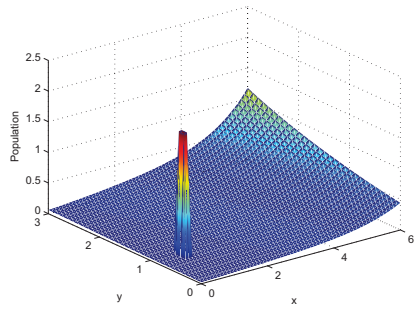


Figure 11: Population after 33 hours.

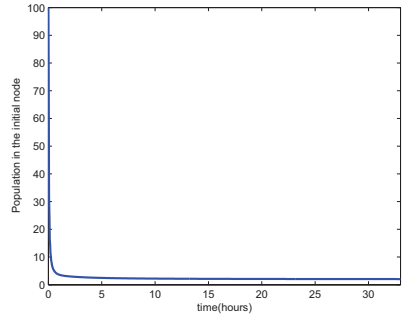


Figure 12: Population in the initial node.

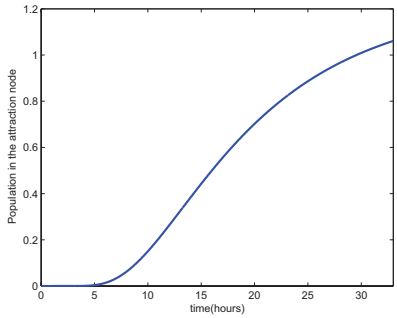


Figure 13: Population in one of the nodes of the region of attraction.

To determine the dispersal in each triangular element, the two FBRS were used: FRBS 1 and FRBS 2, Fig.5. Using FRBS 1, the value for total locomotion in each triangle was determined based on the values of locomotion along x and y. The dispersal in the node of each triangle was determined based on the values of total locomotion and medium population, using FRBS 2.

Thus, x and y components of locomotion were determined only once for each triangle of each region, prior to the first iteration, and total locomotion was determined based on the speed in directions x and y. The dispersion was calculated in each iteration for each triangle, since it depends on the total locomotion and on the population. The latter varies in each iteration.

Fig.10 illustrates the initial condition, i.e., the *A. cephalotes* colony before it occupied the region of attraction. The graph in Fig.11 shows the population after 33 hours, since ants take 33 hours to occupy the considered region of Fig.2. Fig.12 represents ant population in one of the initial nodes as a function of time. As it can be seen, the population in this node is close to two, since there will always be insects at the entrance to the ant nest. Fig. 13 shows the behavior of the number of ants as a function of time in one of the nodes of the region of attraction. Note that the population in this region grows logistically over time.

4 Conclusions

This paper proposes a model for the occupation of an attractive area by leaf-cutting ants, using a form of the diffusion-advection partial differential equation that models the population dispersal using a fuzzy parameter. Dispersion is this fuzzy parameter that depends on the number of the population and on its total locomotion, which is associated with the ease or difficulty of locomotion of the ants as a function of the terrain they traverse. The main difference between the classic models and the fuzzy model (1) is that the latter model exploits uncertainty parameters whereas the classic model does not. Thus, authors believe that, due to the uncertainty of biological phenomena, the combination of differential equations and fuzzy set theory enables computational simulations to much more faithfully portray the phenomenon under study.

Acknowledgement

The first author acknowledges CNPq, the Brazilian National Research Council, for its financial support in the form of a

ISBN: 978-989-95079-6-8

postdoctoral fellowship (152068/2007-4).

References

[1] B. Hölldobler and E.O. Wilson. *The Ants*. Belknap Press, Cambridge, MA, USA, 1990.

[2] H.L. Vasconcelos and H.G. Fowler. *Applied Myrmecology: a World Perspective*, chapter Foraging and fungal substrate selection by leaf-cutting ants, pages 410–419. Westview Press, Boulder, CO, USA, 1990.

[3] S.P. Hubbel and D.F. Wiemer. Host plant selection by an attine ant. In P. Jaisson, editor, *Social insects in the tropics*, volume 2, pages 133–154, Paris: University of Paris Press, 1983.

[4] F.M. Mundim, A.N. Costa, and H.L. Vasconcelos. Leaf nutrient content and host-plant selection by leaf-cutter ants, atta laevigata, in a neotropical savanna. *Entomologia Experimentalis et Applicata*, 130:47–54, 2009.

[5] H.G. Fowler and E.W. Stiles. Conservative resource management by leaf-cutting ants? the role of foraging territories and trails, and environmental patchiness. *Sociobiology*, 5:25–41, 1980.

[6] J.D. Shepherd. Trunk trails and searching strategy of a leaf-cutting ant, atta colombica. *Behavioural Ecology and Sociobiology*, 11:77–84, 1982.

[7] C.M. Blanton and J.J. Ewel. Leaf-cutting ant herbivory in successional and agricultural tropical ecosystems. *Ecology*, 66(3):861–869, 1985.

[8] R.S. Cantrell and C. Cosner. *Spatial Ecology via Reaction-Diffusion Equations*. John Wiley and Sons, Ltd, England, 2003.

[9] L.C. Barros, R.C. Bassanezi, and M.B. Leite. The epidemiological models SI with fuzzy parameter of transmission. *Computers and Mathematics with Applications*, 45:1619–1628, 2003.

[10] R.M. Jafelice, L.C. Barros, R.C. Bassanezi, and F. Gomide. Fuzzy modeling in asymptomatic HIV virus infected population. *Bulletin of Mathematical Biology*, 66(6):1597–1620, 2004.

[11] R.M. Jafelice, L.C. Barros, R.C. Bassanezi, and F. Gomide. Methodology to determine the evolution of asymptomatic HIV population using fuzzy set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 13:39–58, 2005.

[12] N. Ortega, L.C. Barros, and E. Massad. Fuzzy gradual rules in epidemiology. *Kybernetes: The International Journal of Systems and Cybernetics*, 32(4):460–477, 2003.

[13] V. Krivan and G. Colombo. A non-stochastic approach for modeling uncertainty in population dynamics. *Bulletin of Mathematical Biology*, 60, 1998.

[14] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[15] V. Rozin and M. Margaliot. The fuzzy ant. *Computational Intelligence Magazine, IEEE*, 2(4):18–28, 2007.

[16] M. Dorigo, M. Birattari, and T. Stützle. Ant colony optimization: Artificial ants as a computational intelligence technique. *IEEE Computational Intelligence Magazine*, pages 28–39, November 2006.

[17] H.L. Vasconcelos. Foraging activity of two species of leaf-cutting ants (atta) in a primary forest of the central amazon. *Insectes Sociaux*, 37(2):131–145, 1990.

[18] M.N.G. Ribeiro. Aspectos climatológicos de manaus. *Acta Amazonia*, 6:229–233, 1976.

[19] D. Dubois and H. Prade. *Fuzzy sets and systems - Theory and applications*. Academic press, New York, 1980.

[20] C.G. Almeida, R.S.M. Jafelice, and J.F.C.A. Meyer. Modelagem fuzzy em dinâmica populacional. In *Proccedings XXIX Congresso Nacional de Matemática Aplicada e Computacional*, Campinas, Brazil, 2006 (in Portuguese).

[21] L.C. Paraíba and P. Pulino. Simulação numérica da dispersão-advecção de pesticidas sob efeito da temperatura do solo: o modelo da peste. Relatório de Pesquisa 04, IMECC, UNICAMP, Campinas, Brazil, 2003 (in Portuguese).

[22] W. Pedrycz and F. Gomide. *An Introduction to Fuzzy Sets: Analysis and Design*. MIT Press, Cambridge, EUA, 1998.

[23] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

ISBN: 978-989-95079-6-8

Continuous R-implications

Balasubramaniam Jayaram[1] Michał Baczyński[2]

1. Department of Mathematics, Indian Institute of Technology Madras, Chennai 600 036, India
2. Institute of Mathematics, University of Silesia, 40-007 Katowice, ul. Bankowa 14, Poland
Email: jbala@iitm.ac.in, michal.baczynski@us.edu.pl

Abstract*— *In this work we have solved an open problem related to the continuity of R-implications. We have fully characterized the class of continuous R-implications obtained from any arbitrary t-norm. Using this result, we also determine the exact intersection between the continuous subsets of R-implications and (S,N)-implications.

***Keywords*—** R-implication, (S,N)-implication, Łukasiewicz implication, t-norm.

1 Introduction

R-implications and (S,N)-implications are two of the most established families of fuzzy implications. Still, many open problems remain unsolved, see [3, 4]. One of them is related to the continuous subsets of these families. Only recently a characterization of continuous (S,N)-implications was given by the authors in [2].

However, a similar complete characterization regarding the continuous subset of R-implications has not been available so far. It is only known that in the class of R-implications obtained from *left-continuous* t-norms, the only continuous elements are those that are isomorphic to the Łukasiewicz implication, i.e., those R-implications obtained as residuals of nilpotent t-norms. In particular, the following question has remained unanswered so far:

> Does there exist a continuous R-implication obtained from a non-left continuous t-norm?

In this note we show that an R-implication I_T obtained from a t-norm T is continuous if and only if T is a nilpotent t-norm.

Using this result, we are also able to resolve another question related to the intersections between these two families, which is also a generalization of an original result of Smets and Magrez [13], see also [7, 9]. We show that the only continuous (S,N)-implication that is also an R-implication obtained from any t-norm, not necessarily left-continuous, is the Łukasiewicz implication up to an isomorphism.

2 Preliminaries

We assume that the reader is familiar with the classical results concerning basic fuzzy logic connectives, but to make this work more self-contained, we introduce basic notations used in the text and we briefly mention some of the concepts and results employed in the rest of the work.

Definition 1 (cf. [7, 10, 9])**.**

(i) A function $N\colon [0,1] \to [0,1]$ is called a *fuzzy negation*, if it is decreasing and satisfies the boundary conditions $N(1) = 0$ and $N(0) = 1$.

(ii) A fuzzy negation N is called *strong*, if it is an involution, i.e., $N \circ N = \mathrm{id}_{[0,1]}$.

(iii) A function $T\colon [0,1]^2 \to [0,1]$ is called a *t-norm*, if it is increasing in both variables, commutative, associative and has 1 as the neutral element.

(iv) A function $S\colon [0,1]^2 \to [0,1]$ is called a *t-conorm*, if it is increasing in both variables, commutative, associative and has 0 as the neutral element.

(v) A t-norm T is said to be *border continuous*, if it is continuous on the boundary of the unit square $[0,1]^2$, i.e., on the set $[0,1]^2 \setminus (0,1)^2$.

(vi) A t-norm T is said to be *left-continuous*, if it is left-continuous in each component.

(vii) A t-norm T is said to be *nilpotent*, if it is continuous and if each $x \in (0,1)$ is a nilpotent element of T, i.e., if there exists $n \in \mathbb{N}$ such that $x_T^{[n]} = 0$, where

$$x_T^{[n]} := \begin{cases} 1, & \text{if } n = 0, \\ x, & \text{if } n = 1, \\ T(x, x_T^{[n-1]}), & \text{if } n > 1. \end{cases}$$

(viii) A t-norm T is said to be *Archimedean* if for every $x, y \in (0,1)$ there exists $n \in \mathbb{N}$ such that $x_T^{[n]} < y$.

Remark 1 (see. [10, p. 17])**.** For the border continuity of a t-norm T, it is sufficient to require the continuity on the upper right boundary, since from the monotonicity we get

$$\lim_{x\to 0^+} T(x,y) \le \lim_{x\to 0^+} T(x,1) = \lim_{x\to 0^+} x = 0 = T(0,y),$$

for any $y \in [0,1]$.

Remark 2. From the commutativity, the left-continuity of a t-norm T is equivalent to the left-continuity of T with respect to the first or the second variable. Moreover, $T(x,1) = 1$ and $T(x,0) = 0$ for every $x \in [0,1]$, thus a t-norm T is left-continuous if and only if for any $y \in (0,1)$ and every increasing sequence $(x_n)_{n\in\mathbb{N}}$, where $x_n \in [0,1)$, we have

$$\lim_{n\to\infty} T(x_n, y) = T(\lim_{n\to\infty} x_n, y).$$

Proposition 1. *If T is an Archimedean t-norm, then $T(x,y) < \min(x,y)$, for all $x, y \in (0,1)$.*

ISBN: 978-989-95079-6-8

Proof. Let T be an Archimedean t-norm. If, on the contrary, there exist some $x_0, y_0 \in (0,1)$ such that $x_0 \geq y_0$ and $T(x_0, y_0) = y_0 = \min(x_0, y_0)$, then we will prove, by induction, that for every $n \in \mathbb{N}$ we have

$$x_0{}_T^{[n]} \geq y_0. \tag{1}$$

Indeed, firstly see that

$$x_0{}_T^{[0]} = 1 > T(x_0, y_0)) = y_0,$$
$$x_0{}_T^{[1]} = x_0 \geq T(x_0, y_0)) = y_0.$$

Let us assume that (1) holds for some $n \in \mathbb{N}$. Then by the monotonicity of T and our inductive assumption we get

$$x_0{}_T^{[n+1]} = T(x_0, x_0{}_T^{[n]}) \geq T(x_0, y_0) = y_0,$$

which implies that T is not Archimedean, a contradiction. □

By Φ we denote the family of all increasing bijections $\varphi\colon [0,1] \to [0,1]$. We say that two functions $f, g\colon [0,1]^n \to [0,1]$, where $n \in \mathbb{N}$, are Φ-conjugate, if there exists $\varphi \in \Phi$ such that $g = f_\varphi$, where

$$f_\varphi(x_1, \ldots, x_n) := \varphi^{-1}\left(f(\varphi(x_1), \ldots, \varphi(x_n))\right),$$

for all $x_1, \ldots, x_n \in [0,1]$. Equivalently, g is said to be the Φ-conjugate of f.

Definition 2 ([7, 4])**.** A function $I\colon [0,1]^2 \to [0,1]$ is called a *fuzzy implication* if it satisfies the following conditions:

I is decreasing in the first variable, (I1)

I is increasing in the second variable, (I2)

$I(0,0) = 1, \qquad I(1,1) = 1, \qquad I(1,0) = 0.$ (I3)

The set of all fuzzy implications will be denoted by $\mathcal{FI}$.

3 R-implications

Definition 3 (cf. [15, 7, 9])**.** A function $I\colon [0,1]^2 \to [0,1]$ is called an *R-implication*, if there exists a t-norm T such that

$$I(x,y) = \sup\{t \in [0,1] \mid T(x,t) \leq y\}, \tag{2}$$

for all $x, y \in [0,1]$. If an R-implication is generated from a t-norm T, then we will often denote this by I_T.

It is important to note that the name 'R-implication' is a short version of 'residual implication', and I_T is also called as the residuum of T (see e.g. [7, 9, 10]).

Example 1. The Łukasiewicz implication

$$I_{\mathbf{LK}}(x,y) = \min(1, 1 - x + y), \qquad x, y \in [0,1],$$

is an R-implication obtained from the nilpotent (Łukasiewicz) t-norm

$$T_{\mathbf{LK}}(x,y) = \max(x + y - 1, 0), \qquad x, y \in [0,1].$$

For more well-known R-implications along with their t-norms from which they have been obtained, we refer the readers to other sources, notably [7, 10, 4].

Theorem 1 (cf. [7], [3, Theorem 5.5])**.** *If T is any t-norm, then $I_T \in \mathcal{FI}$ and it satisfies the left neutrality property, i.e.,*

$$I_T(1,y) = y, \qquad y \in [0,1], \tag{NP}$$

and the identity principle, i.e.,

$$I_T(x,x) = 1, \qquad x \in [0,1]. \tag{IP}$$

Moreover, if T is left-continuous, then I_T satisfies the exchange principle, i.e.,

$$I_T(x, I_T(y,z)) = I_T(y, I_T(x,z)), \qquad x, y, z \in [0,1],$$

and the ordering property, i.e.,

$$x \leq y \iff I_T(x,y) = 1, \qquad x, y \in [0,1].$$

Proposition 2 ([3, Proposition 5.8])**.** *For a t-norm T the following statements are equivalent:*

(i) T is border continuous.

(ii) I_T satisfies the ordering property.

For R-implications generated form left-continuous t-norms we have the following results.

Proposition 3 (cf. [9, Proposition 5.4.2 and Corollary 5.4.1])**.** *For a t-norm T the following statements are equivalent:*

(i) T is left-continuous.

(ii) T and I_T form an adjoint pair, i.e., they satisfy the residuation property

$$T(x,t) \leq y \iff I_T(x,y) \geq t, \qquad x, y, t \in [0,1].$$

(iii) The supremum in (2) *is the maximum, i.e.,*

$$I_T(x,y) = \max\{t \in [0,1] \mid T(x,t) \leq y\},$$

where the right side exists for all $x, y \in [0,1]$.

Using the above result we are able to obtain the following characterization.

Theorem 2 (cf. [8, Corollary 2])**.** *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I is a continuous R-implication based on some left-continuous t-norm.

(ii) I is Φ-conjugate with the Łukasiewicz implication, i.e., there exists $\varphi \in \Phi$, which is uniquely determined, such that

$$I(x,y) = \varphi^{-1}(\min(1 - \varphi(x) + \varphi(y), 1)), \tag{3}$$

for all $x, y \in [0,1]$.

For more facts related to R-implication see e.g. [7, 3, 4].

ISBN: 978-989-95079-6-8

4 Continuous Partial Functions of R-implications

Note that from Theorem 1 we can consider, for any fixed $\alpha \in [0,1)$, the decreasing partial function $I_T(\cdot,\alpha)\colon [\alpha,1] \to [\alpha,1]$, which we will denote by g_α^T. Observe that g_α^T is decreasing and such that $g_\alpha^T(\alpha) = 1$ and $g_\alpha^T(1) = \alpha$.

Remark 3. If the domain of g_α^T is extended to $[0,1]$, then this is exactly what are called contour lines by Maes and De Baets in [11, 5]. If $\alpha = 0$, then g_0^T is the natural negation associated with the t-norm T (see [3]).

Theorem 3. *Let T be any t-norm and $\alpha \in [0,1)$. If g_α^T is continuous, then g_α^T is strictly decreasing.*

Proof. Let T be any t-norm and $\alpha \in [0,1)$ be fixed. We know that g_α^T is decreasing. On the contrary, let us assume that g_α^T is constant on some interval $[x_0, y_0]$ for some $\alpha < x_0 < y_0 < 1$, i.e., there exists $p \in [\alpha, 1]$ such that

$$g_\alpha^T(x_0) = g_\alpha^T(y_0) = p.$$

Let us fix arbitrarily $z \in (x_0, y_0)$.

Firstly, consider the case $p = 1$. Then

$$g_\alpha^T(z) = I_T(z,\alpha) = \sup\{t \in [0,1] \mid T(z,t) \le \alpha\} = 1,$$

thus $T(z, 1-\varepsilon) \le \alpha$ for any $\varepsilon \in (0,1)$. Hence

$$g_\alpha^T(1-\varepsilon) = \sup\{t \in [0,1] \mid T(1-\varepsilon, t) \le \alpha\} \ge z,$$

for all $\varepsilon \in (0, 1-\alpha)$. However, by the continuity of g_α^T, as $\varepsilon \to 0^+$, we get

$$\alpha = g_\alpha^T(1) = g_\alpha^T(\lim_{\varepsilon\to 0^+} 1-\varepsilon) = \lim_{\varepsilon\to 0^+} g_\alpha^T(1-\varepsilon) \ge \lim_{\varepsilon\to 0^+} z = z,$$

a contradiction to the fact that $\alpha < x_0 < z$.

If $p = \alpha$, then

$$g_\alpha^T(z) = I_T(z,\alpha) = \sup\{t \in [0,1] \mid T(z,t) \le \alpha\} = \alpha,$$

thus $T(z, \alpha+\varepsilon) > \alpha$ for all $\varepsilon \in (0, 1-\alpha)$. Hence

$$g_\alpha^T(\alpha+\varepsilon) = \sup\{t \in [0,1] \mid T(\alpha+\varepsilon, t) \le \alpha\} \le z,$$

for all $\varepsilon \in (0, 1-\alpha)$. Once again, by the continuity of g_α^T we have, as $\varepsilon \to 0^+$, that

$$1 = g_\alpha^T(\alpha) = g_\alpha^T(\lim_{\varepsilon\to 0^+} \alpha+\varepsilon) = \lim_{\varepsilon\to 0^+} g_\alpha^T(\alpha+\varepsilon) \le \lim_{\varepsilon\to 0^+} z = z,$$

a contradiction to the fact that $z < 1$.

Finally, let $p \in (\alpha, 1)$. Then, by the definition of g_α^T, we have

$$T(z, p+\varepsilon) > \alpha \ge T(z, p-\varepsilon),$$

for any $\varepsilon > 0$ such that $p + \varepsilon \le 1$ and $p - \varepsilon \ge \alpha$. Therefore

$$I_T(p+\varepsilon, \alpha) \le z \le I_T(p-\varepsilon, \alpha),$$

hence

$$g_\alpha^T(p+\varepsilon) \le z \le g_\alpha^T(p-\varepsilon).$$

Since g_α^T is continuous, we have, as $\varepsilon \to 0^+$, that

$$g_\alpha^T(p) = z.$$

Now this happens for every $z \in (x, y)$, which contradicts the fact that g_α^T is a function itself. Hence g_α^T is strictly decreasing. □

5 Main results: continuous R-implications

The main result of this work is the generalization of Theorem 2, viz., we show that the *left-continuity* of the underlying t-norm is implied and need not be assumed. Thus we give a complete characterization of the class of all continuous R-implications by showing that it is equivalent to the class of fuzzy implications which are Φ-conjugate to the Łukasiewicz implication.

Theorem 4. *Let T be a t-norm and I_T the R-implication obtained from it. If I_T is continuous, then T is border continuous.*

Proof. On the contrary, let us assume that I_T is continuous and T is not border continuous. Then, by Remark 1, there exist $y_0 \in (0,1)$ and an increasing sequence $(x_n)_{n\in\mathbb{N}}$, where $x_n \in [0,1)$, such that $\lim_{n\to\infty} x_n = 1$, but

$$\lim_{n\to\infty} T(x_n, y_0) = y' < y_0 = T(1, y_0).$$

This implies, in particular, that

$$I_T(y_0, y') = \sup\{t \in [0,1] \mid T(y_0, t) \le y'\} = 1.$$

Now, by (I1) and (IP) of I_T (cf. Theorem 1) we have that

$$1 = I_T(y_0, y') \le I_T(y', y') = 1,$$

i.e., $I_T(x, y') = 1$ for all $x \in [y', y_0]$. Note that $I_T(\cdot, y') = g_{y'}^T$. Since I_T is continuous we have that $g_{y'}^T$ is also continuous and from Theorem 3 we see that it is strictly decreasing. However, from the above, we see that $g_{y'}^T$ is constant on $[y', y_0]$, a contradiction. Thus T is border continuous. □

Theorem 5. *Let T be a t-norm and I_T the R-implication obtained from it. If I_T is continuous, then T is Archimedean.*

Proof. Let T be a t-norm. On the contrary, let us assume that I_T is continuous and T is non-Archimedean. Then, by the Definition 1(viii) there exist $x_0, y_0 \in (0,1)$ such that for all $n \in \mathbb{N}$ we have that $x_0{}_T^{[n]} \ge y_0$.

Let us denote

$$X_0 := \{z \in [0,1] \mid x_0{}_T^{[n]} > z \text{ for all } n \in \mathbb{N}\}.$$

Observe, that $X_0 \neq \emptyset$ since for all $y < y_0$ we have that $x_0{}_T^{[n]} > y$ for all $n \in \mathbb{N}$. Further, let

$$z_0 := \sup X_0.$$

See that $0 < z_0 \le x_0$ and $z_0 - \varepsilon \in X_0$ for all $\varepsilon \in (0, z_0]$. Also, if $t > z_0$, then there exists $m \in \mathbb{N}$ such that $x_0{}_T^{[m]} \le t$, which implies that

$$z_0 - \varepsilon < x_0{}_T^{[m+1]} = T(x_0, x_0{}_T^{[m]}) \le T(x_0, t),$$

for any $t > z_0$. Hence

$$I_T(x_0, z_0-\varepsilon) = \sup\{t \in [0,1] \mid T(x_0, t) \le z_0 - \varepsilon\} \le z_0,$$

for all $\varepsilon \in (0, z_0]$. From the continuity of I_T we get

$$I_T(x_0, z_0) = I_T(x_0, \lim_{\varepsilon\to 0^+} z_0 - \varepsilon) = \lim_{\varepsilon\to 0^+} I_T(x_0, z_0-\varepsilon) \le \lim_{\varepsilon\to 0^+} z_0 = z_0.$$

ISBN: 978-989-95079-6-8

Now, by (I1) and (NP) of I_T (cf. Theorem 1) we have that

$$z_0 \geq I_T(x_0, z_0) \geq I_T(1, z_0) = z_0,$$

i.e., $I_T(x, z_0) = z_0$ for all $x \in [x_0, 1]$. Note that $I_T(\cdot, z_0) = g^T_{z_0}$. Since I_T is continuous we have that $g^T_{z_0}$ is also continuous and from Theorem 3 we see that it is strictly decreasing. However, from the above, we have that $g^T_{z_0}$ is constant on $[x_0, 1]$, a contradiction. Thus T is Archimedean. □

Theorem 6. *Let T be a t-norm and I_T the R-implication obtained from it. If I_T is continuous, then T is left-continuous.*

Proof. Let T be a t-norm such that I_T is continuous. From Theorems 4 and 5 we see that T is border continuous and Archimedean.

On the contrary, let us assume that T is non-left-continuous. From Remark 2 there exist $x_0 \in (0, 1]$, $y_0 \in (0, 1)$ and an increasing sequence $(x_n)_{n\in\mathbb{N}}$, where $x_n \in [0, 1)$, such that $\lim_{n\to\infty} x_n = x_0$, but

$$\lim_{n\to\infty} T(x_n, y_0) = z' < z_0 = T(x_0, y_0).$$

Since T is border continuous it suffices to consider the case when $x_0 \in (0, 1)$.

Firstly observe that

$$I_T(y_0, z') = \sup\{t \in [0, 1] \mid T(y_0, t) \leq z'\} = x_0, \qquad (4)$$

since from the monotonicity of T we have $T(y_0, x_n) \leq z'$ for every $n \in \mathbb{N}$ and $T(y_0, x_0) = z_0 > z'$.

Next, from Proposition 1, by the Archimedeanness and monotonicity of T, we get that for any arbitrary $\varepsilon \in (0, 1-x_0)$ the following inequality holds

$$T(x_0, 1 - \varepsilon) < \min(x_0, 1 - \varepsilon) = x_0. \qquad (5)$$

Now, by (4) and (5) we get

$$T(x_0, 1 - \varepsilon) < I_T(y_0, z'),$$

for any $\varepsilon \in (0, 1 - x_0)$, thus

$$T(x_0, 1 - \varepsilon) < \sup\{t \in [0, 1] \mid T(y_0, t) \leq z'\},$$

hence

$$T(y_0, T(x_0, 1 - \varepsilon)) \leq z'$$

By the associativity of T we get

$$T(T(x_0, y_0), 1 - \varepsilon) \leq z',$$

i.e.,

$$T(z_0, 1 - \varepsilon) \leq z'.$$

for any $\varepsilon \in (0, 1 - x_0)$. This implies that

$$\lim_{\varepsilon\to 0^+} T(z_0, 1 - \varepsilon) \leq z' < z_0 = T(z_0, 1),$$

i.e., T is not border continuous, a contradiction to Theorem 4 and hence T is left-continuous. □

From Theorems 2 and 6 we obtain the following result.

Corollary 1. *For a function $I\colon [0, 1]^2 \to [0, 1]$ the following statements are equivalent:*

(i) I is a continuous R-implication based on some t-norm.

(ii) I is Φ-conjugate with the Łukasiewicz implication, i.e., there exists $\varphi \in \Phi$, which is uniquely determined, such that I has the form (3) *for all $x, y \in [0, 1]$.*

Let us denote the different families of fuzzy implications as follows:

- $\mathbb{I}_{\mathbb{T}}$ – the family of all R-implications;
- $\mathbb{I}_{\mathbb{T}_{\mathbb{LC}}}$ – the family of all R-implications obtained from left-continuous t-norms;
- ${}^{\mathbb{C}}\mathbb{I}_{\mathbb{T}}$ – the family of all continuous R-implications;
- ${}^{\mathbb{C}}\mathbb{I}_{\mathbb{T}_{\mathbb{LC}}}$ – the family of all continuous R-implications obtained from left-continuous t-norms;
- $\mathbb{I}_{\mathbb{LK}}$ – the family of all fuzzy implications Φ-conjugate with the Łukasiewicz implication $I_{\mathbf{LK}}$.

From Corollary 1 we have the following equalities between the above sets:

$${}^{\mathbb{C}}\mathbb{I}_{\mathbb{T}_{\mathbb{LC}}} = {}^{\mathbb{C}}\mathbb{I}_{\mathbb{T}} = \mathbb{I}_{\mathbb{LK}}.$$

6 (S,N)-implications

Definition 4 (cf. [15, 7, 1, 2])**.** A function $I\colon [0, 1]^2 \to [0, 1]$ is called an *(S,N)-implication*, if there exist a t-conorm S and a fuzzy negation N such that

$$I(x, y) = S(N(x), y), \qquad x, y \in [0, 1].$$

If N is a strong negation, then I is called an S-implication. Moreover, if an (S,N)-implication is generated from S and N, then we will often denote this by $I_{S,N}$.

Firstly note that $I_{S,N} \in \mathcal{FI}$ for any t-conorm S and any fuzzy negation N. In the class of continuous (S,N)-implications we have the following important result.

Proposition 4 ([2, Proposition 5.4])**.** *For a function $I\colon [0, 1]^2 \to [0, 1]$ the following statements are equivalent:*

(i) I is a continuous (S,N)-implication.

(ii) I is an (S,N)-implication generated from some continuous t-conorm S and some continuous fuzzy negation N.

Let us denote the different families of fuzzy implications as follows:

- $\mathbb{I}_{\mathbb{S},\mathbb{N}}$ – the family of all (S,N)-implications;
- ${}^{\mathbb{C}}\mathbb{I}_{\mathbb{S},\mathbb{N}}$ – the family of all continuous (S,N)-implications;
- $\mathbb{I}_{\mathbb{S}_{\mathbb{C}},\mathbb{N}_{\mathbb{C}}}$ – the family of all (S,N)-implications obtained from continuous t-conorms and continuous negations.

Observe, that from Proposition 4 we get

$${}^{\mathbb{C}}\mathbb{I}_{\mathbb{S},\mathbb{N}} = \mathbb{I}_{\mathbb{S}_{\mathbb{C}},\mathbb{N}_{\mathbb{C}}}.$$

ISBN: 978-989-95079-6-8

7 Intersection between continuous R- and (S,N)-implications

The intersections between the families and subfamilies of R- and (S,N)-implications have been studied by many authors, see e.g. [6, 13, 7, 3]. As regards the intersection between their continuous subsets only the following result has been known so far.

Theorem 7. *The only continuous (S,N)-implications that are also R-implications obtained from left-continuous t-norms are the fuzzy implications which are Φ-conjugate with the Łukasiewicz implication, i.e.,*

$$^{\mathrm{C}}\mathbb{I}_{\mathrm{S,N}} \cap \mathbb{I}_{\mathrm{T_{LC}}} = \mathbb{I}_{\mathrm{LK}}.$$

Now, from Corollary 1, we can prove the following equivalences.

Theorem 8. *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I is a continuous (S,N)-implication that is also an R-implication obtained from a left-continuous t-norm.

(ii) I is a continuous (S,N)-implication that is also an R-implication.

(iii) I is an (S,N)-implication that is also a continuous R-implication.

(iv) I is Φ-conjugate with the Łukasiewicz implication, i.e., there exists $\varphi \in \Phi$, which is uniquely determined, such that I has the form (3) *for all $x, y \in [0,1]$.*

In summary, we have

$$^{\mathrm{C}}\mathbb{I}_{\mathrm{S,N}} \cap \mathbb{I}_{\mathrm{T_{LC}}} = {}^{\mathrm{C}}\mathbb{I}_{\mathrm{S,N}} \cap \mathbb{I}_{\mathrm{T}} = \mathbb{I}_{\mathrm{S,N}} \cap {}^{\mathrm{C}}\mathbb{I}_{\mathrm{T}} = {}^{\mathrm{C}}\mathbb{I}_{\mathrm{T}} = \mathbb{I}_{\mathrm{LK}}.$$

8 Conclusions

In this paper, we have shown that continuous R-implications cannot be obtained from purely non-left-continuous t-norms and that the only continuous R-implications are those that are Φ-conjugate with the Łukasiewicz implication. Using this result we have been able to answer another question related to the intersection between the continuous sub-families of (S,N)- and R-implications. We believe that this work will further help in solving some of the open problems still remaining with regards these two basic families of fuzzy implications.

References

[1] C. Alsina and E. Trillas. When (S,N)-implications are (T, T_1)-conditional functions? *Fuzzy Sets and Systems* 134 (2003) 305–310.

[2] M. Baczyński and B. Jayaram. On the characterizations of (S,N)-implications. *Fuzzy Sets and Systems* 158 (2007) 1713–1727.

[3] M. Baczyński and B. Jayaram. (S,N)- and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems* 159 (2008) 1836–1859.

[4] M. Baczyński and B. Jayaram. *Fuzzy implications.* (Studies in Fuzziness and Soft Computing, Vol. 231), Springer, Berlin, 2008.

[5] B. De Baets and K.C. Maes. Orthosymmetrical monotone functions. *Bull. Belg. Math. Soc. Simon Stevin* 14 (2007) 99–116.

[6] D. Dubois and H. Prade. A theorem on implication functions defined from triangular norms. *Stochastica* 8 (1984), 267–279.

[7] J. Fodor and M. Roubens. *Fuzzy preference modelling and multicriteria decision support.* Kluwer, Dordrecht, 1994.

[8] J.C. Fodor. Contrapositive symmetry of fuzzy implications. *Fuzzy Sets and Systems* 69 (1995) 141–156.

[9] S. Gottwald. *A treatise on many-valued logic.* Research Studies Press, Baldock, 2001.

[10] E.P. Klement, R. Mesiar and E. Pap. *Triangular norms.* Kluwer, Dordrecht, 2000.

[11] K.C. Maes and B. De Baets. A contour view on uninorm properties. *Kybernetika* 42 (2006) 303–318.

[12] B. Schweizer and A. Sklar. *Probabilistic metric spaces.* North-Holland, New York, 1983.

[13] P. Smets and P. Magrez. Implication in fuzzy logic. *Internat. J. Approx. Reason.* 1 (1987) 327–347.

[14] E. Trillas and L. Valverde, On some functionally expressable implications for fuzzy set theory. In: E.P. Klement (Ed.) *Proc. 3rd Inter. Seminar on Fuzzy Set Theory.* Linz, Austria, 1981, pp. 173–190.

[15] E. Trillas and L. Valverde. On implication and indistinguishability in the setting of fuzzy logic. In: J. Kacprzyk and R.R. Yager (Eds.), *Management decision support systems using fuzzy sets and possibility theory.* TÜV-Rhineland, Cologne, 1985, pp. 198–212.

ISBN: 978-989-95079-6-8

Fuzzy Voxel Object

Derek T. Anderson, Robert H. Luke, Erik Stone, and James M. Keller

Department of Electrical and Computer Engineering
University of Missouri, Columbia, USA
Email: {dtaxtd, rhl3db, erikstone}@mizzou.edu and kellerj@missouri.edu

Abstract—*In this paper, computer vision and fuzzy set theory are merged for the robust construction of three-dimensional objects using a small number of cameras and minimal a priori knowledge about the objects. This work extends our previously defined crisp model, which has been successfully used for recognizing and linguistically summarizing human activity. The objects true features more closely resemble the fuzzy object than those of the crisp object. This is demonstrated both empirically and through the comparison of features used in tracking human activity.*

Keywords—computer vision, human activity analysis, fuzzy objects, fuzzy voxel person.

1 Introduction

Information gathered from three-dimensional models is significantly more useful and robust than that gathered from two-dimensional sources. This is quite apparent in the area of human surveillance. Features, such as the centroid, height, velocity, orientation, etc, in three-dimensional space are not camera view dependent. Object segmentation can also be improved using three-dimensional information about the environment, such as floors, walls, and objects for the removal of additional erroneous artefacts, such as shadows and specular highlights [1][2]. At each step in computer vision, be it construction, shape refinement, etc, various forms of uncertainty are present and should be utilized.

We previously demonstrated a method for constructing a three-dimensional human model, voxel person, in voxel (volume element) space (figure 1) through the use of privacy protected images of the human called silhouettes [1][2].

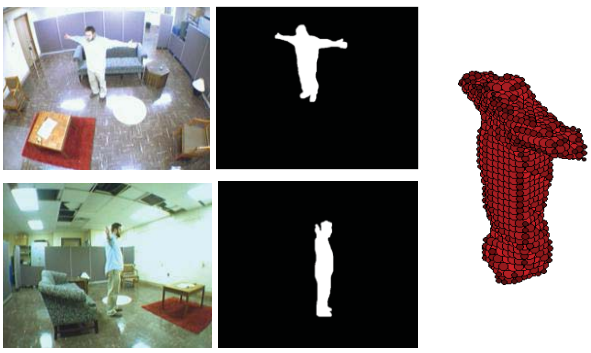


Fig. 1. Crisp voxel person created at a voxel resolution of 1.25"x1.25"x1.25", using 2 cameras and image silhouettes.

A silhouette is a binary map that distinguishes an individual (foreground) from his or her background. Well-known algorithms for computing the silhouette include Mixtures of Gaussians [3], Eigen Backgrounds [4], and Wallflower [5]. Silhouettes are used and we extract linguistic summarizations of human state based on crisp voxel person and fuzzy logic [1]. The resulting information is in a natural format for humans (linguistic) and it also yields succinct descriptions for managing complexity. We extended the work in [1] using a hierarchy of fuzzy logic and linguistic summarization for the inference of activity [2]. The knowledge (rules) is designed under the supervision of nurses for the recognition of falls in elderly populations.

Construction of three-dimensional objects, both solid representations and hulls, from two-dimensional images has been studied intensely in computer graphics, computer vision, biomedicine, and even in the activity analysis domain [6][7][8][9][10]. What separates our object construction related work from most, besides the use of silhouettes for back-projection, is the way in which a small number of low cost cameras, typically two per environment, are used to build a low, but rich for tracking, resolution voxel object; how its shape is refined using environment knowledge; and how features are extracted and used for fuzzy-based activity recognition [1][2]. Segmentation of the object into body parts (torso, head, etc) is not attempted, such as in [11][12]. While advances have been made in body segmentation, no approach to date is either real-time or mature enough to be included in a real-world eldercare system that runs unsupervised for long time periods.

This paper is divided as such. Section 2 describes the crisp voxel model, section 3 is the fuzzy model, and section 4 is the assesment of the quality of construction. Feature extraction is described next, followed by an alpha-cut procedure for an improved crisp voxel person, and section 7 is experiments and qualitative and quantitative results.

2 Crisp Voxel Object

An environment used for tracking is first converted to a voxel representation. Voxels are non-overlapping cubes that discretize a volume. The voxel centered at position $< i, j, l >$ is $v_{(i,j,l)}$. For each pixel in each camera, a list of intersected voxels is pre-computed, $L_{c,(n,m)}$, where c is the camera index, $1 \le c \le C$, and (n,m) is the pixel index [1]. The lists are sorted according to their distance (Euclidean) to the camera. This way, voxels that correspond to regions of change (the foreground, e.g. the human) from the viewpoint of a given camera can be quickly identified. For camera c at time t, the human voxel object is the union of all voxel lists

ISBN: 978-989-95079-6-8

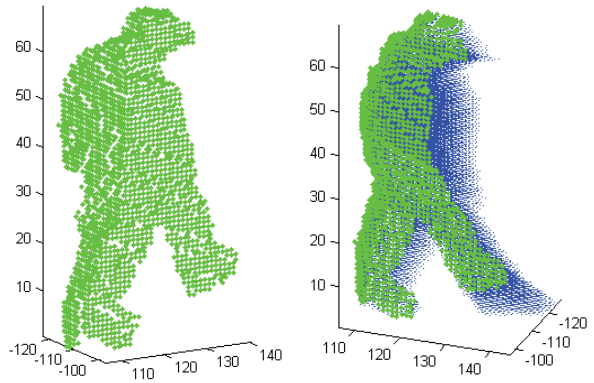
for the foreground pixel set, V_t^c. The object's shape is refined by intersecting the voxel objects acquired from each camera in the scene, i.e. $\{V_t^1,\ldots,V_t^C\}$, thus

$$V'_t = \wedge_{c=1}^{C} V_t^c . \quad (1)$$

The object is then subject to additional morphological operators for noise removal and voxel space processing for removing additional shadows and specular highlights [1].

Another useful object of interest is the visible shell, S_t, which is the set of all voxels that are directly visible according to the C cameras. The algorithm for computing the shell (shown in figure 2) at time step t is

Initialize the set S_t to empty
Compute voxel person, i.e. $V'_t = \wedge_{c=1}^{C} V_t^c$
For each camera $(1 \le c \le C)$
  For each pixel in the foreground set
    Find the closest voxel, $v_{(i,j,l)}$, from $L_{c,(n,m)}$ in V'_t
    Add $v_{(i,j,l)}$ to S_t.



(a) (b)

Fig. 2. (a) Shell (shown in green) and (b) shell along with the rest of voxel object (shown in blue) for a two camera setup.

3 Fuzzy Voxel Object

The quality of voxel person construction varies over the space depending on the object's position and the location of the cameras. As a result, information gathered from the crisp model can be inaccurate. The crisp voxel object is a good real-time initial technique to produce three-dimensional models for tracking. The technique works quite well when there are several cameras viewing a space with overlapping fields of view. Unfortunately, it is rare to have more than two cameras in a given area, due to factors such as cost, processing time of subsequent algorithms, data transmission and storage, and in the case of eldercare, limited space and installation locations, for the seamless integration of a passive video sensor network into the home.

Fuzzy voxel person is an attempt to model and use the different types of uncertainties related to the construction of the object. Each voxel is assigned a membership value that reflects how much it belongs to the actual object. The two types of uncertainty identified and fused in this paper are (a) how reasonable it is to infer that a voxel is part of the object based on the shell and (b) where a voxel is located relative to the distribution of mass across the object.

The first measure considers the minimum distance one must "step" in voxel space until a voxel on the shell is reached. More importance is placed on voxels near the shell. However, this measure has a natural tendency to favor the shell and pull confidence away from the central mass of the object. The second measure determines how confident one is in a voxel, based on how close it is to the central mass. This measure obviously favors the central mass and in return pulls confidence away from the shell. When these two sources of information are combined, the final value is a measure of how dense the region is around the voxel and how reasonable it is to infer that location given its distance to the observed shell. This combined confidence is low for voxels in the tail of the object (the furthest part from the shell), high in the dense close to the shell regions, and relatively higher (than the tail) for voxels in non-dense volumes close to the shell.

The first value is inversely related to a voxel's distance from the shell. The greatest certainty of object intersection occurs at the shell and decreases outwardly. The distance value is found using mathematical morphology, specifically dilation, $\oplus$. The kernel used is K, and $V'_t \oplus K$ is the dilation of the model V'_t by kernel K. A K of 3x3x3 of all ones is used (i.e. the immediate 26 adjacent voxel neighborhood). The distance value is computed quickly by repeatedly dilating the voxels in the shell, then subtracting that set from the set of remaining voxels in the intersected object. All surviving voxels in the intersected set have a related value, $m_{(i,j,l)}$ for $v_{(i,j,l)}$, that is incremented each iteration it survives. Formally, the algorithm is

Initialize all $m_{(i,j,l)}$ to 0
$V' = V'_t$
$S' = S_t$
while $|V'| > 0$
  for each $v'_{(i,j,l)} \in V'$
    add one to $m_{(i,j,l)}$
  end
  $V' = V' - S'$
  $S' = S' \oplus K$
end

The confidence for each voxel is

$$m'_{(i,j,l)} = 1 - \left(\frac{m_{(i,j,l)}}{\max_{m_{(a,b,c)}} m_{(a,b,c)}} \right). \quad (2)$$

The density value is computed using erosion, Θ. Again, a K of 3x3x3 of all ones is used. The voxel object is continually eroded until no voxels remain in the set. For each erosion step that a voxel remains in the set, a separate value, $e_{(i,j,l)}$ for $v_{(i,j,l)}$, is incremented. The algorithm is

ISBN: 978-989-95079-6-8

Initialize all $e_{(i,j,l)}$ to 0
$V' = V'_t$
while $|V'| > 0$
  for each $v'_{(i,j,l)} \in V'$
    add one to $e_{(i,j,l)}$
  end
  $V' = V' \ominus K$
end.

The most direct method for calculating the membership value per voxel is

$$e'_{(i,j,l)} = \frac{e_{(i,j,l)}}{\max_{e_{(a,b,c)}} e_{(a,b,c)}}. \quad (3)$$

However, this assigns low confidence to the object's entire outer shell. Up to this point, no a priori knowledge about the object being tracked has been assumed. For some objects, such as a human with appendages (arms, legs, chest, etc), this calculation might be too drastic, resulting in too little of importance assigned to the appendages (which are less dense than the chest). In order to accommodate such cases, the confidence values can instead be calculated as

$$e'_{(i,j,l)} = (1-\beta)\frac{e_{(i,j,l)}}{\max_{e_{(a,b,c)}} e_{(a,b,c)}} + \beta. \quad (4)$$

This maps the value into [β , 1], where ($0 \le \beta \le 1$) is a free parameter, either user defined or learned. The value β can be thought of as the minimum support one is willing to assign to voxels on, and subsequently near, the shell.

The final value per voxel, $\mu_{(i,j,l)} \in [0,1]$, is found using a t-norm (we use the product),

$$\mu_{(i,j,l)} = m'_{(i,j,l)} \wedge e'_{(i,j,l)}. \quad (5)$$

4 Object Construction Quality

As alluded to earlier, the quality of construction varies over the space with respect to the object's location, the installation locations of the cameras, and their respective orientations. Quantitatively representing the construction quality is of importance for two reasons. First, the value informs us about the reliability of the lower level results. For example, knowledge about the ability to properly construct the human affects the features calculated, and subsequently our decisions regarding activity inferred using those features. Secondly, there may be many locations in a space for which it is impossible to properly construct a voxel object given a particular camera configuration. In many instances, suppression of any object construction in those particular areas is of value.

The best construction generally occurs when the intersecting view vectors are orthogonal. This means that each individual voxel has a different construction quality. The quality of construction for voxel $v_{(i,j,l)}$ is

$$Q_{(i,j,l)} = 1 - \max_{\substack{C_g, C_h \\ g \neq h}} \left| (v_{(i,j,l)} - C_g)^T (v_{(i,j,l)} - C_h) \right|, \quad (6)$$

where C_g and C_h are the center locations for two cameras. Thus, the best case per voxel per camera pair with respect to orthogonality is considered. If the intrinsic camera parameters have been estimated, for example using [13], then the pixel ray in camera c that intersects voxel $v_{(i,j,l)}$ can be used instead. The environment is again converted into a voxel representation, but this time the resolution is lower. We empirically determined that a sampling of once every foot was a good resolution for our application. Figure 3 shows 9 horizontal slices (x-y planes for a varying z) of the sampled voxel space and their respective qualities.

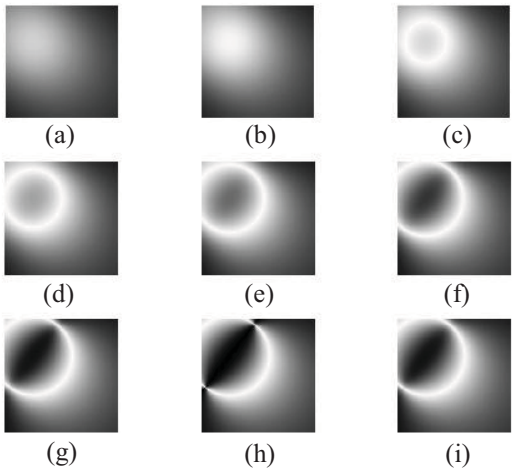


Fig. 3. Nine horizontal (x-y plane) slices of voxel object construction quality for the camera configuration {(a),(b)} shown in figure 4. Brighter values represent higher quality. Map (a) above is at a height of 1 foot, and each consecutive map is 1 foot higher in the z dimension (world up direction).

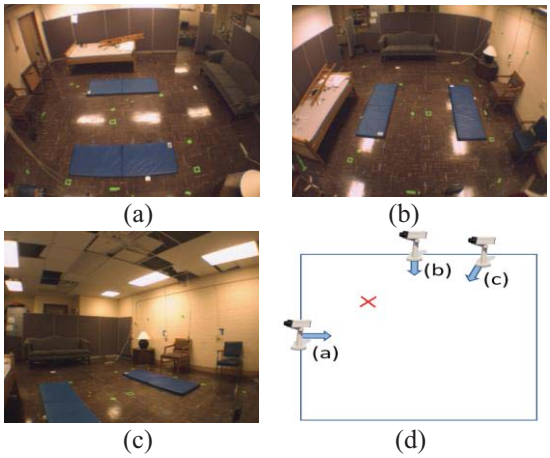


Fig. 4. Camera installation locations (d) and images showing their respective views of the monitored space (a-c).

The next matter is the determination of the quality of construction for an object. One can compute the mentioned

ISBN: 978-989-95079-6-8

quality measure for each voxel and make a decision based on the set of all memberships, or the object can be summarized, according to its centroid, height, or some other measurement, and that point estimate can be used to make a decision for the entire object. We perform the latter. We convert the monitored space into a low resolution voxel configuration (such as demonstrated in figure 3), then the height domain is collapsed and a single voxel plane in the x-y dimension is produced. The voxel qualities are combined using a t-norm (we used the min),

$$Q_{(i,j)} = \min_k Q_{(i,j,k)}. \qquad (7)$$

This pessimistically selects the worst construction case per element in the final x-y plane voxel set. Figure 5 shows the final x-y voxel plane of combined confidences for figure 3.

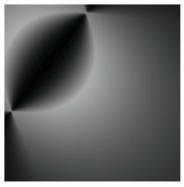
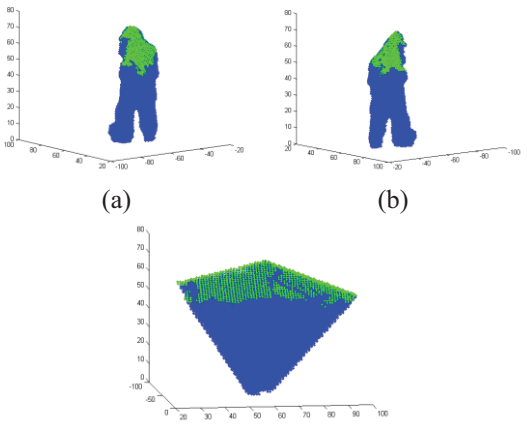


Fig. 5. T-norm produced x-y voxel plane of combined object construction confidences for the nine slices in figure 3.

The voxel object's centroid is mapped to the closest (Euclidean) voxel in the quality map (according to its x-y plane distance), and the respective quality value, ϕ_t, is retrieved. If $\phi_t < \delta_1$ (we empirically determined $\delta_1 = 0.2$), then the object is not constructed and ultimately not tracked. The parameter δ_1 is not specific to any one single tracking case, but it is rather a general parameter related to ray-based back-projection object construction. Figure 6 shows an example construction in which the quality value is 0.05.



(a) (b)

(c)

Fig. 6. Example low quality construction (value of 0.05) using camera configuration {(a),(b)} (the human's location is marked by an X in figure 4). The object appears to be constructed correctly when viewed independently with respect to the two cameras, images (a) and (b). However, its actual shape from another position (c) in three-dimensional space shows that the object was not built correctly.

5 Fuzzy Feature Extraction

After fuzzy voxel person is created, and its quality is assessed and determined to be adequate, one of two possible decisions can be made. The object can remain fuzzy and fuzzy feature extraction can be performed for activity analysis, or the fuzzy object can be used to acquire a better crisp voxel person and standard crisp feature extraction can be performed. This section details the prior.

The fuzzy feature selected for analysis in this paper is the centroid, which is most often used for determining object interaction, velocity, direction, position, and more generally, inferring the overall state of the tracked person (standing, kneeling, on the ground, etc) [1]. The fuzzy centroid is

$$M_{fuzzy} = \frac{\sum_{\dot{v}_{(i,j,l)} \in V'_t} \mu_{(i,j,l)} \dot{v}_{(i,j,l)}}{\sum_{\dot{v}_{(i,j,l)} \in V'_t} \mu_{(i,j,l)}}. \qquad (8)$$

6 Crisp Voxel Object Improvement

An alternative to directly using fuzzy voxel person for tracking is if the volume of the object is known ahead of time, or if it can be sufficiently approximated online, fuzzy voxel person can be used to acquire an improved crisp voxel object. Once the crisp object is obtained, standard crisp tracking and feature extraction can be used. The concept here is that fuzzy voxel person is useful for obtaining the global relative set of memberships, and a search for a better crisp object is guided by the evidence contained in fuzzy voxel person. By using the known or approximated volume, P_{real}, the highest confidence areas can be identified and kept. One possibility is to look for an alpha cut such that the crisp volume (cardinality) of the resultant set is closest to P_{real}. Formally, the search is for an α such that

$$\arg\min_{0 \le \alpha \le 1} \left| \left| {}^{\alpha}V' \right| - P_{real} \right|, \qquad (9)$$

where ${}^{\alpha}V'$ is

$${}^{\alpha}V' = \left\{ \dot{v}_{(i,j,l)} \mid \mu_{(i,j,l)} \ge \alpha \right\}. \qquad (10)$$

One can easily solve this by sorting all voxel memberships in descending order and then selecting the first P_{real} voxels (in the case of a tie, include all ties).

In the case that P_{real} must be approximated, the procedure is as follows. For a user defined time window, T frames, compute crisp voxel person and its associated quality value at each time step. Find the maximum ϕ_t value for the T time steps. If this value is above some threshold δ_2, again this is a global parameter not specific to any one camera configuration, use the crisp voxel-based volume to approximate P_{real}. Figure 7 shows a human built using an approximated P_{real}.

ISBN: 978-989-95079-6-8

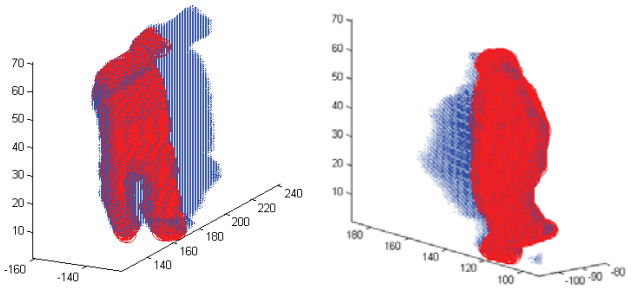


Fig. 7. The proposed alpha cut-based procedure for fuzzy voxel person to obtain an improved crisp object. Red areas are the improved voxel person and the blue areas are the rest of the original crisp voxel person.

In order to quantitatively demonstrate the improvement in the general shape and object's orientation, the following metrics for tracking humans are presented. The voxel covariance matrix is calculated, and its major orientation, the eigenvector with the largest respective eigenvalue, is compared to the world up vector ($<0,0,1>$). For a human standing and for a file cabinet used in the results section the major orientation should be very similar to the world up vector. Next, the difference, according to the variation associated with each eigenvector, between the new and real object is compared. These values indicate how the mass was distributed across the object according to its primary orthonormal basis with respect to variation. The eigen-information is extracted from the covariance matrix of voxel person, and the eigenvalues (φ) at 3 standard deviation are calculated, $\omega = \sqrt{\varphi} * 3$. The eigenvalues are sorted in decreasing order, and the difference between the real object (ω_{real}) and the crisp voxel (ω_{voxel}) is computed, $\omega' = |\omega_{real} - \omega_{voxel}|$. The feature used in the results section is the sum of the three components in ω', denoted by ω_{diff}.

7 Experiments and Quantitative Results

While figure 7 qualitatively demonstrated the benefits of the techniques presented in this paper, comparison between the (a) fuzzy and crisp centroids, (b) eigen-based analysis of the voxel spread (variance), and (c) comparison between the expected and the observed major orientations using eigen-analysis is performed. The subject in this case is a file cabinet that was placed in unique locations and orientations throughout the scene. A file cabinet has easy-to-recognize shape, orientation, known volume, and centroid.

For the file cabinet and the density-based measure, a value of $\beta = 0.5$ is used. Combinations of camera pairs, {(a),(b)} and {(b),(c)} were used. Camera combination {(a),(b)} is an ideal situation, given the near orthogonal installation, while {(b),(c)} is less than ideal. The later configuration results in much larger inaccuracies due to the trailing tail of intersected voxels. Tables 1-6 numerically display the accuracy of the fuzzy-based approaches, figures 8 and 9 show top-down view of the file cabinet at several locations, and figure 10 shows a few perspective views of the constructed object to give a better understanding of camera view induced warping and the resultant object shape.

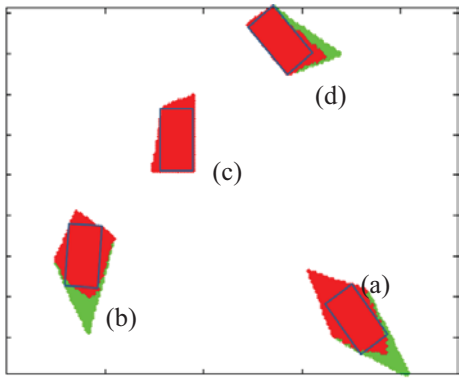


Fig. 8. Crisp (red and green areas) and alpha-cut based crisp objects (red areas) for camera configurations {(a),(b)}. The real object's approximate shape and orientation are overlaid using a rectangular outline.

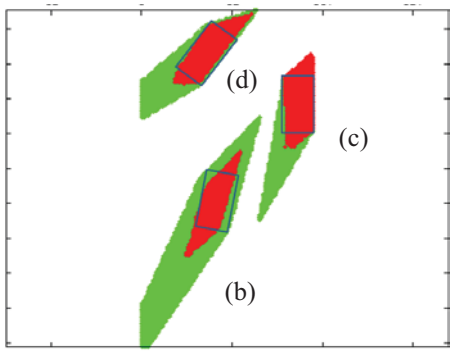


Fig. 9. Crisp (red and green areas) and alpha-cut based crisp objects (red areas) for camera configurations {(b),(c)}. The real object's approximate shape and orientation are overlaid using a rectangular outline.

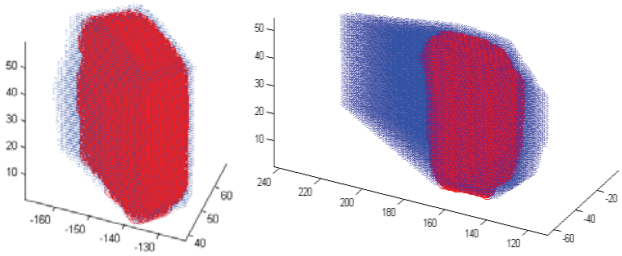


Fig. 10. Improved construction. Red areas are the improved voxel object and the blue areas are the rest of the original crisp voxel object. The left image is for (d) in figure 8, i.e. cameras {(a),(b)}, and the right image is for (b) in figure 9, an extreme case, using cameras {(b),(c)}.

Table 1: Difference between real and computed object centroid. Camera configuration {(a),(b)}. Units are specified in feet. Smaller numbers are better.

	a	b	c	d
Crisp	0.46	0.51	0.49	0.21
Fuzzy	0.19	0.29	0.47	0.27

ISBN: 978-989-95079-6-8

Table 2: Difference between real and computed object centroid. Camera configuration {(b),(c)}. Units are specified in feet. Smaller numbers are better.

	b	c	d
Crisp	2.24	1.18	1.13
Fuzzy	1.47	0.56	0.79

Table 3: Camera configuration {(a),(b)}. Value of ω_{diff}. Smaller numbers are better.

	a	b	c	d
Crisp	87.79	58.60	16.72	42.78
α -cut	4.60	1.72	0.82	0.07

Table 4: Camera configuration {(b),(c)}. Value of ω_{diff}. Smaller numbers are better.

	b	c	d
Crisp	3776.99	411.02	658.13
α -cut	807.52	55.71	17.11

Table 5: Camera configuration {(a),(b)}. Values are the dot product between the object's major orientation (eigenvector with the largest eigenvalue) with the world up vector. Larger numbers are better.

	a	b	c	d
Crisp	0.91	0.96	1.00	0.99
α -cut	1.00	1.00	1.00	1.00

Table 6: Camera configuration {(b),(c)}. Values are the dot product between the object's major orientation (eigenvector with the largest eigenvalue) with the world up vector. Larger numbers are better.

	b	c	d
Crisp	0.15	0.32	0.31
α -cut	0.93	0.96	0.79

The fuzzy centroid is more accurate than the crisp centroid in every case for configuration {(b),(c)} (i.e. extreme viewing condition). For configuration {(a),(b)}, very ideal condition, the centroids are very close. Fuzzy is better overall, however, in one instance the fuzzy centroid is slightly worse (Table 2.d). In that situation, camera (b) was looking at the object from a top down view, and the visible shell resulted in a greater vertical (z direction) pull. Tables 3 and 4 show that the objects shape (according to ω_{diff}) is closer to the real object. Lastly, tables 5 and 6 show that the objects major direction is always equal or more accurate (significant in Table 6) to the world up direction.

8 Conclusion and Future Work

In this paper, a fuzzy-based computer vision approach to the robust construction of three-dimensional objects for human activity analysis using only a few cameras and minimal a priori knowledge of the object is introduced. Extreme joint viewing conditions were considered and the fuzzy acquired results are better than the crisp counterpart in both the quantitative and qualitative regard.

The shell is currently approximated using the visible parts of the voxel object, created using back-projection. As the figures illustrate, this results in some warping in the objects shape. Stereo vision will be used to improve the estimation of the visible shell from depth maps (which will also require the t-norm operator in (5) to be revisited). Additionally, the use of a priori knowledge about the object being tracked will be used, for example, attempting region segmentation, and using that information for improved membership calculation.

Acknowledgment

Robert Luke and Derek Anderson are pre-doctoral biomedical informatics research fellows funded by the National Library of Medicine (T15 LM07089). This work is also supported by the National Science Foundation (ITR award llS-0428420) and the Administration on Aging (90AM3013).

References

[1] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, M Aud, "Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic," *Comp. Vision and Image Understanding*, doi:10.1016/j.cviu.2008.07.006, Vol. 113, pp. 80-89, 2008.

[2] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, M Aud, "Modeling Human Activity From Voxel Person Using Fuzzy Logic," *IEEE Transactions on Fuzzy Systems*, Accepted July 20th 2008.

[3] C. Stauffer, W.E.L. Grimson, Adaptive Background mixture Models for Real-time Tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2), 243-252, 1999

[4] N. Oliver, B. Rosario, A. Pentland, A bayesian computer vision system for modeling human interactions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8), 831-843, 2000

[5] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: Principles and Practice of Background Maintenance. *In Proceedings of ICCV'1999*, 255-261 1999

[6] B. G. Baumgart, Geometric Modeling for Computer Vision. Technical Report AIM-249, Artificial Intelligence Laboratory, Stanford University, 1974.

[7] B. C. Vemuri and J. K. Aggarwal, "3-D model construction from multiple views using range and intensity data," in *Proceedings of IEEE Conf. Computer Vision and Pattern Recognition*, pp. 435-437, 1986.

[8] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, pp. 150-162, 1994.

[9] G. Dudek and D. Daum, "On 3-D surface reconstruction using shape from shadows," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 461-468, 1998.

[10] M. Pardas and J. Landabaso, "Foreground regions extraction and characterization towards real-time object tracking," in *Proc. of Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Lecture Notes in Computer Science. Springer, 2005.

[11] G. Cheung, T. Kanade, J. Bouguet, and M. Holler, "A Real Time System for Robust 3D Voxel Reconstruction of Human Motions", in *Proc. of Computer Vision and Pattern Recognition*, Vol. 2, pp. 714-720, 2000.

[12] N. Thome, D. Merad, and S. Miguet, "Human Body Part Labeling and Tracking Using Graph Matching Theory", in *Proc. of IEEE Intl. Conf. on Video and Signal Based Surveillance*, 2006.

[13] Jean-Yves Bouguet, "Camera Calibration Toolbox for Matlab," Intel Corp, 2004, http://www.vision.caltech.edu/bouguetj/calib_doc/.

ISBN: 978-989-95079-6-8

Fuzzy Logic-Based Image Processing Using Graphics Processor Units

R. H. Luke, D. T. Anderson, J. M. Keller
Department of Computer Engineering
University of Missouri
Columbia, Missouri, USA
{rhl3db, dtaxtd}@mizzou.edu, kellerj@missouri.edu

S. Coupland
Centre for Computational Intelligence
De Montfort University
Leicester, LE1 9BH, UK
simonc@dmu.ac.uk

Abstract—**This paper introduces a parallelization of fuzzy logic-based image processing using Graphics Processor Units (GPUs). Using an NVIDIA 8800 Ultra, a 126 time speed improvement can be made to fuzzy edge extraction making its processing real-time. The GPU can process approximately 42 frames per second at 640x480 image resolution, thus 307,200 inference processes per frame. With a computational speed improvement of over two orders of magnitude, more time can be allocated to higher level computer vision algorithms. This GPU solution is described using NVIDIAs Compute Unified Device Architecture (CUDA).**

Keywords- ***Graphics Processor Units, Fuzzy Image Processing, Fuzzy Edge Extraction, Compute Unified Device Architecture***

I. INTRODUCTION

Edge extraction is a low level technique used in image processing and computer vision. There are many algorithms to perform edge extraction with a wide range of computational complexities and performance. Common approaches include the Sobel [1] and Prewitt [2] convolution kernels and the Canny Edge Detector [3]. Edges are most often used as low level information fed to a higher level process such as segmentation or recognition.

Some more recent algorithms have merged the strengths of image processing and fuzzy logic. These algorithms span a wide range of applications such as image enhancement [4], edge extraction [5], segmentation [6], and contrast adjustment [7]. An edge detection model based on Fuzzy Logic, called Fuzzy Inference Ruled by Else-Action (FIRE), was designed by Russo and Ramponi in [8]. This algorithm checks eight unique edge cases and outputs a value related to the confidence of the best matched case. The significance of this specific edge detection algorithm is its robustness to noise.

Even with simple algorithms, there is a large amount of work required to perform edge detection on moderately sized images. The workload is further compounded for real-time systems processing multiple frames per-second. A stream processing solution based on Graphics Processor Units (GPUs) and NVIDIAs Compute Unified Device Architecture (CUDA) is presented. While the work in this paper focuses on edge detection as the parallel execution of a single Fuzzy Inference System (FIS) for multiple different inputs, the GPU solution can be easily modified to fit a wide range of batch-processing tasks that utilize the standard Mandani-type FIS [9][10][11]. The system's extension is discussed in a later section.

It should be noted that we could have implemented any fuzzy logic-based image processing technique. We chose to demonstrate the proposed GPU-based architecture using FIRE because it is an easily understood algorithm. Our goal is not to show the performance of FIRE, but rather to present a framework for the parallel execution of fuzzy logic based image processing.

Several hardware accelerated FIS solutions have been developed to improve processing speed. For example, FPGA [12][13], and VLSI [14] solutions exist, however the significant portion of these are related to Fuzzy Control, not Image Processing. These specialized solutions are generally expensive and time consuming to modify and adapt to a new problem. In contrast, GPUs are relatively inexpensive, easily integrated with PCs and flexible to general purpose programming.

Using GPUs for general purpose computing is not a new idea. Well-known graphics API-based approaches include: Fast Fourier Transform (FFT) [15], linear algebra operators [16], protein folding [17], Fuzzy C-Means [18][19], and the execution of a single FIS [20]. Researchers have also begun to recently explore CUDA in Image Processing, such as: optical flow estimation [21], segmentation of medical images [22], and the Canny Edge Detector [23].

II. EDGE EXTRACTION USING FIRE

An input image I has luminance values from $\{0,1,\ldots,255\}$. For pixel P_{ij}, let $W_{ij} = \{P_{(i-e)(j-f)},\ldots,P_{(i+e)(j+f)}\}$, where $e,f \in \lfloor 0,\ldots,K/2 \rfloor$ and K is the window size, be the set of pixels in the neighborhood of P_{ij}, with the exception of P_{ij} (hence, $|W_{ij}| = K*K-1$). The inputs to the fuzzy edge extractor are the luminance differences between pixels in W_{ij} to P_{ij},

$$X_{(ij,m)} = W_{(ij,m)} - P_{ij}, \quad (1)$$

$$1 \le m \le (K*K-1),$$

ISBN: 978-989-95079-6-8

where $W_{(ij,m)}$ is the m[th] element in the window centered at P_{ij}. All subsequent operations performed on $X_{(ij,m)}$ are independent of all other pixels, therefore, $X_{(ij,m)}$ will be notationally reduced to X_m, to simplify the indexing and as all discussion is about a specific P_{ij} and its W_{ij}. The domain of X_m is the set $\{-255,\ldots,255\}$.

As stated previously, FIRE is a case-based system used to test eight unique edge orientations on a window. The size of the window used in this system is 3x3. Therefore, X has a cardinality, $|X|$, of 8.

An overly simplified rule is: *If a pixel lies upon an edge Then make it black Else make it white.* Eight unique cases, representing an edge at every 45 degrees of rotation, figure 1, is represented as a rule. Each rule has six antecedents, coinciding with six of the eight pixels in X. Fuzzy sets are used to represent the linguistic terms negative and positive. Similarly, the linguistic uncertainties of the consequent terms white and black are also modeled as fuzzy sets.

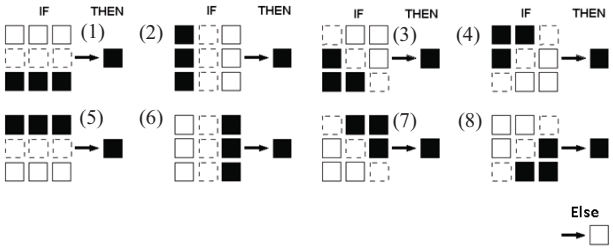


Figure 1. A graphical representation of the eight rules in the fuzzy rule base. Each rule has six antecedents represented as negative (black boxes) and positive (white boxes) luminance differences to the center pixel. The consequent is the same for each rule, a black pixel. A single else case handles the situation of none of the antecedents being matched.

For each pixel, each of the eight rules consisting of six antecedents are fired using the values in X. For each rule, the average of the six confidences is used as the total antecedent firing, f_i, where $1 \le i \le 8$. In this algorithm, it is assumed that a given pixel can belong to, at most, one of the eight cases. Therefore, the rule with the largest firing f_l is used as the antecedent firing strength to the fuzzy consequent set representing black.

$$f_l = \max(f_1,\ldots,f_8) \quad (2)$$

This is a departure from the more common Mamdani FIS which uses all rules in its decision making. The implication of the white fuzzy set uses $f_s = 1 - f_l$, i.e. the complement of the antecedent. The centroid of the aggregation of implication using f_l and f_s is the output of the algorithm. It should be noted that the result of the algorithm, $\{P'_{11},\ldots,P'_{ij},\ldots\}$, is in [0, 127.5], not [0, 255]. This is due to the setup of the antecedent fuzzy sets. In the extreme case, if all pixels in the 3x3 window are the same intensity, (no edge), all antecedent sets for all rules are fired with a 0.5 membership. The output of equation 2 will be 0.5 and therefore, f_s will also be 0.5. The consequents will have the same mass, and the final result will be 127.5. The domain of P'_{ij} is discretized and its range expanded to $\{0,1,\ldots,255\}$ by multiplying a constant, C, by P'_{ij}. In this paper, $C = 2$. Figure 2 displays an input and output image using this system.



Figure 2: Example output image of the algorithm described in [8].

III. CUDA INTRODUCTION

Because the calculation of the edge confidence at each pixel is independent of all other pixel calculations, this algorithm is a good candidate for parallelization. A single FIS is run for each pixel in an image. For a 640x480 image this results in the execution of a single FIS for each of the 307,200 pixels. Parallelization and implementation of the FIRE system to take advantage of the large number of co-processors on a GPU greatly increases the speed over sequentially processing the elements on the CPU.

NVIDIA defines CUDA as "a hardware and software architecture for issuing and managing computations on the GPU as a data-parallel computing device" [24]. The GPU is a highly parallel general purpose co-processor, not just a graphics processor. The CUDA API extends the C language and allows the largest portion of the programming community a quick transition to its use.

CUDA allows multiple programs, kernels, to run sequentially on a single GPU. CUDA organizes a kernel into a grid. A grid is subdivided into blocks. All blocks run the same kernel, but each runs independently from all others. Each block is made up of threads, the smallest divisible unit on the GPU. The actual work of the kernel is performed at the thread level.

The hierarchical structure of CUDA provides a mechanism to take advantage of the underlying hardware structure. This is mainly due to the grouping of processors into multiprocessors and that the memory types are unique to the grid, blocks and threads. Each multiprocessor is comprised of Q processors, Q=8 for the NVIDIA 8800, and each block is loaded into a single multiprocessor. Each multiprocessor is a Single Instruction Multiple Data (SIMD) set, meaning that multiple elements, Q, will be processed in parallel using the same instruction but different data. Threads in a block are partitioned into warps, where a warp size in the NVIDIA 8800 is 32. All threads in a warp are executed using the same

ISBN: 978-989-95079-6-8

instruction. Conditional logic can cause warp divergence and result in performance degredation. Code must be designed to take the warp size and control flow into account.

There are several types of memory in CUDA available to the programmer. Using the proper memory is vital to the efficiency of a program. Therefore, developers must create algorithms with memory access in mind.

The Global memory type can be read from and written to by any thread. Global memory access is the slowest of all types and should be used sparingly. The Constant and Texture memories can be read from any thread in a kernel, but not written to. The access times of texture memory are substantially smaller than Global memory.

Shared memory can be read from and written to by any thread associated with the block. This memory is local to the specific block and cannot be accessed by threads of any other block. Memory access times are shorter than the types mentioned above. Values from Global memory are often brought into shared memory at the beginning of a kernel, all threads are synchronized, and future repetitive access to this data is very efficient.

At the thread level, there is Local Memory and Registers which have the smallest access times. These memory locations are only accessible from a single thread. The entire memory layout is shown in figure 3.

IV. CUDA FIRE IMPLEMENTATION

This GPU implementation of the FIRE algorithm consists of two kernels. The first kernel performs all pixel neighborhood differencing, fuzzification for all rules, computes the average of the antecedent firings for each rule, and finds the maximum rule firing. The second kernel uses the minimum operator for implication, a maximum operator for rule aggregation and two sum reductions [25] to determine the centroid of the aggregated consequent sets for defuzzification. Though a simpler type of defuzzification could have been performed, we elected to use the standard general centroid defuzzification so that readers could easily extend this work to other more complex fuzzy-logic problems.

Even though more parallelism could be performed at each step, this two kernel approach is faster for three reasons. First, if the first kernel is subdivided into more parallel stages, each stage does too little computation relative to the amount of memory access. The GPU is most efficient when there is a large arithmetic to memory access ratio. Secondly, the second stage, implication, takes advantage of shared memory, making it more efficient to be a single kernel. Finally, most fuzzy systems use a small number of rules, less than 32, which does not justify further kernel subdivision.

The first kernel divides the input image into 8x8 non-overlapping pixel blocks. Each image block is processed in a CUDA block. This is done to make the processing of pixels faster by loading blocks of pixel data into shared memory for threads in a CUDA block to work on. As described earlier, the algorithm works on 3x3 windows of data for each pixel. In order to perform these operations on the pixels of the boundary of the 8x8 image blocks, a one pixel border must also be brought into memory. So, a 10x10 window of pixels is brought from texture memory into a block's shared.

Each thread can now perform the operations of the algorithm. An array of length eight, local to each thread, is defined and filled with the pixel neighborhood difference values (X_m , i.e. equation 1). The six antecedents of the eight rules are fired using the values in X . The antecedent firings are then summed across each rule. The max of these eight values is found and divided by six to find the average. This value is then written to the output of this kernel.

The output of the first kernel, f_1 , is the average of the maximum fired antecedent for each pixel of the input image and is in [0, 1]. For the second kernel, a separate CUDA block is created for each f_1 . Each of the T threads in a block performs inference aggregation, and is used to find the centroid of the aggregated consequent sets. Each consequent set is discretized into 256 elements. We empirically picked a T of 128, however, the selection of T is related to reduction, which is detailed later.

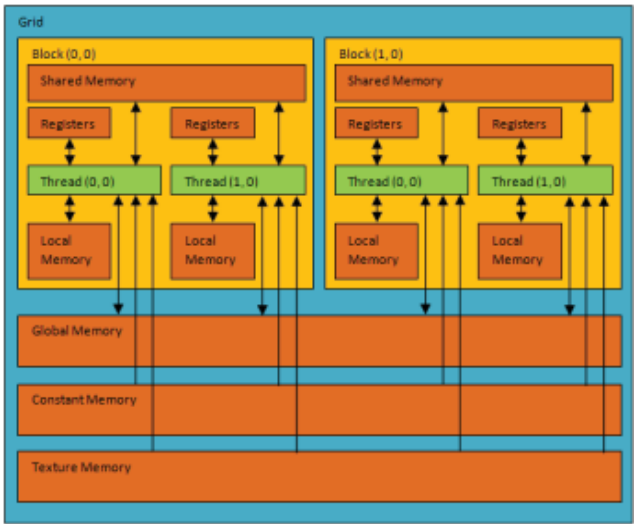


Figure 3: Processing organization scheme and memory layout for the GPU using CUDA. Global, Constant and Texture Memories are accessible by all threads. Shared Memory has is faster, but is local to each Block. Registers and Local Memory are the fastest, but are local to each thread.

Each thread performs inference (min) of f_1 with its corresponding discrete consequent domain membership value in the fuzzy set "black". The minimum of f_s and the appropriate discrete consequent domain membership value in the fuzzy set "white" is calculated, and each thread stores their aggregation (max) value in a shared memory array. Each thread is assigned an identifier in CUDA and this identifier is what is used to select the index into the discrete consequent domain fuzzy sets and the shared memory region. For a discrete consequent domain of size 256, a T of 128 was selected, where T was empirically determined. This means each thread calculates two discrete consequent domain elements instead of one. The reason for doing this is related to reduction, which is used for calculating the centroid.

ISBN: 978-989-95079-6-8

Reduction is the next step and there are many guidelines to follow, [25]. One rule states that it is generally better to calculate a few loads and reduction steps at the beginning versus allocating more threads, which become unused as reduction steps proceed. Other guidelines include sequential addressing and unrolling. Each block calculates the numerator and denominator separately and in parallel. The kernel must now compute the sum of the numerator and denominator. Reduction [25], which is the collapsing of a set of elements by some operator into one scalar, (such as the average, sum or max), is a common GPU operation. This procedure usually takes $\log_2(N)$ time, where N is the number of elements to reduce. For large values of N, such as N=1,000,000, a multiple block reduction would be used. However for 256 elements, a single block method keeping all in shared memory is the most efficient. An example of sum reduction in shown in figure 4.

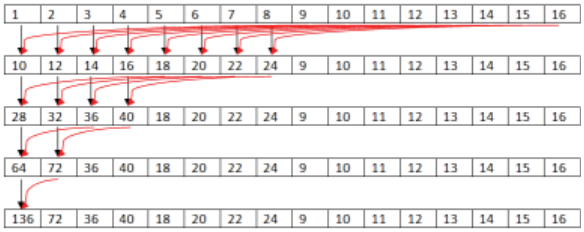


Figure 4: An example of sum reduction. In the first row, eight threads sample two locations in memory, add their values and output the result to a specified location. In the next row, four threads perform the operations. Then, two and one in the final two rows. The total sum is in the first index of the array in the bottom row.

It should be noted that GPU reduction is faster than a single CPU, however, not all processing cores can be used efficiently. This is because an increasing number of processors go unused with each reduction step. For example, reducing 256 elements with 128 threads will reduce to 128 elements in the second step. However, in the second step, only 64 of the threads are involved in reduction. In the third step only 32 are performing work, 16 in the next, and so on.

The final operation of the kernel is the calculation of the centroid. As previously discussed, this value is in [0, 127.5] and is multiplied by 2 (C). It is then discretized so that the final values of the output image, P'_{ij}, are in $\{0,1,\ldots,255\}$. The entire GPU solution is displayed in figure 5.

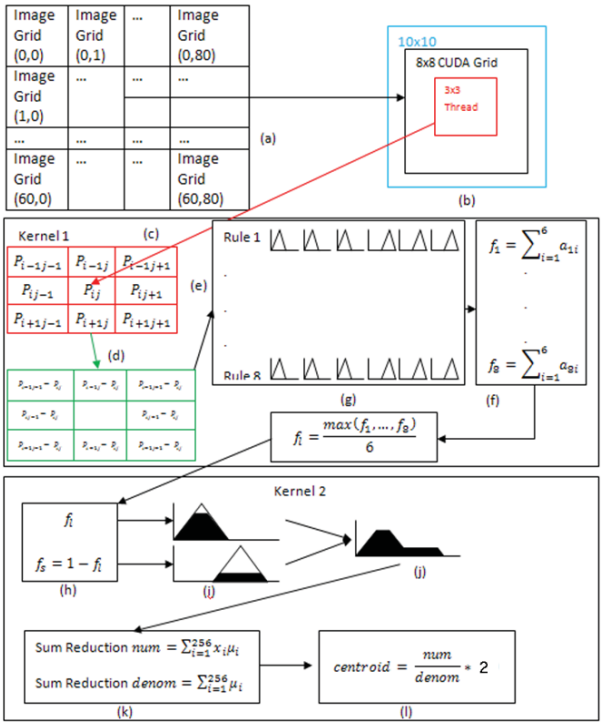


Figure 5: A graphical representation of the system defined in this paper. (a) The 640x480 input image divided into 8x8 blocks. (b) A single CUDA block with a 10x10 shared memory. (c) 3x3 input to a thread. (d) The difference values computed from (c). (e) Antecedent firing over eight rules. (f) Find the average of each rule. (g) Output the largest antecedent firing for kernel 1. (h) Compute else firing. (i) Perform implication on the consequent sets. (j) Aggregate consequent sets. (k) Sum reduction of the numerator and denominator values for centroid computation. (l) Compute the centroid of the aggregated consequents and scale to image output.

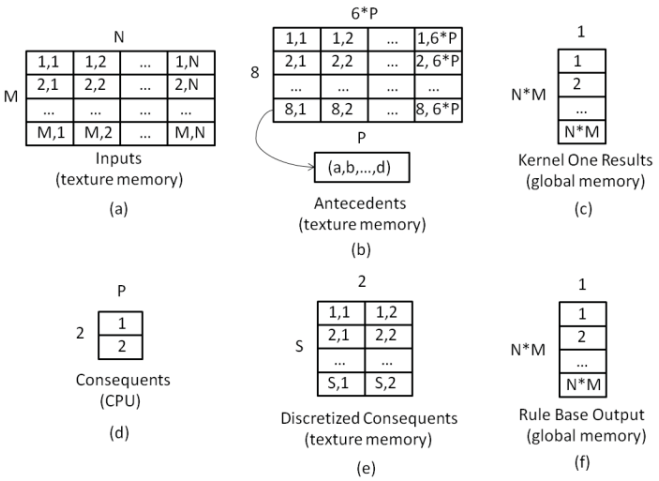


Figure 6: CPU memory and GPU global and texture memory organization for the GPU FIRE solution. For this paper, M = 640, N = 480 and S = 256, however, M, N and S can take on different profiles. (a) The input image in texture memory, (b) antecedents in texture memory, (c) global memory is used to store the output of the first kernel, (d) CPU side parameters of the two consequent sets, (e) GPU texture memory for the discretized consequent sets, and (f) global memory is used to store the output of the second kernel (image of edge confidences).

ISBN: 978-989-95079-6-8

As mentioned above, memory usage and organization is vital to CUDA efficiency. The input image is moved from the CPU to GPU texture memory. In this case, the image size is 640x480. Texture memory is used instead of the slower global memory because the image is read-only on the GPU. Secondly, the antecedent sets for the rule base are stored in texture memory. There are 8 rules with 6 antecedents per rule, and each antecedent set is made up of 4 ordered points, {a,b,c,d} for a trapezoid. Because the rules are read-only, texture memory is used. The output of the first kernel is stored in a one dimensional global memory segment. Next, the discretized consequent sets are transferred from the CPU memory to GPU texture memory. Finally, the output of kernel 2 is stored in a one dimensional global memory segment. The memory organization is displayed in figure 6.

RESULTS

The CPU used for testing was an AMD Athlon 64 FX-55 running at 2.6 GHz. The system had 2GB of RAM and the operating environment was Windows XP. The GPU was an NVIDIA 8800 BFG Ultra with 768 MB of memory [26]. This GPU has 128 processors and connects to the motherboard using PCI express 16X .

The CPU and GPU implementations were then run over a series of different image sizes. Table 1 shows the setups and the associated timings

TABLE I. PROCESSING TIMES AND SPEED INCREASE FROM CPU TO GPU

	320x240 Image 76,800 Pixels	640x480 Image 307,200 Pixels	960x720 Image 691,200 Pixels	1280x960 Image 1,228,800 Pixels	1280x1024 Image 1,310,720 Pixels
CPU	0.73	2.86	6.6	11.5	12.3
GPU	0.006	0.024	0.052	0.091	0.098
Speed Increase	121X	119X	126X	126X	125X



Figure 7: Images processed using the algorithm defined in this paper.

Two output images are shown in figure 7. These images are consistent with the output images from the original paper [8].

As it relates to global memory access, it was found that in the second kernel, the block format and global memory access pattern (index calculation per block) has a large performance impact. For example, two possible block formats for a 640x480 image are (38400, 8) and (8, 38400). A global memory read index (sampling of the f_l value) would then be $b_x * 8 + b_y$ for (38400, 8) and $b_x * 38400 + b_y$ for (8, 38400), where b_x and b_y are the block x and y indices respectively. It was noticed that the (38400, 8) format was three times faster than the (8, 38400) format. In fact, for all profiles (N_{b_x}, N_{b_y}), where N_{b_x} and N_{b_y} are the number of blocks in the x and y dimensions of the grid, it holds that greater speeds are produced if N_{b_y} is a small power of 2 (hence 8, 16, 32), resulting in a smaller calculated offset by $b_x * N_{b_y} + b_y$.

EXTENSION OF SYSTEM

It should be noted that the FIRE system is an unconventional FIS. The Mamdani FIS is the most common model used for Fuzzy Logic. A wide range of Image Processing operations, not just edge extraction, could be performed using a generalized GPU Mamdani FIS implementation.

The implementation defined in this paper requires only a few modifications to perform generalized Mamdani inference. First, instead of finding the mean value of all antecedent firings for a single rule, only the minimum firing would be used. Second, instead of using only the maximum antecedent firing to perform implication on a single consequent set, all fuzzy rules firings would be used. All that is needed is one global memory segment that is $(N * M) \times R$ in size (the FIRE implementation uses an $(N * M) \times 1$ memory segment, fig 6 c), where R is the number of rules. We showed in [27] that fuzzy consequent set aggregation can be efficiently performed as the first step in the reduction kernel, i.e. no need for a separate pass. This is done by storing the discrete consequents in texture memory, a memory segment of size $S \times R$ (the FIRE implementation uses a memory segment of size $S \times 2$, fig 6 e). In the first step of the reduction kernel for implication, aggregation and defuzzification, each thread samples its corresponding discrete consequent values for the R rules. Next, the minimum of each value fetched and their respective rule firings are computed. The final step is to compute the maximum over these minimums. This is a practical assumption, given that the number of rules in a typical rule base is low, e.g., 32 or less, and performing reduction on a set of elements this small is not ideal on the GPU.

CONCLUSION

This paper described a parallelization of fuzzy-logic based image processing on the GPU. The FIRE edge extractor was demonstrated specifically. Using the NVidia 8800 Ultra GPU,

ISBN: 978-989-95079-6-8

a 126 time speed improvement can be made to the original algorithm making its processing real-time. The GPU can process approximately 42 frames per second at 640x480 resolution using this algorithm.

With such large processing power, more time can be spent on higher level processes. The output of this system could be fed into a larger real-time system that performs more complex operations such as object recognition or tracking. The low price of GPUs and the ease of learning and using the CUDA API make this type of parallel programming a legitimate possibility for a large portion of the programming community.

REFERENCES

[1] K. Pringle. "Visual perception by a computer," *In Automatic Interpretation and Classification of Images*, pages 277-284. Academic Press, New York, 1969.

[2] J. Prewitt, "Object enhancement and extraction", in *Picture Processing and Psychopictorics*, Academic Press, pp. 75-149, 1970.

[3] J. Canny, "A computational approach to edge detection," *in IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 679–698, 1986.

[4] Y. S. Choi and R. Krishnapuram, "A robust approach to image enhancement based on fuzzy logic," *in IEEE Trans. Image Process*, 6 (6), pp. 808-825, 1997.

[5] H. R. Tizhoosh, "Fast fuzzy edge detection," in Proceedings NAFIPS 2002 Annual Meeting of the North American Fuzzy Information Processing Society, pp. 239-242, 27-29 June 2002.

[6] L. O. Hall and A. Namasivayam, "Using adaptive fuzzy rules for image segmentation," *in IEEE World Congress on Computational Intelligence*, Vol. 2, pp. 1560-1565, 1998.

[7] M. Craig and M. Schneider, "On the use of fuzzy sets in histogram equalization," *in Fuzzy Sets and Systems*, Vol. 45, pp. 271-278, 1992.

[8] F. Russo and G. Ramponi, "Edge extraction by FIRE operators", *in Proceedings of the Third IEEE International Conference on Fuzzy Systems*, 1994, pp. 249-253.

[9] L. Zadeh, 'Fuzzy sets,' *Information Control*, pp. 338-353, 1965.

[10] L. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes", *in IEEE Trans. on SMC*, Vol. 3, pp. 28-44, 1973.

[11] E. Mamdani, "Application of Fuzzy Algorithms for Simple Dynamic Plant", *in Proc. IEEE*, Vol. 121, pp. 1585-1588, 1974.

[12] M. Cirstea, J. Khor and M. McCormick, "FPGA fuzzy logic controller for variable speed generators," *Proceedings of the 2001 IEEE International Conference on Control Applications*, pp. 2001.

[13] C. Lynch, H. Hagras and V. Callaghan, "Parallel Type-2 Fuzzy Logic Co-Processors for Engine Management," *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International*, 23-26 July 2007 pp.1-6.

[14] Hiroyuki Watanabe, Wayne D. Dettloff, and Kathy E. Yount, "A VLSI Fuzzy Logic Controller with Reconfigurable, Cascadable Architecture," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, 1990, pp. 376-382.

[15] K. Moreland and E. Angel, 'The FFT on a GPU,' in *Proc. SIGGRAPH/EUROGRAPHICS Conf. on Graphics hardware*, 112–119, 2003.

[16] J. Krüger and R. Westermann, "Linear algebra operators for GPU implementation of numerical algorithms," in *Intl. Conf. on Computer Graphics and Interactive Techniques*, 908-916, 2003.

[17] S. Larson, C. Snow, M. Shirts, and V. Pande, 'Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology,' *Comp. Genomics,* 2002.

[18] D. Anderson, R. H. Luke, and J. M. Keller, "Speedup of Fuzzy Clustering Through Stream Processing on Graphics Processing Units", in *IEEE Transactions on Fuzzy Systems*, 2007.

[19] D. Anderson, R. H. Luke, and J. M. Keller, "Incorporation of Non-Euclidean Distance Metrics into Fuzzy Clustering on Graphics Processing Units", in *Proceedings, International Fuzzy Systems Association Conference*, 2007.

[20] N. Harvey, R. H. Luke, J. M. Keller, and D. Anderson, "Speedup of Fuzzy Logic through Stream Processing on Graphics Processing Units", *in IEEE Congress on Evolutionary Computation (CEC), WCCI,* 2008.

[21] Y. Mizukami And K. Tadamura, "Optical Flow Computation on Compute Unified Device Architecture," in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, 10-14 Sept. 2007 pp. 179-184.

[22] L. Pan, L. Gu nad J. Xu, "Implementation of medical image segmentation in CUDA," *Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on*, 30-31 May 2008 pp. 82-85.

[23] Y. Luo andR. Duraiswami, "Canny edge detection on NVIDIA CUDA," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPR Workshops 2008. IEEE Computer Society Conference on*, 23-28 June 2008 pp. 1-8.

[24] Nvidia Corp., 'NVIDIA CUDA Compute Unified Device Architecture,' June 23, 2007, http://developer.download.nvidia.com/compute/cuda/1_0/NVIDIA_CUDA_Programming_Guide_1.0.pdf

[25] M. Harris, "Optimizing Parallel Reduction in CUDA", NVIDIA Whitepaper, March 18, 2008, http://developer.download.nvidia.com/compute/cuda/1_1/Website/projects/reduction/doc/reduction.pdf

[26] Nvidia Corp., 'NVIDIA GeForce 8800 GPU Architecture Overview,' GeForce_8800_GPU_Architecture_Technical_Brief.pdf, November, 2006, http://www.nvidia.com/object/IO_37100.html

[27] D. Anderson and S. Coupland, "Parallelisation of Fuzzy Inference on a Graphics Processor Unit Using the Compute Unified Device Architecture", *UKCI 2008, the 8th Annual Workshop on Computational Intelligence*

ISBN: 978-989-95079-6-8

Dominance-based Rough Set Analysis of Uncertain Data Tables

Tuan-Fang Fan[1] Churn-Jung Liau[2] Duen-Ren Liu[3]

1.Department of Computer Science and Information Engineering, National Penghu University, Penghu 880, Taiwan
2.Institute of Information Science, Academia Sinica, Taipei 115, Taiwan
3.Institute of Information Management, National Chiao-Tung University, Hsinchu 300, Taiwan
Email: dffan@npu.edu.tw, liaucj@iis.sinica.edu.tw, dliu@iim.nctu.edu.tw

Abstract— *In this paper, we propose a dominance-based rough set approach for the decision analysis of a preference-ordered uncertain data table, which is comprised of a finite set of objects described by a finite set of criteria. The domains of the criteria may have ordinal properties that express preference scales. In the proposed approach, we first compute the degree of dominance between any two objects based on their imprecise evaluations with respect to each criterion. This results in a fuzzy dominance relation on the universe. Then, we define the degree of adherence to the dominance principle by every pair of objects and the degree of consistency of each object. The consistency degrees of all objects are aggregated to derive the quality of the classification, with which we can define the reducts of an uncertain data table. In addition, the upward and downward unions of decision classes are fuzzy subsets of the universe. The lower and upper approximations of the decision classes based on the fuzzy dominance relation are thus fuzzy rough sets. By using the lower approximations of the decision classes, we can derive two types of decision rules that can be applied to new decision cases.*

Keywords— Dominance-based rough set approach, multi-criteria decision analysis, preference-ordered data tables, rough set theory, uncertain data tables.

1 Introduction

The theory of knowledge has long been an important topic in many academic disciplines, such as philosophy, psychology, economics, and artificial intelligence, whereas the storage and retrieval of data is the main concern of information science. In modern experimental science, knowledge is usually acquired from observed data, which is a valuable resource for researchers and decision-makers. However, when the amount of data is large, it is difficult to analyze the data and extract knowledge from it. With the aid of computers, the vast amount of data stored in relational data tables can be transformed into symbolic knowledge automatically. Thus, intelligent data analysis has received a great deal of attention in recent years. The rough set theory proposed in [19] provides an effective tool for extracting knowledge from data tables.

When rough set theory is applied to *multi-criteria decision analysis* (MCDA), it is crucial to deal with preference-ordered attribute domains and decision classes [6, 7, 8, 9, 10, 11, 12, 24]. The original rough set theory cannot handle inconsistencies arising from violations of the dominance principle due to its use of the indiscernibility relation. Therefore, in the above-mentioned works, the indiscernibility relation is replaced by a dominance relation to solve the multi-criteria sorting problem; and the data table is replaced by a pairwise comparison table to solve multi-criteria choice and ranking problems. The approach is called the *dominance-based rough set approach* (DRSA). For MCDA problems, DRSA can induce a set of decision rules from sample decisions provided by decision-makers. The induced rules form a comprehensive preference model and can provide recommendations about a new decision-making environment.

A strong assumption about data tables is that each object takes exactly one value with respect to an attribute. However, in practice, we may only have incomplete information about the values of an object's attributes. Thus, more general data tables are needed to represent incomplete information. For example, set-valued and interval-valued data tables have been introduced to represent incomplete information [15, 16, 17, 18, 25]. DRSA has also been extended to deal with missing values in MCDA problems [10, 24]. Since a data table with missing values is a special case of uncertain data tables, we propose further extending DRSA to the decision analysis of uncertain data tables. In this paper, we investigate such an extension based on the fuzzy dominance principle.

In the proposed approach, we first compute the degree of dominance between any two objects based on their imprecise evaluations with respect to each criterion. This results in a fuzzy dominance relation on the universe. Then, we define the degree of adherence to the dominance principle by every pair of objects and the degree of consistency of each object. The consistency degrees of all objects are aggregated to derive the quality of the classification, with which we can define the reducts of the uncertain data tables. In addition, the upward and downward unions of decision classes are fuzzy subsets of the universe. The lower and upper approximations of the decision classes based on the fuzzy dominance relation are thus fuzzy rough sets. By using the lower approximations of the decision classes, we can derive two types of decision rules that can be applied in new decision-making environments.

In the next section, we review the dominance-based rough set approach. Then, in Section 3, we present the extension of DRSA for decision analysis of uncertain data tables. Section 4 contains some concluding remarks.

2 Dominance-based Rough Set Approach

2.1 Rough set theory

The basic construct of rough set theory is an *approximation space*, which is defined as a pair (U, R), where U is a finite universe and $R \subseteq U \times U$ is an equivalence relation on U. A binary relation R is an equivalence relation if it is (1)reflexive (i.e., $(x, x) \in R$ for all $x \in U$); (2) symmetric (i.e., for all $x, y \in U$, if $(x, y) \in R$, then $(y, x) \in R$); and (3) transitive

ISBN: 978-989-95079-6-8

(i.e., for all $x, y, z \in U$, if $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$). An equivalence relation partitions the universe U into a family of equivalence classes so that each element of U belongs to exactly one of these equivalence classes. In other words, there exist $U_1, U_2, \cdots, U_k \subseteq U$ such that $U = \cup_{i=1}^{k} U_i$, $U_i \cap U_j = \emptyset$ for $i \neq j$; and for $x, y \in U$, $(x, y) \in R$ iff there exists i such that both x and $y \in U_i$. Thus, we can write an equivalence class of R as $[x]_R$ if it contains the element x. Note that $[x]_R = [y]_R$ iff $(x, y) \in R$.

According to Pawlak, knowledge is deep-seated in the classification capabilities of human beings. A classification is simply a partition of the universe, so an approximation space can construct basic knowledge about the objects in the universe. In philosophy, the *extension* of a concept is defined as the objects that are instances of the concept. For example, the extension of the concept "bird" is simply the set of all birds in the universe. Pawlak identified a concept by its extension. Thus, a subset of the universe is called a *concept* or a *category* in rough set theory.

Given an approximation space (U, R), each equivalence class of R is called an *R-basic category* or *R-basic concept*, and any union of R-basic categories is called an R-category. Now, for an arbitrary concept $X \subseteq U$, we are interested in defining X by using R-basic categories. We say that X is *R-definable*, if X is an R-category; otherwise X is *R-undefinable*. The R-definable concepts are also called *R-exact* sets, whereas R-undefinable concepts are said to be *R-inexact* or *R-rough*. When the approximation space is explicit from the context, we simply omit the qualifier R and call a set an exact set or a rough set.

A rough set can be approximated from below and above by two exact sets. The lower approximation and upper approximation of X are denoted by $\underline{R}X$ and $\overline{R}X$ respectively, and defined as follows:

$$\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}, \tag{1}$$

$$\overline{R}X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}. \tag{2}$$

2.2 Classical data tables

Although an approximation space is an abstract framework used to represent classification knowledge, it can easily be derived from a concrete data table (DT). The following formal definition of a data table is given in [20].

Definition 1 *A data table*[1] *is a tuple*

$$T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\}), \tag{3}$$

where U is a nonempty finite set, called the universe; A is a nonempty finite set of primitive attributes; for each $i \in A$, V_i is the domain of values for i; and for each $i \in A$, $f_i : U \rightarrow V_i$ is a total function.

An attribute in A is usually denoted by the lower-case letters i or a. In decision analysis, we assume the set of attributes is partitioned into $\{d\} \cup (A - \{d\})$, where d is called the *decision attribute*, and the remaining attributes in $A - \{d\}$ are called *condition attributes*.

[1] Also called knowledge representation systems, information systems, or attribute-value systems

Let $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ be a data table and $B \subseteq A$ be a subset of attributes. Then, we can define an equivalence relation, called the *indiscernibility relation* based on B, as

$$ind(B) = \{(x, y) \mid x, y \in U, f_i(x) = f_i(y) \forall i \in B\}. \tag{4}$$

In other words, x and y are B-indiscernible if they have the same values with respect to all the attributes in B. Consequently, for each $B \subseteq A$, $(U, ind(B))$ is an approximation space.

Let B be a subset of attributes. Then, an object x is *B-consistent* (with respect to the decision attribute d) if $[x]_B \subseteq [x]_d$, where $[x]_B = [x]_{ind(B)}$ and $[x]_d = [x]_{ind(\{d\})}$; otherwise, x is *B-inconsistent*. In other words, x is a B-consistent object in a data table if it satisfies the following *indiscernibility principle* for all $y \in U$:

$$(x, y) \in ind(B) \Rightarrow (x, y) \in ind(\{d\}). \tag{5}$$

That is, the objects that have the same evaluations as x on the condition attributes should have the same decision class assignment as x. Note that a *B-consistent* object corresponds to a decision rule whose decision class can be determined consistently based on the values of the attributes in B.

In [20], a decision logic (DL) is proposed for the representation of knowledge discovered from data tables. The basic alphabet of a DL consists of a finite set of attribute symbols A, and a finite set of value symbols V_i for $i \in \mathsf{A}$. Thus, the syntax of DL can be defined as follows.

Definition 2

1. *An atomic formula of DL is a descriptor (i, v), where $i \in \mathsf{A}$ and $v \in \mathsf{V}_i$.*
2. *The set of well-formed formulas (wffs) of a DL is the smallest set that contains the atomic formulas and is closed under the Boolean connectives $\neg$, $\wedge$, and $\vee$.*
3. *If φ and ψ are wffs of a DL, then $\varphi \longrightarrow \psi$ is a rule in the DL, where φ is called the antecedent of the rule and ψ is the consequent.*

A data table $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ relates to a given DL if there is a bijection $\tau : \mathsf{A} \rightarrow A$ such that, for every $a \in \mathsf{A}$, $V_{\tau(a)} = \mathsf{V}_a$. Thus, by somewhat abusing the notation, we usually denote an atomic formula as (i, v), where $i \in A$ and $v \in V_i$ if the data tables are clear from the context. Intuitively, each element in the universe of a data table corresponds to a data record, and an atomic formula (which is in fact an attribute-value pair) describes the value of some attribute in the data record. Thus, the atomic formulas (and therefore the wffs) can be satisfied or not satisfied by each data record. This results in a satisfaction relation between the universe and the set of wffs.

Definition 3 *Given a DL and a data table $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ related to it, the satisfaction relation $\models_T$ between U and the wffs of the DL is defined inductively as follows (the subscript T is omitted for brevity).*

1. *$x \models (i, v)$ iff $f_i(x) = v$,*
2. *$x \models \neg\varphi$ iff $x \not\models \varphi$,*

ISBN: 978-989-95079-6-8

3. $x \models \varphi \wedge \psi$ *iff* $x \models \varphi$ *and* $x \models \psi$,

4. $x \models \varphi \vee \psi$ *iff* $x \models \varphi$ *or* $x \models \psi$.

If φ is a DL wff, the set $m_T(\varphi)$ defined by

$$m_T(\varphi) = \{x \in U \mid x \models \varphi\} \tag{6}$$

is called the meaning set of the formula φ in T. If T is understood, we simply write $m(\varphi)$.

In terms of DL, each equivalence class of $ind(B)$ is characterized by a DL formula $\wedge_{i\in B}(i, v_i)$ and any formula φ of DL can be considered as a concept $m_T(\varphi)$. Thus, for each B-consistent object x, we can derive a valid rule $\wedge_{i\in B}(i, f_i(x)) \longrightarrow (d, f_d(x))$.

2.3 Preference-ordered data tables

In this subsection, we consider the DRSA in [10]. For MCDA problems, each object in a data table or decision table can be seen as a sample decision, and each condition attribute is a criterion for the decision. Since the domain of values of a criterion is usually ordered according to the decision-maker's preferences, we define a preference-ordered data table (PODT) as a tuple

$$T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\}), \tag{7}$$

where $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ is a classical data table; and for each $i \in A$, $\succeq_i \subseteq V_i \times V_i$ is a binary relation over V_i. The relation $\succeq_i$ is called a *weak preference relation* or *outranking* on V_i, and represents a preference over the set of objects with respect to the criterion i [24]. For $x, y \in U$, $f_i(x) \succeq_i f_i(y)$ means "x is at least as good as y with respect to criterion i". The weak preference relation $\succeq_i$ is supposed to be a complete preorder, i.e., a complete, reflexive, and transitive relation. In addition, we assume that the domain of the decision attribute is a finite set $V_d = \{1, 2, \cdots, n\}$ such that r is strictly preferred to s if $r > s$ for any $r, s \in V_d$. Thus, the weak preference relation $\succeq_d$ is defined as $r \succeq_d s$ iff $r \geq s$; consequently, $\succeq_d$ is a total order.

For a condition criterion i and an object x, $f_i(x)$ denotes the *evaluation* of the object with respect to the criterion i; and for the decision attribute d, $f_d(x)$ is the *assignment* of x to a decision label in V_d. Let P be a subset of criteria. We can then define the *P-dominance relation* $D_P \subseteq U \times U$ as follows:

$$(x, y) \in D_P \Leftrightarrow f_i(x) \succeq_i f_i(y) \forall i \in P. \tag{8}$$

When $(x, y) \in D_P$, we say that x P-dominates y, and that y is P-dominated by x. We usually use the infix notation xD_Py to denote $(x, y) \in D_P$. Although each $\succeq_i$ is a complete preorder, the dominance relation may simply be a preorder. If $P = \{i\}$ is a singleton, we write D_i instead of $D_{\{i\}}$. The most basic principle underlying DRSA is called the *dominance principle*. Let P denote a subset of condition criteria. Then, the dominance principle with respect to P can be expressed for $x, y \in U$ as follows:

$$xD_Py \Rightarrow xD_dy. \tag{9}$$

The principle states that if x P-dominates y (i.e., x is at least as good as y with respect to all criteria in P), then x should be assigned to a decision class at least as good as the class assigned to y.

In the classical rough set approach, a consistent object must satisfy the indiscernibility principle; however, in DRSA, we also require adherence to the dominance principle. Thus, an object x is *P-consistent* in the PODT $T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\})$ if for all $y \in U$, we have

$$(xD_Py \Rightarrow xD_dy) \wedge (yD_Px \Rightarrow yD_dx); \tag{10}$$

otherwise, x is *P-inconsistent*. Note that the dominance principle implies the indiscernibility principle because of the reflexivity of the dominance relation and the antisymmetry of $\succeq_d$.

Given the dominance relation D_P, the P-dominating set and P-dominated set of x are defined, respectively, as

$$D_P^+(x) = \{y \in U \mid yD_Px\} \tag{11}$$

and

$$D_P^-(x) = \{y \in U \mid xD_Py\}. \tag{12}$$

In addition, for each $t \in V_d$, we define the decision class Cl_t as $\{x \in U \mid f_d(x) = t\}$. Then, the *upward and downward unions of classes* are defined as

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s \tag{13}$$

and

$$Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s \tag{14}$$

respectively. Based on the P-dominating sets and P-dominated sets, we can define the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for each $t \in V_d$ as follows:

$$\underline{P}(Cl_t^{\geq}) = \{x \in U \mid D_P^+(x) \subseteq Cl_t^{\geq}\}, \tag{15}$$

$$\overline{P}(Cl_t^{\geq}) = \{x \in U \mid D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}, \tag{16}$$

$$\underline{P}(Cl_t^{\leq}) = \{x \in U \mid D_P^-(x) \subseteq Cl_t^{\geq}\}, \tag{17}$$

$$\overline{P}(Cl_t^{\leq}) = \{x \in U \mid D_P^+(x) \cap Cl_t^{\geq} \neq \emptyset\}. \tag{18}$$

The P-boundaries of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are then defined as

$$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}) \tag{19}$$

and

$$Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}) \tag{20}$$

respectively. Let $\mathbf{Cl} = \{Cl_t \mid t \in V_d\}$ denote the partition of the universe into decision classes. Then, the *quality of the approximation* of the partition $\mathbf{Cl}$ based on the set of criteria P is defined as the ratio

$$\gamma_P(\mathbf{Cl}) = \frac{|U - (\bigcup_{t\in V_d} Bn_P(Cl_t^{\geq}) \cup \bigcup_{t\in V_d} Bn_P(Cl_t^{\leq}))|}{|U|}. \tag{21}$$

Note that $\gamma_P(\mathbf{Cl})$ is equal to the ratio of P-consistent objects in the universe U. Let $C = A - \{d\}$ be the set of all condition criteria; then every minimal subset $P \subseteq C$ such that $\gamma_P(\mathbf{Cl}) = \gamma_C(\mathbf{Cl})$ is a *reduct* of C.

ISBN: 978-989-95079-6-8

2.4 Preference-ordered decision logic

To represent the rules induced from a PODT, preference-ordered decision logic (PODL) is proposed in [5]. The syntax of PODL is the same as that of DL, except for the form of the atomic formulas. An atomic formula in PODL is a descriptor in the form of $(\geq_i, v)$ or $(\leq_i, v)$, where $i \in A$ and $v \in V_i$. The satisfaction relation between U and the set of PODL wffs is defined in the same way as the relation for DL wffs, except that the satisfaction of an atomic formula is defined by $x \models (\geq_i, v)$ iff $f_i(x) \succeq v$, and by $x \models (\leq_i, v)$ iff $v \succeq f_i(x)$.

By using DRSA, two kinds of rules can be induced from a PODT explicitly. Let P be a reduct and $t \in V_d$. Then, for an object $x \in \underline{P}(Cl_t^{\geq})$, we can derive

$$\bigwedge_{i \in P} (\geq_i, f_i(x)) \longrightarrow (\geq_d, t), \tag{22}$$

and for an object $x \in \underline{P}(Cl_t^{\leq})$, we can derive

$$\bigwedge_{i \in P} (\leq_i, f_i(x)) \longrightarrow (\leq_d, t). \tag{23}$$

Note that if x is P-consistent, these two formulas also hold for $t = f_d(x)$.

3 DRSA for Uncertain Data Tables

3.1 Preference-ordered Uncertain Data Tables

Although the PODT can represent multi-criteria decision cases effectively, it inherits the restriction of the classical DT so that uncertain information can not be represented. An uncertain data table is a generalization of a DT such that the values of some or all of its attributes are imprecise [4, 1]. An analogous generalization can be applied to PODT to define preference-ordered uncertain data tables (POUDT). Formally, a POUDT is a tuple

$$T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\}), \tag{24}$$

where $U, A, \{(V_i, \succeq_i) \mid i \in A\}$ are defined as above, and for each $i \in A$, $f_i : U \to 2^{V_i} - \{\emptyset\}$. The intuition about a POUDT is that the evaluation of criterion i for an object x belongs to $f_i(x)$, although the evaluation is not known exactly. When $f_i(x)$ is a singleton, we say that the evaluation is precise. If all evaluations of T are precise, then T is said to be precise. Furthermore, we assume that for each criterion i, the Cartesian plane $V_i \times V_i$ is endowed with a uniform measure μ_i. Thus, for each subset $S \subseteq V_i \times V_i$, $\mu_i(S)$ is a non-negative real number. When V_i is a finite set, we take $\mu_i(S)$ as the cardinality of S; and when V_i is a real interval, we take $\mu_i(S)$ as the area of S.

3.2 Fuzzy dominance relation

In a POUDT, the objects may have imprecise evaluations with respect to the condition criteria and imprecise assignments to decision classes. Thus, the dominance relation between objects can not be determined with certainty. Instead, we define a degree of dominance between two objects with respect to each criterion i based on the associated measures μ_i. Formally, the dominance relation with respect to the criterion i is a fuzzy relation $D_i : U \times U \to [0, 1]$ such that for all $x \neq y$,

$$D_i(x, y) = \frac{\mu_i(\{(v_1, v_2) \mid v_1 \succeq_i v_2, v_1 \in f_i(x), v_2 \in f_i(y)\})}{\mu_i(f_i(x) \times f_i(y))}, \tag{25}$$

and $D_i(x, x) = 1$ for any $x \in U$.

An example of computing the degree of dominance is shown in Figure 1, where the evaluation of x with respect to criterion i, denoted by $s(x)$, is in a continuous interval $f_i(x) = [l_x, u_x]$. In this example, $D_i(x, y)$ is the ratio of the area of ABC over the area of $ABDE$, i.e., $\frac{u_x - l_y}{2(u_y - l_y)}$.

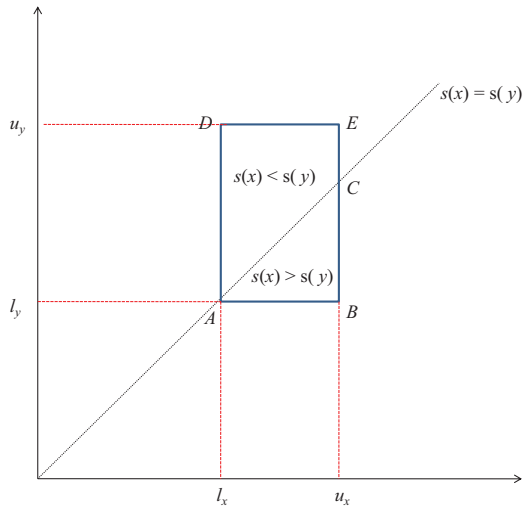


Figure 1: The degree of dominance between x and y

After deriving the fuzzy dominance relation for each criterion, we can aggregate all the relations into P-dominance relations for any subset of criteria P. Let $\otimes$, $\oplus$ and $\rightarrow$ denote, respectively, a t-norm, an s-norm and an implication operation[2] on $[0, 1]$. Then, the fuzzy P dominance relation $D_P : U \times U \to [0, 1]$ is defined as

$$D_P(x, y) = \bigotimes_{i \in P} D_i(x, y). \tag{26}$$

Since the dominance relation is a fuzzy relation, the satisfaction of the dominance principle is a matter of degree. Thus, the *degree of adherence* of (x, y) to the dominance principle with respect to a subset of condition criteria P is defined as

$$\delta_P(x, y) = D_P(x, y) \rightarrow D_d(x, y), \tag{27}$$

and the degree of P-consistency of x is defined as

$$\delta_P(x) = \bigotimes_{y \in U} (\delta_P(x, y) \otimes \delta_P(y, x)). \tag{28}$$

Let T be a POUDT. Then, the *quality of the classification* of T based on the set of criteria P is defined as

$$\gamma_P(T) = \frac{\sum_{x \in U} \delta_P(x)}{|U|}. \tag{29}$$

Note that $\gamma_P(T)$ is monotonic with respect to P, i.e., $\gamma_Q(T) \leq \gamma_P(T)$ if $Q \subseteq P$. Thus, we can define every minimal subset $P \subseteq C$ such that $\gamma_P(T) = \gamma_C(T)$ as a *reduct* of C, where

[2] For the properties of these operations, see a standard reference on fuzzy logic, such as [14]

ISBN: 978-989-95079-6-8

$C = A - \{d\}$ is the set of all condition criteria. In addition, the degree of P-consistency is monotonic with respect to P, so a reduct is also a minimal subset $P \subseteq C$ such that $\delta_P(x) = \delta_C(x)$ for all $x \in U$. However, because $\delta_P(x)$ is less sensitive to individual changes in $\delta_P(x, y)$, we can not guarantee that a reduct will preserve the degree of adherence to the dominance principle for each pair of objects. Thus, an alternative definition of the quality of the classification is

$$\eta_P(T) = \frac{\sum_{x,y \in U} \delta_P(x, y)}{|U|^2}. \tag{30}$$

The reducts can also be defined with respect to this kind of definition.

3.3 Dominance-based fuzzy rough approximations

In a POUDT, the assignment of a decision label to an object may be imprecise, so the decision classes may be fuzzy subsets of the universe. First, for each decision label $t \in V_d$, the decision class $Cl_t : U \to [0, 1]$ is defined by

$$Cl_t(x) = \begin{cases} \frac{1}{|f_d(x)|}, & \text{if } t \in f_d(x), \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

Second, the upward and downward unions of classes are defined by

$$Cl_t^{\geq}(x) = \frac{|f_d(x) \cap \{v \in V_d : v \geq t\}|}{|f_d(x)|} \tag{32}$$

and

$$Cl_t^{\leq}(x) = \frac{|f_d(x) \cap \{v \in V_d : v \leq t\}|}{|f_d(x)|} \tag{33}$$

respectively. Note that $Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$ and $Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$ only hold when we take the Łukasiewicz s-norm as the union operation, i.e., only when $(F \cup G)(x) = F(x) \oplus G(x)$, where $a \oplus b = \min(1, a + b)$. Finally, since our dominance relation is a fuzzy relation and the decision classes are fuzzy sets, the lower and upper approximations of these classes are defined in the same way as those for fuzzy rough sets[3, 21]. More specifically, the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for each $t \in V_d$ are defined as fuzzy subsets of U with the following membership functions:

$$\underline{P}(Cl_t^{\geq})(x) = \bigotimes_{y \in U} (D_P(y, x) \to Cl_t^{\geq}(y)), \tag{34}$$

$$\overline{P}(Cl_t^{\geq})(x) = \bigoplus_{y \in U} (D_P(x, y) \otimes Cl_t^{\geq}(y)), \tag{35}$$

$$\underline{P}(Cl_t^{\leq})(x) = \bigotimes_{y \in U} (D_P(x, y) \to Cl_t^{\leq}(y)), \tag{36}$$

$$\overline{P}(Cl_t^{\leq})(x) = \bigoplus_{y \in U} (D_P(y, x) \otimes Cl_t^{\leq}(y)). \tag{37}$$

3.4 Decision rules

To represent knowledge discovered from a POUDT, we generalize PODL to a kind of preference-ordered uncertain decision logic (POUDL). The syntax of POUDL is same as that of PODL, except that its atomic formulas are of the form $(\geq_i, s_i)$ or $(\leq_i, s_i)$, where $i \in A$ and $s_i \subseteq V_i$. When $s_i = \{v_i\}$ is a singleton, we abbreviate $(\geq_i, s_i)$ (resp. $(\leq_i, s_i)$) as $(\geq_i, v_i)$ (resp. $(\leq_i, v_i)$).

Let P denote a reduct of a POUDT and $t \in V_d$. Then, for each object x where $\underline{P}(Cl_t^{\geq})(x) > 0$ (or above some predetermined threshold), we can derive the first type of fuzzy rule:

$$\underline{P}(Cl_t^{\geq})(x) : \bigwedge_{i \in P} (\geq_i, f_i(x)) \longrightarrow (\geq_d, t); \tag{38}$$

and for each object x where $\underline{P}(Cl_t^{\leq})(x) > 0$ (or above some predetermined threshold), we can derive the second type of fuzzy rule:

$$\underline{P}(Cl_t^{\leq})(x) : \bigwedge_{i \in P} (\leq_i, f_i(x)) \longrightarrow (\leq_d, t), \tag{39}$$

where $\underline{P}(Cl_t^{\geq})(x)$ and $\underline{P}(Cl_t^{\leq})(x)$ are the respective degrees of truth of the rules.

Now, for a new decision case with (possibly imprecise) evaluations on the condition criteria P, we can apply these two types of rules to derive its decision label assignment. Specifically, let x be a new object such that for each criterion $i \in P$, $f_i(x) \subseteq V_i$ is given; and let α be a rule $c : \bigwedge_{i \in P} (\geq_i, s_i) \longrightarrow (\geq_d, t)$ discovered by the above-mentioned approach. Then, we can derive that the degree of satisfaction of $f_d(x) \succeq_d t$, according to the rule α and denoted by $\varepsilon(\alpha, f_d(x) \succeq_d t)$, is

$$c \otimes \bigotimes_{i \in P} \frac{\mu_i(\{(v_1, v_2) \mid v_1 \succeq_i v_2, v_1 \in f_i(x), v_2 \in s_i\})}{\mu_i(f_i(x) \times s_i)}. \tag{40}$$

Let $\mathcal{R}_t^{\geq}$ denote the set of all rules with the consequent $(\geq_d, t)$. Then, the final degree of $f_d(x) \succeq_d t$ is

$$\bigoplus_{\alpha \in \mathcal{R}_t^{\geq}} \varepsilon(\alpha, f_d(x) \succeq_d t). \tag{41}$$

Analogously, we can derive the degree of $f_d(x) \preceq_d t$ from the second type of rule.

4 Conclusion

The work reported in this paper extends DRSA to a dominance-based fuzzy rough set approach (DFRSA), which can be applied to the reduction of criteria and the induction of rules for decision analysis in a POUDT. Unlike other approaches that deal with imprecise evaluations and assignments, DFRSA induces fuzzy rules instead of qualitative rules. Thus, it would be worthwhile to compare DFRSA with other extensions of DRSA for handling uncertain data tables, like those proposed in [2, 13, 22]. In addition, the proposed approach may be useful for sparse data sets[23][3], so we will explore possible applications of DFRSA to such data sets in a future work. Moreover, since DFRSA is a general framework, we do not specify the t-norms used in the aggregation of consistency degrees or the implications used in the definition of adherence to the dominance principle. Consequently, we do not present detailed algorithms for the computation of reducts. The computational aspects of DFRSA for specialized t-norms and implication operations will aslo be addressed in a future work.

[3] We would like to thank the anonymous referee for pointing out this possibility.

ISBN: 978-989-95079-6-8

Acknowledgment

This work was partially supported by the National Science Council of Taiwan under grant 95-2221-E-001-029-MY3.

References

[1] K. Dembczynski, S. Greco, and R. Slowinski. Methodology of rough-set-based classification and sorting with hierarchical structure of attributes and criteria. *Control & Cybernetics*, 31:891–920, 2002.

[2] K. Dembczynski, S. Greco, and R. Slowinski. Second-order rough approximations in multi-criteria classification with imprecise evaluations and assignments. In D. Slezak, G. Wang, M.S. Szczuka, I. Duntsch, and Y. Yao, editors, *Proceedings of The Tenth International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, LNAI 3641, pages 54–63. Springer-Verlag, 2005.

[3] D. Dubois and H. Prade. Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, 17:191–209, 1990.

[4] T.F. Fan, W.C. Hu, and C.J. Liau. Decision logics for knowledge representation in data mining. In *Proceedings of the 25th Annual International Conference on Computer Software and Applications*, pages 626–631. IEEE Press, 2001.

[5] T.F. Fan, D.R. Liu, and G.H. Tzeng. Rough set-based logics for multicriteria decision analysis. *European Journal of Operational Research*, 182(1):340–355, 2007.

[6] S. Greco, B. Matarazzo, and R. Slowinski. Rough set approach to multi-attribute choice and ranking problems. In *Proceedings of the 12th International Conference on Multiple Criteria Decision Making*, pages 318–329, 1997.

[7] S. Greco, B. Matarazzo, and R. Slowinski. Rough approximation of a preference relation in a pairwise comparison table. In L. Polkowski and A. Skowron, editors, *Rough Sets in Data Mining and Knowledge Discovery*, pages 13–36. Physica-Verlag, 1998.

[8] S. Greco, B. Matarazzo, and R. Slowinski. Rough approximation of a preference relation by dominance relations. *European Journal of Operational Research*, 117(1):63–83, 1999.

[9] S. Greco, B. Matarazzo, and R. Slowinski. Extension of the rough set approach to multicriteria decision support. *INFOR Journal: Information Systems & Operational Research*, 38(3):161–195, 2000.

[10] S. Greco, B. Matarazzo, and R. Slowinski. Rough set theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1):1–47, 2001.

[11] S. Greco, B. Matarazzo, and R. Slowinski. Rough sets methodology for sorting problems in presence of multiple attributes and criteria. *European Journal of Operational Research*, 138(2):247–259, 2002.

[12] S. Greco, B. Matarazzo, and R. Slowinski. Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *European Journal of Operational Research*, 158(2):271–292, 2004.

[13] S. Greco, B. Matarazzo, and R. Slowinski. Dominance-based rough set approach as a proper way of handling graduality in rough set theory. *Transactions on Rough sets*, VII:36–52, 2007.

[14] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, 1998.

[15] M. Kryszkiewicz. Properties of incomplete information systems in the framework of rough sets. In L. Polkowski and A. Skowron, editors, *Rough Sets in Knowledge Discovery*, pages 422–450. Physica-Verlag, 1998.

[16] M. Kryszkiewicz and H. Rybiński. Reducing information systems with uncertain attributes. In Z. W. Raś and M. Michalewicz, editors, *Proceedings of the 9th ISMIS*, LNAI 1079, pages 285–294. Springer-Verlag, 1996.

[17] M. Kryszkiewicz and H. Rybiński. Reducing information systems with uncertain real value attributes. In *Proceedings of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 1165–1169, 1996.

[18] W. Lipski. On databases with incomplete information. *Journal of the ACM*, 28(1):41–70, 1981.

[19] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11(15):341–356, 1982.

[20] Z. Pawlak. *Rough Sets–Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, 1991.

[21] A.M. Radzikowska and E.E. Kerre. A comparative study of fuzzy rough sets. *Fuzzy Sets and Systems*, 126(2):137–155, 2002.

[22] H. Sakai, R. Ishibashi, and M. Nakata. Lower and upper approximations of rules in non-deterministic information systems. In C.C. Chan, J.W. Grzymala-Busse, and W. Ziarko, editors, *Proceedings of The 6th International Conference on Rough Sets and Current Trends in Computing*, LNAI 5306, pages 299–309. Springer-Verlag, 2008.

[23] D. Slezak. Rough sets and few-objects-many-attributes problem: The case study of analysis of gene expression data sets. In *Proceedings of the International Conference on Frontiers in the Convergence of Bioscience and Information Technologies*, pages 437–442. IEEE Press, 2007.

[24] R. Slowinski, S. Greco, and B. Matarazzo. Rough set analysis of preference-ordered data. In J.J. Alpigini, J.F. Peters, A. Skowron, and N. Zhong, editors, *Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing*, LNAI 2475, pages 44–59. Springer-Verlag, 2002.

[25] Y.Y. Yao and Q. Liu. A generalized decision logic in interval-set-valued information tables. In N. Zhong, A. Skowron, and S. Ohsuga, editors, *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*, LNAI 1711, pages 285–293. Springer-Verlag, 1999.

ISBN: 978-989-95079-6-8

Overlap index, overlap functions and migrativity

H. Bustince[1] J. Fernandez[1] R. Mesiar[2] J. Montero[3] R. Orduna[1]

1.Dpto. Automatica y Computacion, Universidad Publica de Navarra
Pamplona, Spain
2.Department of Mathematics and Descriptive Geometry, Slovak University of Technology,
SK-813 68 Bratislava, Slovakia
Institute of Information Theory and Automation, Czech Academy of Sciences,
CZ-182 08 Prague, Czech Republic
3.Facultad de Matematicas, Universidad Complutense de Madrid
Madrid, Spain
Email: {bustince,fcojavier.fernandez,raul.orduna}@unavarra.es, mesiar@math.sk, monty@mat.ucm.es

Abstract— *In this work we study overlap degrees expressed in terms of overlap functions. We present the basic properties that from our point of view must satisfy these overlap functions. We study a construction method, we analyze which t-norms are also overlap functions and we prove that if we apply particular aggregations to such functions we recover the overlap index between fuzzy sets as defined by Dubois, and the consistency index of Zadeh. We also consider some properties that can be required to overlap functions, as k-Lipschitzianity or migrativity .*

Keywords— Overlap degree, Overlap function, Overlap index, t-norm, Migrativity.

1 Overlap function. Definition and properties

Zadeh's fuzzy sets theory has been very useful for solving problems which are described by imprecise models and with a large amount of noise. In particular, this theory is very appropriate to study the problem of identifying the objects in an image (see [18, 26]).

To separate the object from the background in an image, the first thing to do is to represent the object by means of a fuzzy set and the background by means of another one. The success of the separation method lies on the correct choice of those fuzzy sets, which do not need to be disjoint in the sense of Ruspini [27] (see [1, 2, 17]).

To build these sets it is necessary to know the exact property that characterizes the pixels belonging to the object (background). This property determines the expression of the membership function associated to the fuzzy set representing the object (background)(see [8, 9]). Usually, this function is not precisely known. There are some pixels for which the expert is sure they belong to the object or the background, but there are other pixels for which the expert hesitates. It is for these last ones that the value of the membership function is not accurately known.

So, suppose that for a given an image, we ask an expert to assign to each pixel of intensity q the following values:
$\mu_B(q)$, representing the membership of the pixel to the background.
$\mu_O(q)$, representing the membership of the pixel to the object.

In Fig.1 we show the two membership functions provided by the expert to represent an image in an L gray-level scale. We can deduce that, for intensities less than q_i, the expert is sure that the pixels do not belong to the object. For intensities greater than q_j he is sure that the pixels do not belong to the background. However, for intensities between q_i and q_j the expert is not sure about the membership of the pixels, with intensity q_k corresponding to the maximum lack of knowledge.

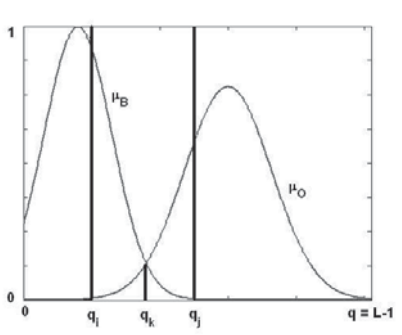


Figure 1: Overlap between two functions

From this analysis we deduce that the overlap degree between the two functions can be understood as a representation of the lack of knowledge of the expert when he has to settle if a given pixel belongs to the background or to the object. So we can define the overlap degree between $\mu_B(q)$ and $\mu_O(q)$ by means of an overlap function

$$G_S : [0,1] \times [0,1] \to [0,1]$$

such that:
(G_S0) G_S depends only on $\mu_B(q)$ and $\mu_O(q)$.
(G_S1) G_S is symmetric. The overlap degree does not depend on the order we consider the membership degrees.
(G_S2)
$G_S(\mu_B(q), \mu_O(q)) = 0$ if and only if
$\mu_B(q) = 0$ or $\mu_O(q) = 0$, (i.e., $\min(\mu_B(q), \mu_O(q)) = 0$) .
(G_S3)
$G_S(\mu_B(q), \mu_O(q)) = 1$ if and only if
$\mu_B(q) = 1$ and $\mu_O(q) = 1$, (i.e., $\min(\mu_B(q), \mu_O(q)) = 1$).
(G_S4) If the membership degrees increase, so does the overlap degree.
(G_S5) Continuity. The overlap degree between the two memberships associated to a given pixel must not react chaotically

ISBN: 978-989-95079-6-8

under small variations of the values of the membership to the background or to the object.

Apart from these five necessary properties, we consider it is also natural to require the following migrative property.

(G_S6) Migrativity. If we decrease $\mu_B(q)$ in a proportion $\alpha \in (0,1]$, the overlap degree G_S should decrease in the same amount as if we decrease $\mu_O(q)$ in the same proportion α. That is:

$$G_S(\alpha\mu_B(q), \mu_O(q)) = G_S(\mu_B(q), \alpha\mu_O(q)) \text{ for all } \alpha \in (0,1].$$

Definition 1 *A mapping $G_S : [0,1]^2 \to [0,1]$ is an overlap function if and only if it satisfies $(G_S0) - (G_S5)$. If G_S satisfies (G_S6) for any $0 < \alpha \leq 1$, we say that it is a migrative overlap function.*

Let's denote by $\mathcal{G}$ the set of overlap functions in the sense of Definition 1. Then the following result is immediate.

Theorem 1 *$(\mathcal{G}, \leq_{\mathcal{G}})$ with the ordering $\leq_{\mathcal{G}}$ defined for $G_1, G_2 \in \mathcal{G}$ by*

$$G_1 \leq_{\mathcal{G}} G_2 \text{ if and only if } G_1(x,y) \leq G_2(x,y) \qquad (1)$$

for all $x, y \in [0,1]$ is a lattice.

It is clear that the lattice $(\mathcal{G}, \leq_{\mathcal{G}})$ is not complete, neither has it top or bottom elements. On the other hand, it is closed under the action of appropriate aggregation functions, as shown next.

Definition 2 *An aggregation function of dimension n ([7, 11, 20, 16]) is a mapping $M : [0,1]^n \to [0,1]$ satisfying (see also [12, 3]):*
$M1$. $M(0, \cdots, 0) = 0$ and $M(1, \ldots, 1) = 1$.
$M2$. For any $(x_1, \cdots, x_n)$, $(y_1, \cdots, y_n) \in [0,1]^n$, if $x_i \leq y_i$ for any $i \in \{1, \cdots, n\}$, then $M(x_1, \cdots, x_n) \leq M(y_1, \cdots, y_n)$; that is, M is monotone increasing in all its arguments.

Theorem 2 *Let M be a binary aggregation function without zero divisors (that is, $M(x,y) = 0$ implies $min(x,y) = 0$) and such that $M(x,y) = 1$ only if $\max(x,y) = 1$. Then, $M(G_1, G_2) \in \mathcal{G}$ for any $G_1, G_2 \in \mathcal{G}$.*

2 Construction

The following theorem provides both a characterization and a construction method of overlap functions.

Theorem 3 *A mapping $G_S : [0,1]^2 \to [0,1]$ is an overlap function if and only if it can be written as*

$$G_S(x,y) = \frac{f(x,y)}{f(x,y) + h(x,y)} \qquad (2)$$

for some $f, h : [0,1]^2 \to [0,1]$ such that
1) f and h are symmetric;
2) f is non decreasing and h is non increasing;
3) $f(x,y) = 0$ if and only if $\min(x,y) = 0$;
4) $h(x,y) = 0$ if and only if $\min(x,y) = 1$;
5) f and h are continuous;

Example 1.
1) If $f(x,y) = \min(x,y)$ and $h(x,y) = \max(1-x, 1-y)$, then $G_S(x,y) = \min(x,y)$ is an overlap function.
2) If we take $f(x,y) = \sqrt{x.y}$ and $h(x,y) = \max(1-x, 1-y)$, then the construction proposed in Theorem 3 provides an overlap function.
3) If $f(x,y) = \sqrt{x \cdot y}$ and $h(x,y) = 1 - x \cdot y$, expression

$$G_S(x,y) = \frac{\sqrt{x \cdot y}}{\sqrt{x \cdot y} + 1 - x \cdot y} \qquad (3)$$

is an overlap function.

Corollary 1 *In the setting of Theorem 3, assume that G_S can be expressed in two different ways:*

$$G_S(x,y) = \frac{f_1(x,y)}{f_1(x,y) + h_1(x,y)} = \frac{f_2(x,y)}{f_2(x,y) + h_2(x,y)} \qquad (4)$$

for any $x, y \in [0,1]$. Let M be a binary continuous aggregation function that is homogeneous of order one, that is, such that

$$M(\lambda x, \lambda y) = \lambda M(x,y) \qquad (5)$$

for any $x, y \in [0,1]$ and any $\lambda \geq 0$ such that $\lambda x, \lambda y \in [0,1]$. Then, if we define $f(x,y) = M(f_1(x,y), f_2(x,y))$ and $h(x,y) = M(h_1(x,y), h_2(x,y))$ it also holds that

$$G_S(x,y) = \frac{f(x,y)}{f(x,y) + h(x,y)} \, . \qquad (6)$$

Proof. First observe that $f_i = h_i \frac{G_S}{1-G_S}$ for $i = 1, 2$. By the homogeneity condition on M, also $f = h \frac{G_S}{1-G_S}$ and the result follows.

2.1 Specific case: t-norms

In this subsection we study under which conditions we can assure a t-norm satisfies the properties required to an overlap function.

We know that a t-norm is a commutative, associative, increasing mapping $T : [0,1]^2 \to [0,1]$ such that $T(x,1) = x$ for any $x \in [0,1]$. So t-norms satisfy (G_S1) and (G_S4). They also satisfy (G_S3), since if $T(x,y) = 1$, as $T(x,y) \leq \min(x,y)$, it must be $x = y = 1$. The reciprocal is direct, taking into account that 1 is the neutral element of any t-norm. So it is only required to study condition (G_S2). Observe that when dealing with t-norms, the necessary condition in this property coincides with the definition of positive t-norm. Besides, the analysis of the conditions under which a t-norm satisfies (G_S2) leads to the following classification result.

Theorem 4 *If a t-norm T is an overlap function, then T is of one of the following three types:*
1) $T = \min$;
2) T is strict;
3) T is the ordinal sum of the family $\{([a_m, b_m], T_m)\}$, with all the T_m continuous Archimedean t-norms and such that if for some m_0 we have $a_{m_0} = 0$ then necessarily T_{m_0} is a strict t-norm.

Proof. By hypothesis, T is continuous. From the classification of continuous t-norms given in page 11 of [15] (see [25],

ISBN: 978-989-95079-6-8

[21]), we know that for a continuous t-norm T there are three possibilities:
1.- $T = \min$;
2.- T is Archimedean;
3.- There exists a family $\{([a_m, b_m], T_m)\}$ such that T is the ordinal sum of this family in the sense of [15].

As, by hypothesis, T is an overlap function, (G_S2) holds. If T is also Archimedean, we have that T is strict.

Suppose now that T is the ordinal sum of the family $\{([a_m, b_m], T_m)\}$; that is:

$$T(x,y) = \begin{cases} a_m + (b_m - a_m)T_m(\frac{x-a_m}{b_m-a_m}, \frac{y-a_m}{b_m-a_m}) \\ \qquad \text{if } (x,y) \in [a_m, b_m]^2 \ ; \\ \min(x,y) \text{ otherwise.} \end{cases} \tag{7}$$

We know that for any t-norm T, if $\min(x,y) = 0$ then $T(x,y) = 0$. Since our t-norm is an overlap function, also the reciprocal is true. So, if $T(x,y) = 0$ two things can happen:
a) (x,y) does not belong to any $[a_m, b_m]^2$. Then we have $T(x,y) = \min(x,y)$ for that (x,y).
b) (x,y) belongs to $[a_m, b_m]^2$. As $T(x,y) = 0 = a_m + (b_m - a_m)T_m(\frac{x-a_m}{b_m-a_m}, \frac{y-a_m}{b_m-a_m})$, we have $a_m = 0$ and $b_m \neq 0$ since otherwise the interval would be $[0,0]$ and $x = y = 0$. We know that T verifies (G_S2) and if $T(x,y) = 0$, then $T_m(\frac{x}{b_m}, \frac{y}{b_m}) = 0$, so T_m also verifies (G_S2). Hence, the continuous and Archimedean t-norm T_m associated to the interval $[0, b_m]$ also satisfies (G_S2), so it is strict.

Example 2. 1) In the construction of the following overlap function we use item 3) of Theorem 4 . We take as t-norm for the corresponding interval $[0, b_m]$ the product, which is strict, continuous and Archimedean.

$$G_S(x,y) = \begin{cases} 2xy \text{ if } (x,y) \in [0, 0.5]^2 \ ; \\ \min(x,y) \text{ in other case.} \end{cases} \tag{8}$$

2) In the construction of the following overlap function we take the Lukasiewicz and the product t-norms (see page 84 in [19]). Observe that now we do not consider any interval of the type $[0, b_m]$.

$$G_S(x,y) = \\ \begin{cases} 0.1 + 2.5(x - 0.1)(y - 0.1) \text{ if } (x,y) \in [0.1, 0.5]^2; \\ 0.7 + \max(x + y - 1.6, 0) \text{ if } (x,y) \in [0.7, 0.9]^2; \\ \min(x,y) \text{ in other case.} \end{cases} \tag{9}$$

3) The following t-norm does not satisfy (G_S2). This is due to the fact that in $[0, 0.25]^2$ we take Lukasiewicz t-norm, which is continuous and Archimedean, but it is not strict.

$$T(x,y) = \begin{cases} \max(x + y - 0.25, 0) \text{ if } (x,y) \in [0, 0.25]^2 \\ \min(x,y) \text{ in other case.} \end{cases} \tag{10}$$

3 Overlap functions and k-Lipschitzianity

In this section we consider a particular type of overlap functions, satisfying a sort of stronger continuity. We start adapting the definition of k-Lipschitz functions to the overlap function case.

Definition 3 *Let $k \geq 1$. An overlap function G_S is k-Lipschitz if for any $x, y, z, t \in [0,1]$ it holds*

$$|G_S(x,y) - G_S(z,t)| \leq k(|x - z| + |y - t|) \ . \tag{11}$$

It is worth remarking that the usual definition of k-Lipschitziniaty allows any value of k greater than zero. But, in the case of overlap functions, just by taking $x = y = z = 1$ and $t = 0$ the restriction to $k \geq 1$ becomes justified.

The set of k-Lipschitz overlap functions is bounded and its supremum can be easily determined, as the next result shows.

Theorem 5 *Let $k \geq 1$. Then the supremum of the set of k-Lipschitz overlap functions is given by the mapping $\min(kx, ky, 1)$, whereas the infimum is given by $\max(kx + ky - 2k + 1, 0)$.*

Proof. Suppose that $G_S(x,y) > \min(kx, ky, 1)$ for some $x, y \in [0,1]$. Since $G_S(x,y) \leq 1$, this means that $\min(kx, ky, 1) = kx$ or $\min(kx, ky, 1) = ky$. In the first case, $y = t = 1$ and $z = 0$ in Eq. (11), we arrive at

$$kx < G_S(x, 1) \leq kx \tag{12}$$

which is a contradiction. The second case is analogous. On the other hand, by defining for $\epsilon > 0$ the mappings

$$\max(x \cdot y, (1 - \epsilon))(\min(kx, ky, 1)) \tag{13}$$

we get a sequence of overlap functions which converges uniformly to $\min(kx, ky, 1)$ as $\epsilon \to 0$. The proof for the lower bound is similar.

The mapping $\max(kx + ky - 2k + 1, 0)$ is never an overlap function. On the contrary, although in general, the mapping $\min(kx, ky, 1)$ for $k > 1$ and $x, y \in [0,1]$ such that $kx, ky \in [0,1]$ does not define an overlap function, (since by taking $x = y = \frac{1}{k}$ we see that it does not fulfill condition (G_S3)), $\min(x,y)$ is an overlap function, so we have the following corollary.

Corollary 2 *The mapping $\min(x,y)$ is the strongest 1-Lipschitz overlap function, in the sense that for any other 1-Lipschitz overlap function G_S the inequality*

$$G_S(x,y) \leq \min(x,y) \tag{14}$$

holds for any $x, y \in [0,1]$.

For associative k-Lipschitz overlap functions we have the next result which can be derived from [23, 24].

Theorem 6 *If G_S is an associative k-Lipschitz overlap function, then G_S is a t-norm of the form given in Theorem 4, where each involved strict t-norm T (see item 2) or item 3)) has a k-convex additive generator t, i.e.,*

$$t(y + k\epsilon) - t(y) \leq t(x + \epsilon) - t(x) \tag{15}$$

for all $0 \leq y \leq x < 1$ and $\epsilon \in]0, \min(1 - x, (1 - y)/k)[$.

4 Overlap index. Construction from overlap functions

In this section we are going to build overlap indexes by aggregating overlap degrees. We start recalling the concepts of overlap index and consistency. Then we justify why most of the overlap indexes in the literature do not satisfy one of the four conditions required by Dubois and Prade. Next we show the properties we have to demand to aggregation functions so

ISBN: 978-989-95079-6-8

that, when applied to the previously studied overlap degrees, we get overlap indexes. We finish showing how to construct such indexes and, in particular, how to recover the two most used expressions.

We denote by $F(U)$ the set of all fuzzy sets over the finite, non empty referential U ($Cardinal(U) = n$). We are going to represent the fuzzy sets over U in the following way:

$$A = \{(u, \mu_A(u)) | u \in U\} \tag{16}$$

In 1978 Zadeh [29] presented the natural extension to the fuzzy set theory of the classical concept of overlap, which he called consistency:

$$O(A, B) = \sup_{i=1}^{n}(\min(\mu_A(u_i), \mu_B(u_i))). \tag{17}$$

Clearly, $O(A, B) = 0$ if A and B are completely disjoint, and $O(A, B) = 1$ if there is $u_i \in U$ such that $\mu_A(u_i) = \mu_B(u_i) = 1$.

In 1982 Dubois and Prade [13] presented the following axiomatization for the overlap index:

Definition 4 *An overlap index is a function $O(A, B)$ from $F(U) \times F(U)$ on the unit interval such that:*
$(O1)$ $O(A, B) = 0$ if and only if A and B have disjoint supports;
$(O2)$ $O(A, B) = 1$, if ($\mu_A(u_i) = 0$ or $\mu_B(u_i) = 1$) or ($\mu_A(u_i) = 1$ or $\mu_B(u_i) = 0$);
$(O3)$ $O(A, B) = O(B, A)$;
$(O4)$ If $B \leq C$, then $O(A, B) \leq O(A, C)$.

Condition $(O2)$ in this definition presents the advantage of that, if A is not fuzzy, then $O(A, A) = 1$. But Dubois, Ostasiewicz and Prade in [14] settled the following:

1. For subnormal fuzzy sets (i.e., $\mu_A(u_i) < 1$ and $\mu_B(u_i) < 1$ for any $u_i \in U$), $(O2)$ *must be ignored.*
2. The ROC index(see [13]) does not fulfill $(O2)$.

It is also interesting to notice that if $A = \{(u_i, \mu_A(u_i) = 0) | u_i \in U\}$, then $\min(A, A) = \{(u_i, \mu_{\min(A,A)}(u_i) = 0) | u_i \in U\}$ and from $(O1)$, $O(A, A) = 0$. If we also impose $(O2)$, then $O(A, A) = 1$. So we get a contradiction.

Due to all these considerations, usually only conditions $(O1)$,$(O3)$ and $(O4)$ from Def. 4 are required to overlap indexes.

In the following theorem we present a construction method of overlap indexes, by means of aggregation functions.

Theorem 7 *Let $M : [0, 1]^n \to [0, 1]$ be an aggregation function being idempotent and such that $M(x_1, \ldots, x_n) = 0$ if and only if $x_1 = \ldots = x_n = 0$.*
Let $G_S : [0, 1]^2 \to [0, 1]$ be a mapping and consider:
$O : F(U) \times F(U) \to [0, 1]$ *defined as*
$O(A, B) = M_{i=1}^{n}(G_S(\mu_A(u_i), \mu_B(u_i)))$
Then the following items hold:
i) O verifies $(O1)$ if and only if G_S verifies (G_S2);
ii) O verifies $(O3)$ if and only if G_S verifies (G_S1);
iii) If G_S verifies (G_S4), then O verifies $(O4)$.

Example 3.

1. $O(A, B) = \max_{i=1}^{n} G_S(\mu_A(u_i), \mu_B(u_i))$
2. $O(A, B) = \left(\frac{1}{n}\sum_{i=1}^{n}(G_S(\mu_A(u_i), \mu_B(u_i))^\beta\right)^{\frac{1}{\beta}}, \beta \neq 0$

Notes for the Example

1. If in 1. we take $G_S(x, y) = \min(x, y)$, then we recover Zadeh's consistency index (see [29]).
2. If there is a single u_i such that $G_S(\mu_A(u_i), \mu_B(u_i)) = 1$, then, by 1. we have $O(A, B) = 1$. This fact of a single element making $O(A, B) = 1$ suggest us to use 2. instead of 1.
3. Expression 2. satisfies:

   $O(A, B) = 1$ if and only if $G_S(\mu_A(u_i), \mu_B(u_i)) = 1$

   for all $i = 1, \ldots, n$. In these conditions, if $G_S(x, y) = \min(x, y)$ or $G_S(x, y) = x \cdot y$ or $G_S(x, y) = \sqrt{x \cdot y}$, then $O(A, B) = 1$ if and only if

   $$A = B = \{(u_i, \mu_A(u_i) = \mu_B(u_i) = 1) | u_i \in U\} \tag{18}$$

   so 2., as most of the expressions of overlap indexes (see [4, 5, 6, 14, 13]), does not fulfill condition $(O2)$ in Definition 2.

Corollary 3 *In the setting of Theorem 7, if we demand M to satisfy $M(x_1, \ldots, x_n) = 1$ if and only if $x_1 = \ldots = x_n = 1$, then, if G_S verifies (G_S3) we have*
$O(A, B) = 1$ if and only if $A = B = \{(u_i, \mu_A(u_i) = 1) | u_i \in U\}$.

5 Construction of migrative overlap functions

As we have already said, migrativity seems to be quite a natural property to be required to overlap functions. In [10], an in-deep study of the migrativity property is carried on for general aggregation functions. In this paper we use the following results that are proved there.

Lemma 1 *Let $H : [0, 1]^2 \to [0, 1]$ be a binary function. Then H is migrative if and only if $H(x, y) = H(1, x \cdot y)$, for all $x, y \in [0, 1]$.*

Lemma 2 *A function $H : [0, 1]^2 \to [0, 1]$ is migrative if and only if there exists $h : [0, 1] \to [0, 1]$ such that $H(x, y) = h(x \cdot y)$, for all $x, y \in [0, 1]$.*

In this section we present a characterization theorem of migrative overlap functions. Clearly there are overlap functions that are not migrative (for instance, those in Ex. 1). In the following results we prove that there exist also overlap functions which are migrative.

Theorem 8 *Let $H : [0, 1]^2 \to [0, 1]$ be a migrative mapping (not necessarily an overlap function). Then*

$$H(x, 1) = H(\sqrt{x}, \sqrt{x}) \tag{19}$$

for each $x \in [0, 1]$.

ISBN: 978-989-95079-6-8

Proof. By Lemmas 1 and 2 there exists a mapping $h : [0,1] \to [0,1]$ such that $H(x,1) = h(x) = h(\sqrt{x} \cdot \sqrt{x}) = H(\sqrt{x}, \sqrt{x})$.

Theorem 9 *A mapping $G_S : [0,1] \to [0,1]$ is a migrative overlap function if and only if there exists a non decreasing function $g : [0,1] \to [0,1]$ satisfying $g^{-1}((0,1)) = (0,1)$ such that*

$$G_S(x,y) = g(x \cdot y). \tag{20}$$

Proof. (Necessity) Since G_S is migrative, by Lemma 2 we know that there exists a function $g : [0,1] \to [0,1]$ such that $G_S(x,y) = g(x \cdot y)$ for all $x, y \in [0,1]$. As G_S is an overlap function, g is not decreasing and continuous. Besides G_S satisfies $(G_S 2)$ and $(G_S 3)$, so:

$$g(x) = g(x \cdot 1) = G_S(x,1) = 0 \text{ if and only if } x = 0 \tag{21}$$

$$g(x) = g(x \cdot 1) = G_S(x,1) = 1 \text{ if and only if } x = 1 \tag{22}$$

(Sufficiency) By Lemma 2 we have that $G_S(x,y) = g(x,y)$ satisfies $(G_S 6)$. From the migrativity it is clear that $(G_S 1)$ holds. On the other hand:

$$G_S(x,y) = 0 = g(x \cdot y) \text{ if and only if } x \cdot y = 0 \tag{23}$$

$$G_S(x,y) = 1 = g(x \cdot y) \text{ if and only if } x \cdot y = 1 \tag{24}$$

Clearly, G_S satisfies $(G_S 4)$ and $(G_S 5)$ since g is non decreasing and continuous.

Acknowledgements. H. Bustince, J. Fernandez and R. Orduna have been supported by project TIN2007-65981. R. Mesiar has been supported by grants APVV-0012-07 and GACR 402/08/0618. J. Montero has been supported by project TIN2006-06190.

References

[1] A. Amo, D. Gómez and J. Montero, Spectral fuzzy classification: a supervised approach, *Mathware and Soft Computing*, **10**, Pages 141-154, 2003.

[2] A. Amo, J. Montero, G. Biging and V. Cutello, Fuzzy classification systems, *European Journal of Operational Research*, **156**, Pages 459-507, 2004.

[3] A. Amo, J. Montero, D. Gómez and E. Molina, Representation of consistent recursive rules, *European Journal of Operational Research*, **130**, Pages 29-53, 2001.

[4] H. Bustince, M. Pagola, E. Barrenechea, Construction of fuzzy indices from fuzzy DI-subsethood measures: Application to the global comparison of images, *Information Sciences*, 177, **3**, Pages 906-929, 2007.

[5] H. Bustince, E. Barrenechea, M. Pagola, et al., Weak fuzzy S-subsethood measures. Overlap index, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, 14 **5**, Pages. 537-560, 2006.

[6] H. Bustince, V. Mohedano, E. Barrenechea, et al. Definition and construction of fuzzy DI-subsethood measures, *Information Sciences*, 176, **21**, Pages 3190-3231, 2006.

[7] H. Bustince H., J. Montero, E. Barrenechea and M. Pagola, Semiautoduality in a restricted family of aggregation operators, *Fuzzy Sets and Systems*, 158, **12**, Pages 1360-1377, 2007.

[8] H. Bustince, E. Barrenechea, M. Pagola, Restricted equivalence functions, *Fuzzy Sets and Systems*, 157 **17**, Pages 2333-2346, 2006.

[9] H. Bustince, E. Barrenechea, M. Pagola, Image thresholding using restricted equivalence functions and maximizing the measures of similarity, *Fuzzy Sets and Systems*, 158, **5**, Pages 496-516, 2007.

[10] Bustince H., Montero J. and Mesiar R., Migrativity of Aggregation Operators, *Fuzzy Sets and Systems*, 160, **6**, Pages 766-777, 2008.

[11] T. Calvo, A. Kolesárová, M. Komorníková and R. Mesiar: Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor and R. Mesiar (Eds.): *Aggregation Operators New Trends and Applications* (Physica-Verlag, Heidelberg); Pages 3-104, 2002.

[12] V. Cutello and J. Montero, Recursive connective rules, *International Journal of Intelligent Systems*, 14, Pages 3-20, 1999.

[13] D. Dubois and J.L. Koning, Social choice axioms for fuzzy set aggregation. *Fuzzy Sets and Systems* 58, Pages 339-342, 1991.

[14] D. Dubois, W. Ostasiewicz, H. Prade, Fuzzy Sets: History and Basic Notions, in: D. Dubois, H. Prade (Eds), *Fundamentals of Fuzzy Sets*, Kluwer, Boston, MA, Pages 21-124, 2000.

[15] J. Fodor and M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support, in: *Theory and Decision Library* (Kluwer Academic Publishers) 1994.

[16] D. Gómez and J. Montero: A discussion on aggregation operators. *Kybernetika* 40, Pages 107-120, 2004.

[17] D. Gómez, J. Montero and H. Bustince, On the relevance of some families of fuzzy sets, *Fuzzy Sets and Systems*, 158, Pages 2429-2442, 2007.

[18] L.K. Huang and M.J. Wang, Image thresholding by minimizing the measure of fuzziness, *Pattern recognition*, 28, **1**, Pages 41-51, 1995.

[19] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Trends in Logic. *Studia Logica Library*, Kluwer Academic Publishers, Dordrecht, 2000.

[20] G.J. Klir and T.A. Folger: *Fuzzy Sets, Uncertainty and Information.* Prentice Hall, Englewood Cliffs, NJ, 1988.

[21] C.H. Ling, Representation of associative functions, *Publ. Math. Debrecen*, 12, Pages 189-212, 1965.

[22] R. Mesiar, V. Novák, Open problems from the 2nd International conference of fuzzy sets theory and its applications, *Fuzzy Sets and Systems*, 81, Pages 185-190, 1996.

ISBN: 978-989-95079-6-8

[23] A.Mesiarová, A note on two open problems of Alsina, Frank and Schweizer. *Aequationes Math.*, 72, no. 1-2, Pages 41-46, 2006.

[24] A. Mesiarová, $k - l_p$-Lipschitz t-norms. *Int. J. Approximate Reasoning*, 46, Pages 596-604, 2007.

[25] P.S. Mostert, A.L. Shields, On the structure of semigroups on a compact manifold with boundary, *Ann. of Math.*, 65, Pages 117-143, 1957.

[26] N. R. Pal and S. K. Pal, A review of image segmentation techniques, *Pattern recognition*, 26, Pages 1277-1294, 1993.

[27] E.H. Ruspini, A new approach to clustering, *Information and Control*, 15, Pages 22-32, 1969.

[28] L.A. Zadeh, Fuzzy sets, *Information Control*, 8, Pages 338-353, 1965.

[29] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, 1,Pages 3–28, 1978.

ISBN: 978-989-95079-6-8

On 'family resemblances' with fuzzy sets*

E. Trillas, C. Moraga[1] A. Sobrino[2]

[1] European Centre for Soft Computing
Mieres (Asturias), Spain
[2] University of Santiago de Compostela
Santiago de Compostela, Spain
Email: {enric.trillas, claudio.moraga}@softcomputing.es, alejandro.sobrino@usc.es

Abstract— *This paper takes into account the Wittgenstein's idea on family resemblances as a particular crisp relation between some fuzzy sets, that is, between some predicates representable from its use. It is shown that all uses of the same predicate actually do have family resemblance, that some pairs of predicates cannot, and a numerical degree of family resemblance is introduced.*

Keywords— Degree of family resemblance, Family resemblance, Fuzzy sets.

1 Introduction

Ludwig Wittgenstein headed two influential traditions in the so-called philosophy of language, that were originated by his famous books, *Tractatus logico-philosophicus* (1922, [1]), and *Philosophical Investigations* (1953, [2]), respectively.

The *Tractatus* does not properly deal with ordinary language, but with the logical analysis of propositions built up from atomic propositions, considered as 'pictures' of facts and keeping a strict correspondence with the world, understood through the totality of facts. Instead, the *Philosophical Investigations* in which Wittgenstein abandoned logical analysis, meant a shift in conferring a main role to the ways of designating facts, as an activity-oriented perspective on language. What is central at this respect, is that language does not primarily consist on describing the facts, but on playing 'language's games', or ways of dynamically using words to define their meaning, and that are to be described. In order to fix the meaning of a word, to show how a word works, it should be placed in the context and environment it is used. Language is not yet defined through propositions that, now, in Wittgenstein's view, come from their function in a language's game, and to note the absence of boundaries for describing such use of words, Wittgenstein introduced the term 'family resemblances'. Of course, Wittgenstein's idea on 'family resemblances', is broader than the relation of *family resemblance* in next section.

Fuzzy logic manages the extensional meaning of predicates through its use, once captured by the corresponding membership functions. This paper is nothing else than a first approach, in the path towards Zadeh's *Computing With Words*, to introduce 'family resemblances' between full-normalized fuzzy sets, and a numerical degree for measuring such relation.

Usually, full-normalized fuzzy sets, $\mu_P \in [0,1]^X$, those for which there are $x, y \in X$ such that $\mu_P(x) = 1$, $\mu_P(y) = 0$, appear as 'data' in the modeling of fuzzy systems, and they are neither self-contradictory ($\mu_P \leq \mu'_P$), nor negatively self contradictory ($\mu'_P \leq \mu_P$). Of course, μ'_P denotes the fuzzy set corresponding to 'not p'. Points x verifying $\mu_P(x) = 1$ can be taken as *the prototypes of* P *in* X, and points y verifying $\mu_P(y) = 0$ as *the anti-prototypes of* P *in* X.

In its own nature, this paper is not a conclusive one, but only a tentative to reflect the potentially interesting subject of the family resemblances shown by 'data' fuzzy sets. That is, by fuzzy sets with prototypes and anti-prototypes.

2 Family resemblance of fuzzy sets

Let $P, Q, ...$ be predicates on a universe of discourse X, such that their use, or meaning, is described by fuzzy sets $\mathbb{P}, \mathbb{Q}, ...$, given by membership functions $\mu_P, \mu_Q, ...$ in $[0,1]^X$.
For each membership function μ in $[0,1]^X$, define the sets of

- its 0-points, $Z(\mu) = \{x \in X; \mu(x) = 0\}$
- its 1-points, $S(\mu) = \{x \in X; \mu(x) = 1\}$

Definition 2.1 *With* $X \subset \mathbb{R}$*, the relation of family resemblance,* **fr** $\subset [0,1]^X \times [0,1]^X$*, is defined by* $(\mu, \sigma) \in$ **fr** *if and only if,*

1. $Z(\mu) \cap Z(\sigma) \neq \emptyset$, $S(\mu) \cap S(\sigma) \neq \emptyset$
2. μ *is non-decreasing in* $A \subset X$ *iff* σ *is non-decreasing in* A.
3. μ *is decreasing in* $A \subset X$ *iff* σ *is decreasing in* A.

In this definition, both 'decreasing' and 'non-decreasing', are not in strict sense, but allowing some constant pieces that can be taken as the first, or the second, by following what happens before or after.

Notice that the binary relation **fr** is only predicable between full-normalized fuzzy sets, that is, such that $Z(\mu) \neq \emptyset$, and $S(\mu) \neq \emptyset$. Denote

$$\mathfrak{F}^*(X) = \{\mu \in [0,1]^X - \{0,1\}^X; Z(\mu) \neq \emptyset, S(\mu) \neq \emptyset\}.$$

*This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain), and CICYT (Spain) under project TIN2008-06890-C02-01

ISBN: 978-989-95079-6-8

It should be pointed out that fuzzy sets μ in $\{0,1\}^X$ (crisp sets) are excluded, since such functions μ are neither decreasing, nor non-decreasing, but only piecewise constant with the values 0 or 1. In addition, in many cases if $(\mu,\sigma) \in \mathbf{fr}$, with $\mu,\sigma \in \{0,1\}^X$, it should be $\mu = \sigma$. Obviously, also constant fuzzy sets μ_r ($\mu_r(x) = r$, with $r \in [0,1], x \in X$) are not in $\mathfrak{F}^*(X)$, since $Z(\mu_r) = \emptyset$, or $S(\mu_r) = \emptyset$.
Of course, if σ results from a translation of μ keeping $Z(\mu) \cap Z(\sigma) \neq \emptyset$, and $S(\mu) \cap S(\sigma) \neq \emptyset$, it is obvious that $(\mu,\sigma) \in \mathbf{fr}$.

Remark 2.2 *Any non full-normalized fuzzy set, represented by a non-constant but continuous membership function $\mu \in [0,1]^{\mathbb{R}}$, can be re-scaled to a full-normalized one μ^* by*

$$\mu^* = \frac{\mu - \min(\mu)}{\max(\mu) - \min(\mu)}$$

Obviously, it is $Z(\mu^*) \neq \emptyset, S(\mu^*) \neq 0$, that is $\mu^* \in \mathfrak{F}^*(\mathbb{R})$, and μ^* is non-decreasing (decreasing) if and only if μ is non-decreasing (decreasing), but μ^* and μ can differ in the respective slopes (provided they have derivatives, it is $\mu^{*\prime} = \mu' / \max(\mu) - \min(\mu)$.

Example 2.3

1. *Fuzzy sets μ,σ in figure 1 verify $(\mu,\sigma) \in \mathbf{fr}$, since $Z(\mu) \cap Z(\sigma) = [0,2] \neq \emptyset$, and $S(\mu) \cap S(\sigma) = \{10\} \neq \emptyset$, and both are non-decreasing in $X = [0,10]$.*
2. *Fuzzy set μ in figure 1, and fuzzy set λ in figure 2, verify $(\mu,\lambda) \notin \mathbf{fr}$, since $S(\mu) \cap S(\lambda) = \emptyset$, and λ is decreasing in $[0,2]$, but μ is not.*

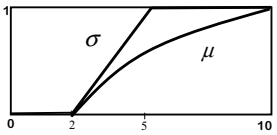


Figure 1:

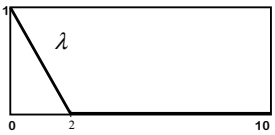


Figure 2:

Theorem 2.4 **fr** *is a reflexive and symmetric relation in $\mathfrak{F}^*(X)$.*

Proof. It is immediate to check that **fr** does verify the reflexive and symmetric properties. □

As it is intuitive, crisp relation **fr** is not transitive. A counterexample is given by membership functions μ,σ,λ in the figures

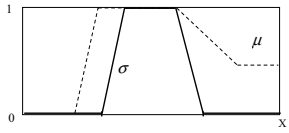


Figure 3:

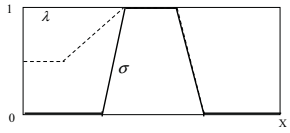


Figure 4:

that verify $(\mu,\sigma) \in \mathbf{fr}$, and $(\sigma,\lambda) \in \mathbf{fr}$, but $(\mu,\lambda) \notin \mathbf{fr}$, since $Z(\mu) \cap Z(\lambda) = \emptyset$.

Hence, if sets $[\mu] = \{\sigma \in \mathfrak{F}^*(X); (\mu,\sigma) \in \mathbf{fr}\}$ are not empty and cover $\mathfrak{F}^*(X)$, because it is $\bigcup_{\mu \in \mathfrak{F}^*(X)} [\mu] = \mathfrak{F}^*(X)$, they do not give a partition of $\mathfrak{F}^*(X)$, and the quotient $\mathfrak{F}^*(X)/\mathbf{fr}$ does not exist. For example, in the above figures, it is $\sigma \in [\mu] \cap [\lambda']$. Anyway, $[\mu]$ can be called the *family of* $\mu \in \mathfrak{F}^*(X)$, and $\sigma \in [\mu]$ a *relative* of μ, although μ could have relatives in other families, like it happens in people's families.

Theorem 2.5 *For no complement, μ' of $\mu \in \mathfrak{F}^*(X)$, is $(\mu,\mu') \in \mathbf{fr}$.*

Proof. If for $x,y \in A$, $x \leq y$, and $\mu(x) \leq \mu(y)$, it is $\mu'(y) \leq \mu'(x)$. □

Theorem 2.6 *For no opposite, or antonym, $'\mu$ of $\mu \in \mathfrak{F}^*(X)$, given by a symmetry α in X, $'\mu = \mu \circ \alpha$, is $(\mu, {}'\mu) \in \mathbf{fr}$.*

Proof. If $x \leq y$, and $\mu(x) \leq \mu(y)$, it follows $\alpha(y) \leq \alpha(x)$, and $\mu(\alpha(y)) \leq \mu(\alpha(x))$, or $\mu \circ \alpha(y) \leq \mu \circ \alpha(x)$, that is ${}'\mu(y) \leq' \mu(x)$. (see [3]) □

Theorem 2.7 *Let $u : X \to X$ be a bijective mapping such that*

- *If $x \leq y$, then $u(x) \leq u(y)$*
- $u^{-1}(Z(\mu)) \subset Z(\mu)$
- $u^{-1}(S(\mu)) \subset S(\mu)$

for $\mu \in \mathfrak{F}^*(X)$, $X \subset \mathbb{R}$. It is $(\mu, \mu \circ u^{-1}) \in \mathbf{fr}$.

Proof. It is $Z(\mu) \cap Z(\mu \circ u^{-1}) \neq \emptyset$, since $x \in Z(\mu)$, or $\mu(x) = 0$, implies $u^{-1}(x) \in Z(\mu)$, or $\mu(u^{-1}(x)) = \mu \circ u^{-1}(x) = 0$, it is, $x \in Z(\mu \circ u^{-1})$. Analogously, $x \in S(\mu)$, or $\mu(x) = 1$, that is, $x \in S(\mu \circ u^{-1})$. Hence, $Z(\mu) \cap Z(\mu \circ u^{-1}) \neq \emptyset$, and $S(\mu) \cap S(\mu \circ u^{-1}) \neq \emptyset$. If μ is non-decreasing, '$x \leq y \Rightarrow \mu(x) \leq \mu(y)$, from $u^{-1}(x) \leq u^{-1}(y)$, follows $\mu(u^{-1}(x)) \leq \mu(u^{-1}(y))$. Analogously, if μ is decreasing, so it is $\mu \circ u^{-1}$. □

In particular, if Q represented by $\mu_Q = \mu_P \circ u^{-1}$ is a synonym of P (see[4]), then $(\mu_P, \mu_Q) \in \mathbf{fr}$.

Example 2.8

1. *Fuzzy sets μ,σ in figure 5 verify $\sigma = \mu \circ \alpha$, with the symmetry $\alpha(x) = 10 - x$, and represent two antonyms. Obviously, μ and $\mu \circ \alpha$ do not show family resemblance.*

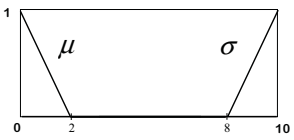


Figure 5:

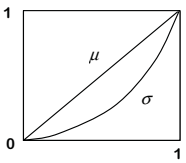


Figure 6:

2. *Fuzzy sets μ,σ in figure 6 verify $\sigma = \mu \circ u^{-1}$, with $u(x) = \sqrt{x}$, and show family resemblance.*

3 Degree of family resemblance

Obviously, if σ_1, σ_2 are in the family $[\mu]$, they do verify $(\mu, \sigma_1) \in$ **fr**, and $(\mu, \sigma_2) \in$ **fr**, but this does not mean that σ_1, σ_2 show the same extent of family resemblance with μ. A way of measuring such extent is by means of a convenient T-indistinguishability (see([5])) DR : **fr** $\rightarrow [0,1]$, defining,

If $(\mu, \sigma) \in$ **fr**, the degree up to which μ resembles σ is $DR(\mu, \sigma) \in [0,1]$.

3.1

Any prod-indistinguishability E in $\mathfrak{F}^*(X)$, can be represented by Ovchinnikov's theorem (see [6])

$$E(\mu,\sigma) = \inf_{f\in F}(\frac{f(\mu)}{f(\sigma)}, \frac{f(\sigma)}{f(\mu)}),$$

with F a family of functions $f : \mathfrak{F}^*(X) \rightarrow \mathbb{R}^+ - \{0\}$.

Provided X is a closed interval in $\mathbb{R}$, consider only the functions $\mu \in \mathfrak{F}^*(X)$ that are Riemann-integrable in X, and such that both $Z(\mu)$ and $S(\mu)$ are intervals in X. Denote by $\mathfrak{F}^{**}(X)$ this subset of functions in $\mathfrak{F}^*(X)$, and take

$$f_1(\mu) = \int_X \mu(x)dx, \quad f_2(\mu) = \text{length of the interval } X - Z(\mu).$$

Then,

$$Ovch(\mu,\sigma) = \min(\frac{f_1(\mu)}{f_1(\sigma)}, \frac{f_1(\sigma)}{f_1(\mu)}, \frac{f_2(\mu)}{f_2(\sigma)}, \frac{f_2(\sigma)}{f_2(\mu)}),$$

for all $\mu, \sigma \in \mathfrak{F}^{**}(X)$, is a prod-indistinguishability in the part of **fr** in $\mathfrak{F}^{**}(X) \times \mathfrak{F}^{**}(X)$.

3.2

If $X = [a,b] \subset \mathbb{R}$, define

$$I(\mu,\sigma) = 1 - \frac{1}{b-a}|\int_X \mu(x)dx - \int_X \sigma(x)dx|,$$

for all $\mu, \sigma \in \mathfrak{F}^{**}(X)$.

Function I is a W-indistinguishability (W is the Łukasiewicz t-norm), since:

- $I(\mu,\mu) = 1$, and $I(\mu,\sigma) = I(\sigma,\mu)$, for all $\mu, \sigma \in \mathfrak{F}^{**}(X)$.
- $W(I(\mu,\sigma), I(\sigma,\lambda)) = \max(0, 1 - \frac{1}{b-a}[|\int_X \mu(x)dx - \int_X \sigma(x)dx| + |\int_X \sigma(x)dx - \int_X \lambda(x)dx|]) \leq \max(0, 1 - \frac{1}{b-a}|\int_X \mu(x)dx - \int_X \lambda(x)dx|) = 1 - \frac{1}{b-a}|\int_X \mu(x)dx - \int_X \lambda(x)dx| = I(\mu,\lambda)$, for all $\mu, \sigma, \lambda \in \mathfrak{F}^{**}(X)$, because of the triangular inequality, $|\int_X \mu(x)dx - \int_X \lambda(x)dx| \leq |\int_X \mu(x)dx - \int_X \sigma(x)dx| + |\int_X \sigma(x)dx - \int_X \lambda(x)dx|$.

3.3

Since it is $Ovch(\mu,\sigma) \neq I(\mu,\sigma)$ and, in general, $Ovch(\mu,\sigma) < I(\mu,\sigma)$, a better degree could be obtained with a mean like,

$$DR(\mu,\sigma) = \frac{r_1 Ovch(\mu,\sigma) + r_2 I(\mu,\sigma)}{r_1 + r_2},$$

$$\text{with } r_1, r_2 \in \mathbb{R}^+ - \{0\},$$

of which the only DR that is symmetric is the one with $r_1 = r_2 = \frac{1}{2}$ (arithmetic mean). In addition, and since $W \leq prod$, $Ovch$ is also W-transitive: $W(Ovch(\mu,\sigma), Ovch(\sigma,\lambda)) \leq Ovch(\mu,\sigma) \cdot Ovch(\sigma,\lambda) \leq Ovch(\mu,\lambda)$. Thus, also DR with $r_1 = r_2 = \frac{1}{2}$ is W-transitive, since:

$$W(DR(\mu,\sigma), DR(\sigma,\lambda)) =$$

$$\max(0, \frac{Ovch(\mu,\sigma) + Ovch(\sigma,\lambda) - 1}{2} +$$

$$\frac{I(\mu,\sigma) + I(\sigma,\lambda) - 1}{2}) \leq$$

$$\max(0, \frac{Ovch(\mu+\lambda)}{2} + \frac{I(\mu+\lambda)}{2}) \leq$$

$$\max(0, DR(\mu,\lambda)) = DR(\mu,\lambda),$$

because $Ovch(\mu,\sigma) + Ovch(\sigma,\lambda) - 1 \leq \max(0, Ovch(\mu,\sigma) + Ovch(\sigma,\lambda) - 1)$, and $I(\mu,\sigma) + I(\sigma,\lambda) - 1 \leq \max(0, I(\mu,\sigma) + I(\sigma,\lambda) - 1)$. Hence, we will take

$$DR(\mu,\sigma) = \frac{Ovch(\mu,\sigma) + I(\mu,\sigma)}{2},$$

as the *index of family resemblance*.

Remark 3.1 *Obviously, functions DR are not only applicable to pairs in* **fr***, but also to all pairs in $\mathfrak{F}^{**}(X) \times \mathfrak{F}^{**}(X)$*

3.4

Let E be a T-indistinguishability relation. The crisp relation

$$\mu \equiv \sigma \Leftrightarrow E(\mu,\sigma) > 0,$$

is reflexive, and symmetric. Obviously, and provided $T = \min$, or $T = prod$, it is also transitive. Nevertheless, if $T = W$, since "$E(\mu,\sigma) > 0$ and $E(\sigma,\lambda) > 0$" is equivalent to the existence of $\varepsilon > 0$ such that "$E(\mu,\sigma) \geq \varepsilon$ and $E(\sigma,\lambda) \geq \varepsilon$", it is

$$W(E(\mu,\sigma), E(\sigma,\lambda)) \geq W(\varepsilon,\varepsilon) = \max(0, 2\varepsilon - 1), \text{ or}$$

$$E(\mu,\lambda) \geq \max(0, 2\varepsilon - 1).$$

To have $E(\mu,\lambda) > 0$, it is sufficient that $2\varepsilon - 1 > 0$, or $\varepsilon > 0.5$. In this case, $\mu \equiv \sigma$ and $\sigma \equiv \lambda$ imply $\mu \equiv \lambda$ (transitivity). Hence $E(\mu,\sigma) > 0.5$ allows to take μ and σ as 'equivalent'.

If $0.5 < Ovch(\mu,\sigma)$, and $0.5 < I(\mu,\sigma)$, it is $\mu \equiv \sigma$ for both T-indistinguishabilities $E = Ovch$ and $E = I$. Thus, if in addition to $(\mu,\sigma) \in$ **fr**, is $0.5 < \varepsilon \leq Ovch(\mu,\sigma)$, and $0.5 < \delta \leq I(\mu,\sigma)$, it is also $0.5 < \frac{\varepsilon+\delta}{2} \leq DR(\mu,\sigma)$, implying $\mu \equiv \sigma$ for $E = DR$. For example if $0.7 \leq Ovch(\mu,\sigma)$, and $0.7 \leq I(\mu,\sigma)$, it is $0.7 \leq DR(\mu,\sigma)$. In these cases, *it can be said that μ and σ show high family resemblance.*

ISBN: 978-989-95079-6-8

When $(\mu,\sigma) \in$ **fr**, if e.gr., $Ovch(\mu,\sigma) \geq 0.8$ and $I(\mu,\sigma) \geq 0.8$ implying $DR(\mu,\sigma) \geq 0.8$, there is a so high family resemblance between μ and σ that μ and σ can be taken as strongly representing similar uses of the same predicate (linguistic label).

Example 3.2

1. *Functions μ,σ in figure 7, do obviously verify $(\mu,\sigma) \in$* **fr**. *For them, $f_1(\mu) = 5$, $f_2(\mu) = 8$, $f_1(\sigma) = 4$ and $f_2(\sigma) = 7$,*

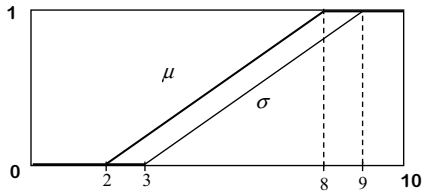


Figure 7:

hence, $Ovch(\mu,\sigma) = \min(\frac{5}{4}, \frac{4}{5}, \frac{8}{7}, \frac{7}{8}) = 0.8$.
Since, $I(\mu,\sigma) = 1 - \frac{1}{10}|5-4| = 0.9$, it results

$$DR(\mu,\sigma) = \frac{0.8+0.9}{2} = 0.85,$$

μ *and* σ show *high family resemblance.*

2. *Functions μ,σ in figure 8, verify $(\mu,\sigma) \in$* **fr**. *For then, $f_1(\mu) = 4 + 2{\cdot}0.5 = 5$, $f_2(\mu) = 6$, $f_1(\sigma) = 2$, $f_2(\sigma) = 4$,*

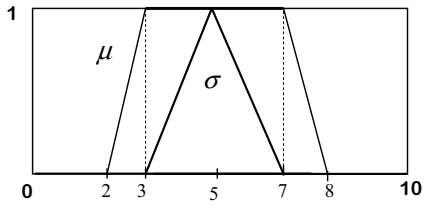


Figure 8:

hence, $Ovch(\mu,\sigma) = \min(\frac{5}{2}, \frac{2}{5}, \frac{6}{4}, \frac{4}{6}) = \frac{2}{5} = 0.4$.
Since, $I(\mu,\sigma) = 1 - \frac{1}{10}(|5-2|) = 0.7$, it results

$$DR(\mu,\sigma) = \frac{0.4+0.7}{2} = 0.55,$$

μ and σ *do not show* high family resemblance.

3. *Functions μ,σ in figure 9, verify $(\mu,\sigma) \in$* **fr**. *For them $f_1(\mu) = 5+1 = 6$, $f_2(\mu) = 7$, $f_1(\sigma) = 4+2 = 6$ and $f_2(\sigma) = 8$, hence, $Ovch(\mu,\sigma) = \min(\frac{6}{6}, \frac{7}{8}, \frac{8}{7}) = \frac{7}{8} = 0.88$.*

Since $I(\mu,\sigma) = 1 - \frac{1}{10}(|6-6|) = 1$, it results

$$DR(\mu,\sigma) = \frac{0.88+1}{2} = 0.94,$$

μ and σ *show* high family resemblance.

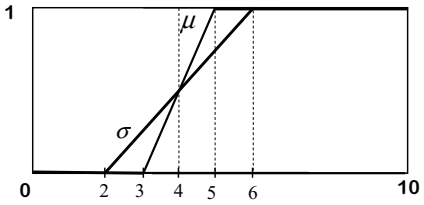


Figure 9:

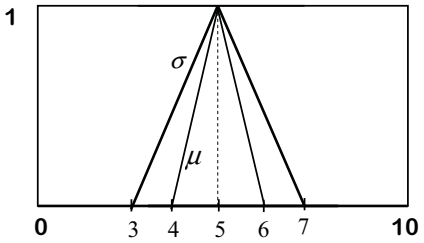


Figure 10:

4. *Functions μ,σ in figure 10, verify $(\mu,\sigma) \in$* **fr**. *For them $f_1(\mu) = 1$, $f_2(\mu) = 2$, $f_1(\sigma) = 2$ and $f_2(\sigma) = 4$,*
hence, $Ovch(\mu,\sigma) = \min(\frac{1}{2}, \frac{2}{1}, \frac{2}{4}, \frac{4}{2}) = 0.5$.
Since $I(\mu,\sigma) = 1 - \frac{1}{10}(2-1) = 0.9$, it results

$$DR(\mu,\sigma) = \frac{0.5+0.9}{2} = 0.7,$$

μ and σ *show* high family resemblance.

5. *Functions μ,σ in figure 11, verify $(\mu,\sigma) \in$* **fr**. *For them $f_1(\mu) = 3 + 2{\cdot}0.5 + 1 = 5$, $f_2(\mu) = f_2(\sigma) = 10$ and $f_1(\sigma) = 0.5 + 3{\cdot}0.5 + \frac{3{\cdot}0.5}{2} + 5 = 7.75$, hence, $Ovch(\mu,\sigma) = \frac{5}{7.75} = 0.64$.*

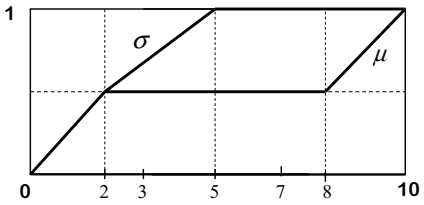


Figure 11:

Since $I(\mu,\sigma) = 1 - \frac{1}{10}(7.75-5) = 0.725$, it results

$$DR(\mu,\sigma) = \frac{0.64+0.725}{2} = 0.6825,$$

μ and σ *do not show* high family resemblance.

6. *Functions μ,σ in figure 12, verify $(\mu,\sigma) \in$* **fr**. *It is:*
$f_1(\mu) = 3$, $f_2(\mu) = 4$, $f_1(\sigma) = 3$ *and* $f_2(\sigma) = 5$.*Then,*

$$Ovch(\mu,\sigma) = \frac{4}{5}, \;\; I(\mu,\sigma) = 1$$

That is,

$$DR(\mu,\sigma) = 0.9,$$

and μ, σ *show* high family resemblance.

ISBN: 978-989-95079-6-8

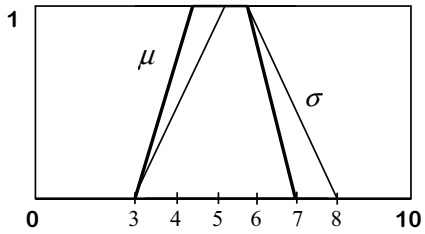


Figure 12:

4 The case of membership functions with the same predicate.

Given a predicate, or linguistic label, P on X, there is not a single fuzzy set generated by P, but each use of P in X, once reflected by its membership function μ_P (see [4]), defines a fuzzy set $\mathbb{P}$ by

$$x \in_r \mathbb{P} \Leftrightarrow \mu_P(x) = r, \text{ for all } x \in X, r \in [0,1].$$

Nevertheless, all functions μ_P do have some 0-points, and some 1-points, in common, as well as they should either decrease or non-decrease simultaneously in the same parts of X. Hence, *all μ_P should show some degree of family resemblance.*

This is the case for example, with P =*small* in $X = [0, 10]$, if two of its uses are represented by the membership functions μ_P^1 and μ_P^2 in figure 13.

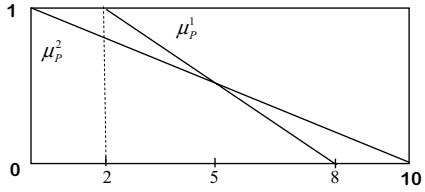


Figure 13:

It is clear that $(\mu_P^1, \mu_P^2) \in$ **fr**, and:

- $Ovch(\mu_P^1, \mu_P^2) = \min(\frac{5}{5}, \frac{10}{8}, \frac{8}{10}) = 0.8$
- $I(\mu_P^1, \mu_P^2) = 1 - \frac{1}{10}(|5-5|) = 1.$

  Hence,

$$DR(\mu_P^1, \mu_P^2) = \frac{0.8+1}{2} = 0.9,$$

  and μ, σ *show high family resemblance*.

Examples 1 and 4 in section 3, do correspond to uses of P =*big* and P =*around five*, respectively.

Remark 4.1 *In natural language, families $[\mu]$ must be 'open', but not 'closed' like they were defined in section 2. These families are here static (sets), but in natural language they should be dynamical. With time, an element σ that was not in $[\mu]$, but possibly with a not too low degree $DR(\sigma, \mu)$, could be included in $[\mu]$ and thus generating a new family, indeed, changing the relation* **fr**. *Actually,* **fr** *is not a permanent relation, in natural language it is a changing one. Only with families of resemblance taken as classical sets, human thought seems to be impossible (see [7]).*

5 Last remarks

5.1

In his *Philosophical Investigations*, Wittgenstein conceived language in a way close to how people manages it. Thus, the meaning of an imprecise predicate is not given by necessary and sufficient conditions, but is built up by similarity with its prototypes in the universe of discourse, like in the case of *big* in [0,10] with, at least, the prototype 10. Notwithstanding, for more complex predicates like P =*beautiful* in a set of art's objects, it could be not clear the existence of prototypes and, since in such cases it could be $S(\mu_P) = \emptyset$, the study of the family resemblance for these predicates remains an open problem.

5.2

Certainly, to consider $X \subset \mathbb{R}$ is a restriction for this paper's results. Nevertheless, most of the predicates fuzzy logic considers are those exhibiting a numerical characteristic, allowing to translate them into an interval in $\mathbb{R}$. This is the case, for example, of $P =$ *tall* in a big population X, that is translated into the interval $[0, 2]$, in meters, by the numerical characteristic 'Height', and the predicate $Q = big$ in such interval. That is, by the identification $\mu_{tall} = \mu_{big} \circ Height$, or $\mu_{tall}(x) = \mu_{big}(Height(x))$, for all x in X, once the use of *big* in $[0, 2]$ is chosen accordingly with that of *tall* interpreted, as it turn, by *big height*.

There are not so obvious cases that, also, can be translated into an interval through a more complex process, consisting in identifying each x in X with an n-tuple of significative parts in x, $x := (x_1, \ldots, x_n)$, with numerical characteristics $Ch_i(x_i) \in [a_i, b_i]$, and taking $Ch(x) = A(Ch_1(x_1), \ldots, Ch_n(x_n)) \in [a, b]$, with some n-place aggregation function A. An example is given by P= *beautiful* in a set X of paintings, if each painting x can be partitioned in $x = x_1 \cup \ldots \cup x_n (x_i \cap x_j = \emptyset, i \neq j)$, allowing to interpret the statement 'x is P' as the composite one (x_1 is P,$\ldots, x_n$ is P). Provided each component 'x_i is P' is numerically evaluable by a clearly explicitable characteristic $Ch_i(x_i) \in \mathbb{R}^+$, the values of μ_P could be obtained by

$$\mu_P(x) = \frac{r_1 Ch_1(x_1) + \ldots + r_n Ch_n(x_n)}{r_1 + \ldots + r_n} (r_i \geq 0),$$

that are in the interval $[\min \mu_P, \max \mu_P]$.

Of course, $A(a_1, \ldots, a_n) = \frac{r_1 a_1 + \ldots + r_n a_n}{r_1 + \ldots + r_n}$, is not the only way of reasonably aggregating the n-tuples $(Ch_1(x_1), \ldots, Ch_n(x_n)) \in \mathbb{R}^n$. For example, if '$x$ is P' is interpretable by 'x_1 is P' *and* $\ldots$ *and* 'x_n is P', A can be taken as a t-norm, provided $Ch_i(x_i) \in [0, 1], 1 \leq i \leq n$.

6 Conclusions

6.1

The concept of family resemblances is here introduced as a crisp-binary relation for only a particular type of fuzzy sets in the real line, with 0-points and 1-points. Even more restrictive is the class of Riemann-integrable membership functions, to which pairs a degree of family resemblance is assigned. Anyway, this paper should be viewed as only a first trial to consider Wittgenstein's idea on family resemblances with fuzzy

ISBN: 978-989-95079-6-8

sets . For example, the definition of **fr** could be, perhaps, extended to other fuzzy sets without 1-points (non-normalized fuzzy sets), or without 0-points, as well as to pairs (μ, σ) with either μ, or σ in $\{0,1\}^X$. That is, to non-data fuzzy sets, to fuzzy sets resulting from computations with data ones. In addition, the selected degree DR cannot be considered as the definitive definition to measure to what extent there is family resemblance. More study on the subject is deserved since, for example, DR can be applied to pairs $(\mu, \sigma) \notin$ **fr**, and more accurate values can be obtained in some cases with non symmetric functions $DR(\mu,\sigma) = \frac{r_1 \cdot Ovch(\mu,\sigma) + r_2 \cdot I(\mu,\sigma)}{r_1 + r_2}$.

6.2

Of course, definition 2.1 could also be applied to crisp sets by simply allowing either μ or σ to only have 'constant pieces'. At this respect, examples like the one in figure 14 are of some interest. In such figure, μ is the trapezoidal fuzzy set $(3.8, 4, 6, 6.2)$, and σ does represent the crisp subset $[4, 6]$.

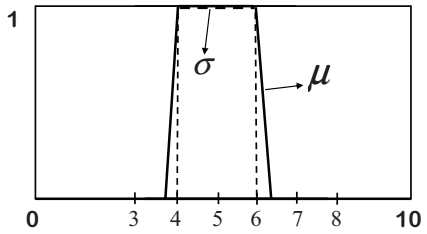


Figure 14:

Since $Ovch(\mu,\sigma) = 0.83$, $I(\mu,\sigma) = 0.98$, it follows $DR(\mu,\sigma) = 0.90$, a big value suggesting a so high resemblance between μ and σ that, perhaps, could allow to identify μ with μ_P, $P =$ *Almost between 4 and 6*. That is, μ could be 'linguistically approached' by the imprecise predicate P.

6.3

Maybe what is here introduced could be useful to approach the (pending) problem of 'linguistic approximation'. Namely, given the output $f : X \to [0,1]$ of a system of fuzzy rules, how to find a predicate P in X such that f could be identified with μ_P? A possible way could come once identified a membership function μ_P (representing a use of the known predicate P), such that, $DR(f, \mu_P) > \varepsilon$, and $Sup|f - \mu_P| < \delta$ (for fixed $\varepsilon > 0, \delta > 0$), even if $(f, \mu_P) \notin$ **fr**. Then 'P' can be called an (ε, δ)-linguistic approximation of f. For example, in 6.2, *Almost between 4 and 6* is an $(0.89, 0.2)$-linguistic approximation of μ.

6.4

In this paper, the threshold 0.7 is taken for the sake of illustrating the idea of 'high family resemblance'. Nevertheless, and although 0.7 seems to be a good enough value for the examples shown in figures 7 to 14, such threshold's value still remains to be studied.

Acknowledgment

The authors express their thanks to the two anonymous reviewers for their interesting and constructive remarks.

References

[1] L. Wittgenstein. *Tractatus logico-philosophicus*. Routledge & Kegan Paul, London, 1961.

[2] L. Wittgenstein. *Philosophical Investigations*. Basil Blackwell, London, 1973.

[3] E. Trillas, C. Moraga, S. Guadarrama, S. Cubillo, and E. Castiñeira. Computing with Antonyms. In *Forging New Frontiers: Fuzzy Pioneers I*, volume 217 of *Studies in fuzziness and soft computing*, pages 133–153. Springer, Berlin, 2007.

[4] E. Trillas. On a model for the meaning of predicates. A naïve approach to the genesis of fuzzy sets. In Rudolf Seising, editor, *Fuzzy sets theory. Philosophy and criticism*, Studies in fuzziness and soft computing. Springer Verlag, forthcoming.

[5] L. Valverde E. Trillas. An inquiry into indistinguishability operators. *Aspects of Vagueness. H.J. Skala, S. Termini, E. Trillas (Eds.)*, pages 231–256. Reidel, Dordrecht, 1984.

[6] S.V. Ovchinnikov. Representation of transitive fuzzy relations. *Aspects of Vagueness. H.J. Skala, S. Termini, E. Trillas (Eds.)*, pages 105–118. Reidel, Dordrecht, 1984.

[7] S. Pinker. *Words and Rules*. Basic Books, N.Y., 1999.

ISBN: 978-989-95079-6-8

Convergence theorems for generalized random variables and martingales

Andrew L Pinchuck

Department of Mathematics (Pure and Applied),

Rhodes University, PO Box 94,Grahamstown, 6140,

South Africa

Email: a.pinchuck@ru.ac.za.

Abstract— We examine generalizations of random variables and martingales. We prove a new convergence theorem for set-valued martingales. We also generalize a well known characterization of set-valued random variables to the fuzzy setting.

Keywords— Banach space, fuzzy, martingale, random variable, set-valued.

1 Set-valued random variables and martingales

The theory of conditional expectations and martingales has been established for Banach space-valued, Bochner-integrable functions. Hiai and Umegaki have then generalized random variables, conditional expectations and martingales to the set-valued (multivalued) setting in [1]. We will assume that the reader is familiar with the basic ideas of random variables and provide a brief introduction to the set-valued generalization and also to the related concept of conditional expectations which has played an important role in probability theory, ergodic theory and quantum statistical mechanics. We also contribute to the theory of set-valued martingales and since martingales are important in probability theory, we feel that there are potential applications to be developed from this work. This paper can be divided into two main results. The first result is an original convergence theorem for set-valued martingales. The second result is a generalization of a useful characterization of measurable set-valued random variables by Hiai and Umegaki in the fuzzy setting. The original theorem by Hiai and Umegaki was an essential tool used to obtain many of the subsequent results in [1] and it is clear that we can apply this new generalized theorem in an analagous way in the fuzzy setting.

We will be again considering a measure space (Ω, Σ, μ), a separable Banach space $(X, \|\cdot\|)$ with K a field of scalars (either $\mathbb{R}$ or $\mathbb{C}$). We remind the reader that functions f, g from Ω into X are said to be equal *almost everywhere*, denoted $f = g$ a.e. (μ) or $f(\omega) = g(\omega)$ $(\forall \omega \in \Omega)$ a.e. (μ) if $f(\omega) = g(\omega)$ for all $\omega \in \Omega$ except on a set of measure zero. We can omit reference to μ and Ω if there is no confusion. That is $f = g$ a.e. if $\mu(\{\omega \in \Omega : f(\omega) \neq g(\omega)\}) = 0$. We will denote by $M[\Omega, X]$ and $L^p[\Omega, X]$ the collections of measurable and p-integrable functions $f : \Omega \to X$ respectively, for $1 \leq p \leq \infty$. Due to the completeness of X we have several characterizations of this class of functions (see [1] Theorem 1.0). As in [1], $K(X)$ shall denote the collection of nonempty compact subsets of X and $K_k(X)$ the collection of convex sets in $K(X)$. We denote the collection of (weakly) measurable set-valued functions $F : \Omega \to K_k(X)$ by $\mathcal{M}[\Omega, K_k(X)]$. That is $F \in \mathcal{M}[\Omega, K_k(X)]$ if and only if for all O open, $O \subset X$, $F^{-1}(O) \in \Sigma$. Let p be chosen such that $1 \leq p \leq \infty$ then we denote by $\mathcal{L}^p[\Omega, K_k(X)] = \mathcal{L}^p[\Omega, \Sigma, \mu, K_k(X)]$ the space of all p-integrable functions in $\mathcal{M}[\Omega, K_k(X)]$, where two functions $F_1, F_2 \in \mathcal{L}^1[\Omega, K_k(X)]$ will be considered identical if $F_1(\omega) = F_2(\omega)$ a.e.

The topology on $K_k(X)$ is induced by the Hausdorff metric defined by

$$d_H(A, B) = \max\Big\{ \sup_{a\in A} \inf_{b\in B} \|a - b\|, \sup_{b\in B} \inf_{a\in A} \|a - b\| \Big\}, \tag{1.1}$$

for $A, B \subseteq X$. It is well known $(K_k(X), d_H)$ is a complete separable metric space. It now follows that in this topology $(F_n)_{n\in N} \subseteq \mathcal{M}[\Omega, K_k(X)]$ then $F_n \longrightarrow F$ as $n \longrightarrow \infty$ if and only if $d_H(F_n(\omega), F(\omega)) \longrightarrow 0$ for all $\omega \in \Omega$ a.e.

We can consider $(X, \|\cdot\|)$ to be a topological space since a norm defines a natural metric on X which induces the *metric topology* on X. We define the *closure* of a set $A \subseteq X$, denoted $\mathrm{cl}(A)$ to be the *norm closure* of A. That is, the smallest closed set containing A with respect to the metric topology induced by the norm $\|\cdot\|$.

A measurable function $f : \Omega \to X$ is called a *measurable selection* of F if $f(\omega) \in F(\omega)$ for all $\omega \in \Omega$ a.e. That is f is a measurable selection of F if $f(\omega) \in F(\omega)$ for all $\omega \in \Omega$ except on a set of measure zero. We define the set $S^p_F = \{f \in L^p[\Omega, X] : f(\omega) \in F(\omega)\ a.e.\}$. It is easy to show that S^p_F is a closed subset of $L^p[\Omega, X]$ for $1 \leq p \leq \infty$.

Lemma 1.1 *[1] Let $F \in \mathcal{M}[\Omega, K_k(X)]$ and $1 \leq p \leq \infty$. If S^p_F is nonempty, then there exists a sequence $(f_i)_{i\in N}$ contained in S^p_F such that $F(\omega) = \mathrm{cl}\{f_i(\omega)\}$ for $\omega \in \Omega$.*

Corollary 1.2 *[1] Let $F_1, F_2 \in \mathcal{M}[\Omega, K_k(X)]$ and $1 \leq p \leq \infty$. If $S^p_{F_1} = S^p_{F_2} \neq \emptyset$ then $F_1(\omega) = F_2(\omega)$, for all $\omega \in \Omega$.*

A brief discussion of the generalization of conditional expectations is necessary before we reach our first main result. This material is covered more comprehensively in [1].

Let W be a sub-σ-algebra of Σ,

$$S^1_F(W) = \{f \in L^1[\Omega, W, \mu, X] : f(\omega) \in F(\omega)\, a.e.\}. \tag{1.2}$$

We define

ISBN: 978-989-95079-6-8

$$\int_{\Omega}^{(W)} F d\mu = \Big\{ \int_{\Omega} f d\mu : f \in S_F^1(W) \Big\}. \quad (1.3)$$

If $f \in L^1[\Omega, X]$ then the *conditional expectation* $E[f|W]$ *of* f *relative to* W is defined as a W-measurable function $E[f|W] \in L^1[\Omega, W, \mu, X]$ such that

$$\int_A E[f|W] d\mu = \int_A f d\mu, \quad (1.4)$$

for all $A \in W$.

The following theorem provides a natural extension of the notion of a conditional expectation to the set-valued setting.

Theorem 1.3 *[1] Let $F \in \mathcal{L}^1[\Omega, K_k(X)]$. Then there exists a unique $E[F|W] \in \mathcal{L}^1[\Omega, W, \mu, K_k(X)]$ such that*

$$S^1_{E[F|W]}(W) = \text{cl}\{E[f|W] : f \in S_F^1\} \quad (1.5)$$

where the closure is taken with respect to the norm in $\mathcal{L}^1[\Omega, K_k(X)]$. $E[F|W]$ is called the conditional expectation of F relative to W.

The concept of martingale in probability theory was introduced by Paul Pierre Lévy. Part of the motivation for that work was to show the impossibility of successful betting strategies. A martingale was originally devised as an indexed sequence of random variables with the index representing time. If t is a later time than s then the idea is that the conditional expected value at time t given the same observations as at time s will be equal to the expected value at time s. The notion of a martingale has proved to be a useful tool in modeling various events in probability theory. We present a precise definition of the classical notion followed by the generalization to the set-valued setting.

We remind the reader that a *filtration* on (Ω, Σ, μ) is a sequence of σ-algebras $(\Sigma_n)_{n\in N}$ such that $\Sigma_n \subseteq \Sigma$ and $\Sigma_n \subseteq \Sigma_{n+1}$ for all $n \in \mathbb{N}$.

Definition 1.4 *If $(f_n)_{n\in N}$ is a sequence of random variables and $(\Sigma_n)_{n\in N}$ a filtration then the sequence $(f_n, \Sigma_n)_{n\in N}$ is called a martingale if we have for each $n \in \mathbb{N}$:*

(a) f_n is Σ_n-measurable and

$$\int_{\Omega}^{(\Sigma_n)} \|f_n\| d\mu < \infty, \quad (1.6)$$

(b)

$$E[f_{n+1}|\Sigma_n] = f_n. \quad (1.7)$$

Alternatively, if property (b) is replaced by

(b')

$$E[f_{n+1}|\Sigma_n] \geq f_n (E[f_{n+1}|\Sigma_n] \leq f_n) \quad (1.8)$$

then $(f_n, \Sigma_n)_{n\in N}$ is called a submartingale (supermartingale), respectively.

Definition 1.5 *Let $(F_n)_{n\in N}$ be a sequence of set-valued random variables and $(\Sigma_n)_{n\in N}$ a filtration then the sequence $(F_n, \Sigma_n)_{n\in N}$ is called a set-valued martingale if we have, for each $n \geq 1$:*

(a) F_n is Σ_n-measurable and

$$\int_{\Omega}^{(\Sigma_n)} \|F_n\| d\mu < \infty, \quad (1.9)$$

(b)

$$E[F_{n+1}|\Sigma_n] = F_n. \quad (1.10)$$

Alternatively, if property (b) is replaced by

(b')

$$E[F_{n+1}|\Sigma_n] \geq F_n (E[F_{n+1}|\Sigma_n] \leq F_n) \quad (1.11)$$

then $(F_n, \Sigma_n)_{n\in N}$ is called a fuzzy submartingale (supermartingale), respectively.

Theorem 1.6 *If $(F_n, \Sigma_n)_{n\in N}$ is a set-valued martingale and $1 \leq p \leq \infty$ then $F_n \longrightarrow F$ as $n \longrightarrow \infty$ if and only if for each choice of $f_n \in S^p_{F_n}$, $n \in N$, we have that*

$$\lim_{n\longrightarrow\infty} f_n \in S_F^p. \quad (1.12)$$

Proof: Let p be such that $1 \leq p \leq \infty$. For brevity, we will henceforth denote $\lim_{n\longrightarrow\infty} f_n(\omega)$ by $f(\omega)$ for each $\omega \in \Omega$. Now $f_n \in S^p_{F_n}$ for all $n \in \mathbb{N}$ but $f \notin S_F^p \iff$ for all $n \in \mathbb{N}$, there exists $A \in \Sigma$, $\mu(A) \neq 0$ such that $f(\omega) \notin F(\omega)$ for all $\omega \in A$. Now for a given $A \in \Sigma$ we have $\forall \omega \in A$, $f_n(\omega) \neq f(\omega) \iff \forall \omega \in A$, $f(\omega) - f_n(\omega) \neq 0 \iff \forall \omega \in A$, $\|f(\omega) - f_n(\omega)\| \geq \delta$ for some $\delta > 0$. Thus $f \notin S_F^p \iff \exists A \in \Sigma, \mu(A) \neq 0, \forall \omega \in A, \forall n \in \mathbb{N}$,

$$d_H(F(\omega), F_n(\omega)) =$$

$$\max\{ \sup_{f(\omega)\in F(\omega)} \inf_{f_n(\omega)\in F_n(\omega)} \|f(\omega) - f_n(\omega)\|,$$

$$\sup_{f_n(\omega)\in F_n(\omega)} \inf_{f(\omega)\in F(\omega)} \|f(\omega) - f_n(\omega)\| \}$$

$\geq \delta \iff d_H(F(\omega), F_n(\omega) \not\longrightarrow 0$ for all $\omega \in A \in \Sigma$, $\mu(A) \neq 0 \iff d_H(F, F_n) \not\longrightarrow 0 \iff F_n \not\longrightarrow F$ as $n \longrightarrow \infty$.

2 Fuzzy sets

We now provide a brief introduction to fuzzy sets with a particular focus on the tools that we will need in the subsequent section. These notions are all well known and standard. For further reading on the topic, the reader is referred to [2, 3, 4, 5, 8](by no means an exhaustive list). Throughout this paper, we will be considering fuzzy sets $A \in I^X$ where X is a Banach Space and I the unit interval $[0, 1]$.

For a given fuzzy set we associate collections of crisp subsets of X with it. This concept is highly useful in relating statements in the fuzzy setting to the classical setting.

ISBN: 978-989-95079-6-8

Definition 2.1 *If $A \in I^X$ and $\alpha \in I$ we define,*

$$A^\alpha = \{x \in X : A(x) > \alpha\}; \tag{2.13}$$

$$A_\alpha = \{x \in X : A(x) \geq \alpha\}. \tag{2.14}$$

These crisp sets are referred to as α-levels (or cuts), strong and weak respectively. We call $\text{supp}A = A^0$ *the support of A.*

We shall work mainly with the weak cuts in this paper although similar results can be obtained using the strong cuts.

Throughout this paper we will denote the characteristic function of a crisp set $A \subseteq X$ by χ_A. That is $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ elsewhere. We immediately obtain a useful characterization of fuzzy sets.

Theorem 2.2 *If $A \in I^X$ and $x \in X$ then*

$$A(x) = \sup_{n \in N}\{\alpha_n \cdot \chi_{A_{\alpha_n}}(x)\} \tag{2.15}$$

where $\{\alpha_n : n \in \mathbb{N}\}$ is dense in I.

We remind the reader of the standard definitions of addition and scalar multiplication of crisp sets. $A + B$ denotes the collection $\{a + b : a \in A, b \in B\}$ and for a scalar α, we denote by $\alpha \cdot A$ the collection $\{\alpha \cdot a : a \in A\}$. There are natural extensions of these concepts in the fuzzy setting and it should be noted that the following two definitions are consequences of the well known concept of a fuzzy image.

Definition 2.3 *Let $A, B \in I^X$ we define*

$$(A + B)(x) = \sup_{a+b=x} \{A(a) \wedge B(b)\} \tag{2.16}$$

for all $x \in X$.

Definition 2.4 *Let $A \in I^X, t \in K$ and $x \in X$. Then we define $t \odot A(x) = A(\frac{x}{t})$ for $t \neq 0$ and*

$$t \odot A(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ \sup A & \text{if } x = 0 \end{cases} \tag{2.17}$$

if $t = 0$.

The $\odot$ is simply used to distinguish between this type of scalar multiplication and the $\cdot$ representing ordinary pointwise multiplication of functions as used in Theorem 2.2.

Due to the lattice structure of I, we have that pointwise ordering is a natural partial ordering that can be defined on I^X. That is $\leq$ given by

$$A \leq B \Leftrightarrow A(x) \leq B(x) \tag{2.18}$$

for all $x \in X$ and $A, B \in I^X$. This leads to the following straightforward relationships between the concepts of α-levels and this ordering respectively. The proofs follow trivially.

Lemma 2.5 *If $A, B \in I^X$ then for all $\alpha \in [0, 1]$ the following statements hold*

1.

$$[A \vee B]_\alpha = A_\alpha \cup B_\alpha, \tag{2.19}$$

2.

$$[A \wedge B]_\alpha = A_\alpha \cap B_\alpha, \tag{2.20}$$

3.

$$A \leq B \Longleftrightarrow A_\alpha \subseteq B_\alpha. \tag{2.21}$$

We have the following immediate consequence of the previous lemma.

Lemma 2.6 *Let A and B be fuzzy sets. Then*

$$A = B \Longleftrightarrow A_\alpha = B_\alpha \tag{2.22}$$

for all $\alpha \in [0, 1]$.

Definition 2.7 *A fuzzy set A on X is convex if $A(kx+(1-k)y) \geq A(x) \wedge A(y)$* whenever *$x, y \in X$ and $0 \leq k \leq 1$.*

The following proposition is a useful characterization of fuzzy convexity which justifies the definition and relates it to the corresponding crisp definition.

Proposition 2.8 *Let A be a fuzzy set in X then the following three assertions are equivalent:*

1. *A is convex.*
2. *$\forall k \in [0, 1], kA + (1 - k)A \leq A$.*
3. *$\forall \alpha \in I, A_\alpha$ is convex.*

We will later require the next theorem which illustrates the relationship between the scalar multiplication and addition of fuzzy sets and the corresponding concepts in the crisp setting.

Lastly we discuss the natural concept of fuzzy points since this will be necessary for the next section. We use $\mathcal{P}(X)$ to denote the *power set* of X - that is, $\mathcal{P}(X) = \{A : A \subseteq X\}$.

If $A \in \mathcal{P}(X)$ and $\alpha \in I$ we define

$$\alpha \cdot \chi_{\{x\}} = \begin{cases} \alpha & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{2.23}$$

$$\alpha \cdot \chi_{\{x\}} = \begin{cases} \alpha & \text{on } x \\ 0 & \text{elsewhere} \end{cases} \tag{2.24}$$

So

We call $\alpha \cdot \chi_{\{x\}}$ a *fuzzy point* with *support* at x and *value* α. We will denote the set of fuzzy points in I^X by $\hat{X}$.

3 Fuzzy set-valued random variables

Our next focus is on the generalization of set-valued random variables to the fuzzy setting. The following overview of fuzzy set-valued random variables is taken mainly from [5] and [9]. These concepts are also discussed in [6, 7] amongst other works.

Definition 3.1 *Let* $\mathrm{F(X)}$ *denote the collection of fuzzy sets such that*

(a) A is uppersemicontinuous,

(b) $\{x \in X : A(x) = 1\} \neq \emptyset$,

(c) $\text{supp}A \in K(X)$.

ISBN: 978-989-95079-6-8

We define $F_k(X)$ to be the collection of fuzzy sets in $F(X)$ that are fuzzy convex. We denote the collection of measurable functions $f : \Omega \to \hat{X}$ by $M[\Omega, \hat{X}]$, that is, $f \in M[\Omega, \hat{X}]$ if and only if for each $\alpha \in (0, 1]$, $f_\alpha \in M[\Omega, X]$ with $f_\alpha(\omega) = [f(\omega)]_\alpha$ for all $\omega \in \Omega$. Let p be chosen such that $1 \leq p \leq \infty$ then $L^p[\Omega, \hat{X}]$ will denote the p-integrable functions from Ω into $\hat{X}$. A fuzzy set-valued random variable (f.r.v.) or a fuzzy random set is a function $F : \Omega \to \mathrm{F_k(X)}$ such that $F_\alpha(\omega) = \{x \in X : F(\omega)(x) \geq \alpha\}$ is a set-valued random variable for all $\alpha \in (0, 1]$. A fuzzy set-valued random variable F is called measurable (integrable) if for each $\alpha \in (0, 1]$, F_α is measurable (integrable). Let $\mathcal{M}[\Omega, F_k(X)]$ be the collection of all measurable fuzzy random variables and $\mathcal{L}^p[\Omega, \mathrm{F_k(X)}] = \mathcal{L}^p[\Omega, \Sigma, \mu, \mathrm{F_k(X)}]$ denote the set of all p-integrable fuzzy random variables for $1 \leq p \leq \infty$. Clearly, a fuzzy set-valued random variable can be considered to be a generalized set-valued random variable. F_1, F_2 fuzzy set-valued random variables then $(F_1 + F_2)(\omega) = F_1(\omega) + F_2(\omega)$ for all $\omega \in \Omega$ (the normal addition of fuzzy sets defined earlier). Similarly for a fuzzy set-valued random variable F and measurable real valued function ζ on Ω, $(\zeta F)(\omega) = \zeta(\omega)F(\omega)$ for all $\omega \in \Omega$. We will also generalize Theorem 2.8 to the fuzzy setting. There are several ways to generalize d_H to the fuzzy setting but due to certain desirable properties we will restrict our attention to d_∞ defined by

$$d_\infty(F_1, F_2) = d_H(\mathrm{supp}F_1, \mathrm{supp}F_2). \tag{3.25}$$

It follows trivially from our earlier discussion of the Hausdorff metric that $(F_k(X), d_\infty)$ is a metric space. Under the topology induced by d_∞ we have that $F_n \longrightarrow F$ if and only if $d_\infty(F_n(\omega), F(\omega)) \longrightarrow 0$ as $n \longrightarrow \infty$ for all $\omega \in \Omega$ a.e.

Based on the notation used in [1] we now generalize the definition of S_F^p to the fuzzy setting.

Definition 3.2 *A measurable function $f \in M[\Omega, \hat{X}]$ such that $f(\omega) \in F(\omega)$ for all $\omega \in \Omega$ a.e. is called a measurable selection of F. Let $F \in \mathcal{M}[\Omega, F_k(X)]$ and let $1 \leq p \leq \infty$ then $S_F^p = \{f \in L^p[\Omega, \hat{X}] : f(\omega) \in F(\omega)$ a.e. $\}$.*

We now state the second of our main results. This is a new generalization of Lemma 1.1 - notice that the collection of fuzzy points $\hat{X}$ replaces the set X in this version of the theorem. This theorem opens the way for similar results to those obtained in [1].

Theorem 3.3 *If $F \in \mathcal{M}[\Omega, F_k(X)]$ then there exists a countable collection $\{f^i\}_{i \in N} \subseteq S_F^p$, such that $F(\omega) = \mathrm{cl}_{i \in N}\{f^i(\omega)\}$ for all $\omega \in \Omega$.*

Proof: Let $F \in \mathcal{M}[\Omega, F_k(X)]$ and let $\omega \in \Omega$. Then for $n \in \mathbb{N}$ we have $F_{\alpha_n} \in \mathcal{M}[\Omega, K_k(X)]$ with $\{\alpha_n\}_{n \in N}$ a countable set dense in I as in Theorem 2.2. Therefore by Lemma 1.1, for each $n \in \mathbb{N}$, $\exists \{f^i_{\alpha_n}\}_{i \in N}$ such that

$$\mathrm{cl}_{i \in N}\{f^i_{\alpha_n}(\omega)\}(x) = F_{\alpha_n}(\omega).$$

For each $i \in \mathbb{N}$ we now define

$$f^i(\omega) = \bigvee_{n \in N} \alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}.$$

Then $\{f^i\}_{i \in N}$ is the required collection. Let $x \in X$ and now notice that for each $i \in \mathbb{N}$, $f^i : \Omega \to \hat{X}$. We now have that $\mathrm{cl}_{i \in N}\{[f_i(\omega)]\}(x) = \mathrm{cl}_{i \in N}\{[f_i(\omega)](x)\} = \mathrm{cl}_{i \in N}\{\bigvee_{n \in N} \alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}(x)\} = \lim_{j \longrightarrow \infty} b_j$ where for each $j \in \mathbb{N}$, $b_j \in \left\{ \bigvee_{n \in N} \alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}(x) : i \in \mathbb{N} \right\}$. Now for each $j \in \mathbb{N}$, $b_j = \lim_{k \longrightarrow \infty} b_j^k$ with $b_j^k \in \{\alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}(x) : i, n \in \mathbb{N}\} \subseteq I$. Thus

$$\mathrm{cl}_{i \in N}\{f_i(\omega)\}(x) = \lim_{j \longrightarrow \infty} \lim_{k \longrightarrow \infty} b_j^k, \tag{3.26}$$

for $\{b_j^k\}_{j,k \in N} \subseteq \{\alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}(x) : i, n \in \mathbb{N}\} = \left\{[f^i_{\alpha_n}(\omega)](x) : i, n \in \mathbb{N}\right\}$ and b_j^k is nondecreasing with respect to k.

We also have by Theorem 2.2 that $[F(\omega)](x) = \left(\bigvee_{n \in N} \alpha_n \cdot \chi_{F_{\alpha_n}(\omega)}\right)(x)$ $= \left(\bigvee_{n \in N} \alpha_n \cdot \chi_{\mathrm{cl}_{\{f^i_{\alpha_n}(\omega)\}}}\right)(x) = \left(\bigvee_{n \in N} \alpha_n \cdot \chi_{\mathrm{cl}_{\{f^i_{\alpha_n}(\omega)\}}}(x)\right) \in I$.

By definition $\left(\bigvee_{n \in N} \alpha_n \cdot \chi_{\mathrm{cl}_{\{f^i_{\alpha_n}(\omega)\}}}(x)\right) = \lim_{k \longrightarrow \infty} b^k$ where $(b^k)_{k \in N} \subseteq \{[f^i_{\alpha_n}(\omega)](x) : n \in \mathbb{N}\} \subseteq I$ and b^k is nondecreasing with respect to k. Also for each $k \in \mathbb{N}$ we have that $b^k = \lim_{j \longrightarrow \infty} b_j^k$ with $b_j^k \in \left\{[f^i_{\alpha_n}(\omega)](x) : i, n \in \mathbb{N}\right\}$ for each $j, k \in \mathbb{N}$. Therefore we have

$$[F(\omega)](x) = \lim_{k \longrightarrow \infty} \{ \lim_{j \longrightarrow \infty} b_j^k\} \tag{3.27}$$

This concludes the proof since it is well known from real analysis that

$$\lim_{k \longrightarrow \infty} \{ \lim_{j \longrightarrow \infty} b_j^k\} = \lim_{j \longrightarrow \infty} \{ \lim_{k \longrightarrow \infty} b_j^k\}$$

for all $b_j^k \in I$ with $j, k \in \mathbb{N}$.

References

[1] F. Hiai, H Umegaki. Integrals, Conditional Expectations, and Martingales of Multivalued Functions, *J. Multivariate Anal 7*, no. 1 (1977), 149–182.

[2] A. K. Katsaras & D. B. Liu. Fuzzy Vector Spaces and Fuzzy Topological Vector Spaces. J.Math.Anal.Appl. 58 (1977), 135-146.

[3] A. K. Katsaras. Fuzzy Topological Vector Spaces I. Fuzzy Sets and Systems 6 (1981) 85-95.

[4] A. K. Katsaras. Fuzzy Topological Vector Spaces II. Fuzzy Sets and Systems 12 (1984) 143-154.

[5] S. Li, Y. Ogura, V. Kreinovich. *Limit theorems and applications of set-valued and fuzzy set-valued random variables*, Kluwer Academic Press, Dordrecht-Boston-London, 2002.

ISBN: 978-989-95079-6-8

[6] S. Li, Y. Ogura. Convergence of set-valued and fuzzy-valued martingales. Fuzzy Sets and Systems 101 (1999) 453-461.

[7] S. Li, Y. Ogura. A Convergence theorem of fuzzy-valued martingales in the extended Hausdorff metric H_∞. Fuzzy Sets and Systems 135 (2003) 391-399.

[8] A. L. Pinchuck. Extension Theorems on L-Topological Spaces and L-Fuzzy Vector Spaces. Masters thesis, Rhodes University (2001).

[9] M. L. Puri, D. A. Ralescu. Convergence Theorem for Fuzzy Martingales. *Journal of Mathematical Analysis and Applications.* **160** (1991), 107-122.

ISBN: 978-989-95079-6-8

Production and Transportation Planning – A fuzzy Approach for Minimizing the Total Cost

Heinrich J. Rommelfanger

Goethe University Frankfurt am Main
Faculty of Economics and Business Administration
Rommelfanger@wiwi.uni-frankfurt.de

Abstract— *In this paper, we deal with the production and transportation planning of a household appliances manufacturer that has production facilities and central stores for resellers in several sites in Europe. Each store can receive products from all production plants and it is not necessary that all products are produced in all production units. The transport between any two bases is done by trucks. For simplicity we assume, that each truck has the same capacity of M EURO-pallets, and for each product the unit is EURO-pallet. The target of this paper is to determine a production and transportation plan that minimize the total sum of the production cost and the transportation cost. For working in a realistic environment we assume that the production capacities in the different plants and the demand in the sales bases are not known exactly but the management can describe the data in form of fuzzy numbers. By using an inter-active algorithm for solving the fuzzy linear programming system we achieve a stable production and a satisfactory supply of the products. Moreover, we demonstrate that this integer programming problem can adequately be solved without using computation-intensive integer programming algorithms. Additionally, in the course of the inter-active solution process the production bottlenecks get clearly visible. A numerical example illustrates the efficiency of the proposed procedure.*

Keywords — *Fuzzy LP-systems, integer programming, FULPAL, Production and transportation planning, inter-active process*.

1 Introduction

Due to globalisation and the involved international expansion of companies numerous firms in Europe produce their products in several European and Overseas countries. On the one hand they take advantage of the different production cost and of higher lot sizes, on the other hand they endeavour to minimize the total cost of manufacturing and transportation.

In literature it is usually proposed to determine a minimal cost production plan by means of an integer programming system. But this procedure is very computationally intensive. Moreover, it is neglected that real production capacities in the different plants and the demand for the products are not known exactly. In this paper we look on the more realistic case that the management can describe the production capacities in the different plants and the demand in the sales bases in form of fuzzy numbers.

The paper is organized as follows: In Section 2 the problem is formulated in detail. In Section 3 we demonstrate that the integer programming problem can adequately be solved by using an inter-active algorithm for solving the fuzzy linear programming system. A numerical example in Section 4 illustrates the efficiency of the proposed procedure. Finally, conclusions and possible extensions are presented in Section 5.

2 Problem formulation

We deal with the production and transportation planning of a household appliances manufacturer HAM that has production facilities and central stores for resellers in several sites in Europe. Each store can receive products from all production plants and it is not necessary that all products are produced in all production units.

- The number of production and sales bases of HAM is N. Obviously, it is not necessary to differ between production plants and sales stores. A base without any demand is a pure production location and a base where nothing is produced is a pure sales shop.
- The number of products of HAM is K.
- The transport between any two bases is done by trucks. For simplicity we assume, that each truck has the same capacity of M EURO-pallets independent of the sort of products.
- For each product the unit is EURO-pallet.

In the production and transportation planning model, we use the following notations:

x_{ki} output of the product k at the base i,

y_{kij} total number of units of the product k transported from the base i to the base j,

w_{ij} number of the trucks from base i to base j,

$\tilde{D}_{ki}$ demand for the product k at the base i,

$\tilde{P}_{ki}$ production capacity for the product k at the base i,

C_{ki} production cost of one unit of product k at the base i,

T_{ij} cost for transporting a truck from the base i to the base j, where $T_{ii} = 0$.

$k \in \{1,\ldots,K\}$, $i, j \in \{1,\ldots,N\}$

The items $x_{ki}, y_{kij}, w_{ij}, \tilde{D}_{ki}, \tilde{P}_{ki}$ are referred to the same time period, e. g. one week or one month.

Now, a production and transportation plan that minimize the total sum of the production cost and the transportation cost can be determined by solving the following integer programming problem:

ISBN: 978-989-95079-6-8

$$z(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{N}\sum_{k=1}^{K} C_{ki}x_{ki} + \sum_{j=1}^{N}\sum_{i=1}^{N} T_{ji}w_{ji} \;\rightarrow\; \text{Min}$$

subject to (1)

$$x_{ki} \mathrel{\tilde{\leq}} \tilde{P}_{ki}, \quad k=1,...,K;\ i=1,...,N$$

$$s_{ik} = x_{ki} - \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kji} \mathrel{\tilde{\geq}} \tilde{D}_{ki}, \quad k=1,...,K;\ i=1,...,N$$

$$\sum_{k=1}^{K} y_{kij} \leq M\, w_{ij}, \quad i,j=1,...,m$$

$$x_{ij}, y_{kij}, w_{ij} \in \mathbf{N}\cup\{0\}, \quad k=1...,K,\ i,j=1,...,N$$

$$\mathbf{x} = (x_{11},\ldots,x_{K1},x_{12},\ldots,x_{K2},\ldots,x_{1N},\ldots,x_{KN})$$

$$\mathbf{w} = (w_{11},\ldots,w_{N1},w_{12},\ldots,w_{N2},\ldots,w_{1N},\ldots,w_{NN})$$

The first constraint expresses that the output of the product k at the base i is smaller than or equal to its production capacity $\tilde{P}_{ki}$; the second constraint means that the supply s_{ki} of the product k at the base i is larger than or equal to the its demand $\tilde{D}_{ki}$; the third constraint indicates that the total amount of products transported from the base i to the base j is smaller than or equal to the transportation capacity of w_{ij} trucks; the fourth constraint means that the output of the product k at the base i, the number of products transported from the base i to the base j and the number of trucks from the base i to base j are nonnegative integer.

Concerning the imprecise right-hand sides $\tilde{P}_{ki}$ and $\tilde{D}_{ki}$ we assume that management is able to describe the production capacities in the different plants and the demand in the sales bases in form of fuzzy numbers $\tilde{P}_{ki} = (p_{ki}, 0, \pi_{ki})$ and $\tilde{D}_{ki} = (d_{ki}, \delta_{ki}, 0)$. Here, p_{ki} and $d_{ki} - \delta_{ki}$ d_{ki} are the production capacities and the demands respectively that are expected in any case. Furthermore the management estimate the maximal production capacities and the maximal demands as $p_{ki} + \pi_{ki}$ and d_{ki} respectively.

3 Solution process

As shown in Rommelfanger [3], fuzzy integer programming LP-problem can be effectively solved by interactive algorithms for solving fuzzy LP-systems. The present system (1) is a relative simple model with one crisp objective function and soft constraints. Therefore we can use a special form of the algorithm FULPAL (**FU**zzy **L**inear **P**rogramming Based on **A**spiration **L**evels) for a stepwise calculation of an efficient compromise solution of the system (1).

At first, we have to calculate the smallest and the highest total cost by solving the two crisp LP-systems:

$$\underline{z} = \text{Min}\,(\sum_{i=1}^{N}\sum_{k=1}^{K} C_{ki}x_{ki} + \sum_{i=1}^{N}\sum_{j=1}^{N} T_{ji}w_{ji})$$

subject to (2)

$$x_{ki} \leq p_{ki} + \pi_{ki}, \quad k=1,...,K;\ i=1,...,N$$

$$s_{ki} = x_{ki} - \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kji} \geq d_{ki} - \delta_{ki}, \quad \begin{matrix} k=1,...,K \\ i=1,...,N \end{matrix}$$

$$\sum_{k=1}^{K} y_{kij} - M\, w_{ij} \leq 0, \quad i,j=1,...,m$$

$$x_{ij}, y_{kij}, w_{ij} \geq 0, \quad k=1...,K,\ i,j=1,...,N$$

and

$$\overline{z} = \text{Min}\,(\sum_{i=1}^{N}\sum_{k=1}^{K} C_{ki}x_{ki} + \sum_{i=1}^{N}\sum_{j=1}^{N} T_{ji}w_{ji})$$

subject to (3)

$$x_{ki} \leq p_{ki}, \quad k=1,...,K;\ i=1,...,N$$

(In case of $\sum_{i=1}^{N} p_{ki} < \sum_{i=1}^{N} d_{ki}$ the capacities p_{ki} must be increased up to the existence of an feasible solution of (2), starting with the plants that have the highest production costs; $k=1,\ldots,K$.)

$$s_{ki} = x_{ki} - \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kji} \geq d_{ki}, \quad k=1,...,K;\ i=1,...,N$$

$$\sum_{k=1}^{K} y_{kij} - M\, w_{ij} \leq 0, \quad i,j=1,...,N$$

$$x_{ij}, y_{kij}, w_{ij} \geq 0, \quad k=1...,K,\ i,j=1,...,N$$

In accordance with FULPAL, the objective function and the soft constraints of the system (1) are transformed in utility functions, where $z^A[r], p_{ki}^A[r], d_{ki}^A[r]$ with $r=1$ are the crisp aspiration levels that are specified by the management for the time being.

In the course of the inter-active solution process the management can change the aspiration levels step by step. For getting an effective comparability of the utilities, the same utility (membership degree) λ_A is assigned to all crisp aspiration levels. The goal of the management is to get a solution that satisfies all aspiration levels.

$$\mu_z(\mathbf{x},\mathbf{w}) = \quad (4)$$

$$\begin{cases} 1 & \text{if} \quad z(\mathbf{x},\mathbf{w}) < \underline{z} \\ 1 - \dfrac{z(\mathbf{x},\mathbf{w}) - \underline{z}}{z^A[r] - \underline{z}} \cdot (1-\lambda_A) & \text{if} \quad \underline{z} \leq z(\mathbf{x},\mathbf{w}) < z^A[r] \\ \lambda_A + \dfrac{z(\mathbf{x},\mathbf{w}) - z^A[r]}{\overline{z} - z^A[r]} \cdot (1-\lambda_A) & \text{if} \quad z^A[r] \leq z(\mathbf{x},\mathbf{w}) \leq \overline{z} \\ 0 & \text{if} \quad \overline{z} < z(\mathbf{x},\mathbf{w}) \end{cases}$$

$$\mu_{xki}(x_{ki}) = \quad (5)$$

$$\begin{cases} 1 & \text{if} \quad x_{ki} < p_{ki} \\ 1 - \dfrac{x_{ki} - p_{ki}}{p_{ki}^A[r] - p_{ki}} \cdot (1-\lambda_A) & \text{if} \quad p_{ki} \leq x_{ki} \\ \lambda_A + \dfrac{x_{ki} - p_{ki}^A[r]}{p_{ki} + \pi_{ki} - p_{ki}^A[r]} \cdot (1-\lambda_A) & \text{if} \quad p_{ki}^A[r] < x_{ki} \\ 0 & \text{if} \quad p_{ki} + \pi_{ki} < x_{ki} \end{cases}$$

ISBN: 978-989-95079-6-8

$$\mu_{ski}(s_{ki}) = \qquad (6)$$

$$\begin{cases} 0 & \text{if} \quad s_{ki} < d_{ki} - \delta_{ki} \\ \dfrac{s_{ki} - (d_{ki} - \delta_{ki})}{d^A_{ki}[r] - (d_{ki} - \delta_{ki})} \cdot \lambda_A & \text{if} \quad d_{ki} - \delta_{ki} \le s_{ki} \le d^A_{ki}[r] \\ \lambda_A + \dfrac{s_{ki} - d^A_{ki}[r]}{d_{ki} - d^A_{ki}[r]} \cdot (1 - \lambda_A) & \text{if} \quad d^A_{ki}[r] < s_{ki} \le d_{ki} \\ 1 & \text{if} \quad d_{ki} < s_{ki} \end{cases}$$

$$k = 1,...,K;\ i = 1,...,N$$

For calculating a compromise solution of the multi-objective system

$$(\mu_Z(\mathbf{x},\mathbf{w}), \mu_{x11}(x_{11}),\ldots,\mu_{xKN}(x_{KN}), \mu_{s11}(s_{11}),\ldots,\mu_{sKN}(s_{KN})) \rightarrow \text{Max}$$

subject to (7)

$$x_{ki} \le p_{ki} + \pi_{ki}, \quad k = 1,...,K;\ i = 1,...,N$$

$$s_{ki} = x_{ki} - \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kji} \ge d_{ki} - \delta_{ki},$$

$$k = 1,...,K;\ i = 1,...,N$$

$$\sum_{k=1}^{K} y_{kij} - M w_{ij} \le 0, \ i, j = 1,...,m$$

$$x_{ij}, y_{kij}, w_{ij} \ge 0, \ k = 1...,K, \ i, j = 1,...,N$$

we use the compromise objective function

$$\mu = \text{Min}\ (\mu_Z(\mathbf{x},\mathbf{w}), \mu_{x11}(x_{11}),\ldots,\mu_{xKN}(x_{KN}), \mu_{s11}(s_{11}),\ldots,\mu_{sKN}(s_{KN})) \qquad (8)$$

Then, we get a pareto-optimal compromise solution of (7) by solving the crisp mathematical programming system (9).

$$\lambda \rightarrow \text{Max}$$

subject to (9)

$$\lambda \le \mu_Z(\mathbf{x},\mathbf{w})$$

$$\lambda \le \mu_{xki}(x_{ki}), \quad k = 1,...,K;\ i = 1,...,N$$

$$\lambda \le \mu_{ski}(s_{ki}), \quad k = 1,...,K;\ i = 1,...,N$$

$$x_{ki} \le p_{ki} + \pi_{ki}, \quad k = 1,...,K;\ i = 1,...,N$$

$$s_{ki} = x_{ki} - \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kji} \ge d_{ki} + \delta_{ki},$$

$$k = 1,...,K;\ i = 1,...,N$$

$$\sum_{k=1}^{K} y_{kij} - M w_{ij} \le 0, \ i, j = 1,...,m$$

$$x_{ij}, y_{kij}, w_{ij} \ge 0, \ k = 1...,K, \ i, j = 1,...,N$$

As the management is only interested in a solution that satisfies all aspiration levels, it is sufficient to solve the following crisp linear LP-system, where $r = 1$; for details see [1; 2].

$$\lambda \rightarrow \text{Max}$$

subject to (10)

$$(z^A[r] - \underline{z})\lambda + (1 - \lambda_A)\left(\sum_{i=1}^{N}\sum_{k=1}^{K} C_{ki}x_{ki} + \sum_{i=1}^{N}\sum_{j=1}^{K} T_{ji}w_{ji}\right) \le z^A[r] - \lambda_A \underline{z}$$

$$(p^A_{ki}[r] - p_{ki})\lambda + (1 - \lambda_A)x_{ki} \le p^A_{ki}[r] - \lambda_A p_{ki}, \quad k = 1,...,K;\ i = 1,...,N$$

$$(d_{ki} - d^A_{ki}[r])\lambda - (1 - \lambda_A)\left(x_{ki} - \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kji}\right) \le \lambda_A d_{ki} - d^A_{ki}[r]$$

$$k = 1,...,K;\ i = 1,...,N$$

$$\sum_{k=1}^{K} y_{kij} - M w_{ij} \le 0, \ i, j = 1,...,m$$

$$\lambda, x_{ij}, y_{kij}, w_{ij} \ge 0, \ k = 1...,K, \ i, j = 1,...,N$$

Obviously, a solution of the system (10) with $\lambda \ge \lambda_A$ is a pareto-optimal solution of the systems (7) and (1) and fulfils all aspiration levels $z^A[1]$, $p^A_{ki}[1]$, $d^A_{ki}[1]$.

Moreover, if λ is greater than λ_A, the management can increase some of the aspiration levels to $z^A[2]$, $p^A_{ki}[2]$, $d^A_{ki}[2]$ and can calculate a new pareto-optimal solution by means of the revised system (10), and so on. When the management is satisfied with the non-integer solution, it is time for looking for an integer solution. Due to the soft constraints several integer solution will exist in the neighborhood of the non-integer solution.

4 Numerical example

The manufacturer HAM produces four products P1, P2, P3 and P4 in three product plants in **E**ssen, **K**rakow and **L**yon.

Table 1: Production costs in € per unit

	P1	**P2**	**P3**	**P4**
E	200	400	330	500
K	180	300	280	
L	220		300	570

Table 2: Transportation costs in € per truck

	E	**K**	**L**
E		1000	600
K	1000		1200
L	600	1200	

Table 3: Secure production capacities p_{ki}

	p_{1i}	p_{2i}	p_{3i}	p_{4i}
E	600	450	250	250
K	350	200	250	
L	250		400	150

ISBN: 978-989-95079-6-8

Table 4: Maximal production capacities $p_{ki} + \pi_{ki}$

	$p_{1i} + \pi_{1i}$	$p_{2i} + \pi_{2i}$	$p_{3i} + \pi_{3i}$	$p_{4i} + \pi_{4i}$
E	750	500	300	300
K	500	300	300	
L	300		500	200

Table 5: Minimal demands $d_{ki} - \delta_{ki}$

	$d_{1i} - \delta_{1i}$	$d_{2i} - \delta_{2i}$	$d_{3i} - \delta_{3i}$	$d_{4i} - \delta_{4i}$
E	650	300	300	150
K	450	100	200	80
L	150	150	250	100

Table 6: Maximal demands d_{ki}

	d_{1i}	d_{2i}	d_{3i}	d_{4i}
E	750	400	350	200
K	350	150	200	120
L	200	200	400	130

Using the LP-systems (2) and (3), we get the minimal total cost $\underline{z}$ = 864,100 € and the maximal cost $\overline{z}$ = 1,091,700 €. With all these information the management of HAM specifies for the total cost the aspiration level z[1] = 950,000 € and for the outputs and the demands the following aspiration levels.

Table 7: Aspiration levels p_{ki} [1]

output	p_{1i} [1]	p_{2i} [1]	p_{3i} [1]	p_{4i} [1]
E	720	480	280	280
K	450	270	280	
L	280		470	180

Table 8: Aspiration levels d_{ki} [1]

demand	d_{1i} [1]	d_{2i} [1]	d_{3i} [1]	d_{4i} [1]
E	670	350	330	180
K	280	120	200	100
L	180	170	300	110

As $\lambda = 0.53 > \lambda_A = 0.5$, the solution of the corresponding LP-system (10) fulfills all aspiration levels. E. g. the total cost of this plan is 945,082.50 €.
Moreover this result indicates that it is possible to improve the aspiration levels. For simplification we assume that the management is only interested in lower cost and decides to reduce the total cost to z[2] = 940,000 €.

Even the solution of the revised system (10) fulfils all aspiration levels. We have $\lambda = 0.5077$, z = 938,712.20 € and the following production and transportation plan.

Table 9: Production and transportation plan

x1e	x1k	x1l	x2e	x2k	x3e
671,05	280,92	180,26	372,24	269,08	250,79
x3k	x3l	x4e	x4l	wek	wel
279,61	301,32	279,61	111,18	4,97	1,08
wke	wkl	wle	wlk	y1ek	y1el
3,98	7,43	0,00	0,05	0,00	0,00
y1ke	y1kl	y1le	y1lk	y2ek	y2el
0,00	0,00	0,00	0,00	0,00	21,58
y2ke	y2kl	y3ek	y3el	y3ke	y3kl
0,00	148,68	0,00	0,00	79,61	0,00
y3le	y3lk	y4ek	y4el	y4le	y4lk
0,00	0,00	99,34	0,00	0,00	0,92

As we have soft constraints it is very simple to derive an integer plan that fulfils all aspiration levels and actually leads to lower total cost z= 937,100 €.

Table 10: Integer production and transportation plan

x1e	x1k	x1l	x2e	x2k	x3e
670	280	180	370	270	250
x3k	x3l	x4e	x4l	wek	wel
280	300	280	110	5	1
wke	wkl	wle	wlk	y1ek	y1el
4	8	0	0	0	0
y1ke	y1kl	y1le	y1lk	y2ek	y2el
0	0	0	0	0	20
y2ke	y2kl	y3ek	y3el	y3ke	y3kl
0	150	0	0	80	0
y3le	y3lk	y4ek	y4el	y4le	y4lk
0	0	100	0	0	0

5 Conclusions

Fuzzy mathematical programming systems offer the possibility to model real problems as precisely as a decision maker is able to describe them. In doing so non-satisfying modeling can be avoided. Another advantage of fuzzy models is the fact that (mixed) integer programming problems can be solved very easily because the boundaries are not crisp but fuzzy and points with integer variables in the neighborhood of an optimal solution are feasible in general.

In this paper we have assumed that only the right-hand sides of some constraints are not known exactly. The model can be extended to the case that additionally coefficients of the objective function or of the constraints are described in form of fuzzy intervals. Moreover, it is possible to allow several objective functions. All these systems can adequately be solved by means of the inter-active algorithm FULPAL, see [1; 2; 3].

References

[1] H. Rommelfanger. FULPAL 2.0 - An Interactive Algorithm for Solving Multicriteria Fuzzy Linear Programs Controlled by Aspiration Levels. In: Scheigert D. (Ed.) Methods of Multicriteria Decision Theory. Pfalzakademie Lamprecht, 21-34, 1995.

[2] H. Rommelfanger and R. Slowinski. Fuzzy Linear Programming with single or multiple Objective Functions. In: Slowinski R. (Ed.): Fuzzy Sets in Decision Analysis, Operations Research and Statistics, Kluwer Academic Publishers Norwell, Massachusetts, USA,179-213, 1998.

[3] H. Rommelfanger. The Advantages of Fuzzy Optimization Models in Practical Use. *Fuzzy Optimization and Decision Making 3*, 295-309, 2004.

ISBN: 978-989-95079-6-8

The Use of Interval-Valued Probability Measures in Fuzzy Linear Programming: A Constraint Set Approach

Phantipa Thipwiwatpotjana
Department of Mathematical sciences
University of Colorado at Denver
Phantipa.Thip@ucdenver.edu

Weldon A. Lodwick
Department of Mathematical sciences
University of Colorado at Denver
Weldon.Lodwick@ucdenver.edu

***Abstract*—*This paper uses a constraint set approach to linear programming problems with equality constraints whose coefficients and/or right-hand side values could be uncertain. We consider three types of uncertainty: probabilistic, fuzzy/possibilistic, and interval. The constraint set generated by the linear constraints under uncertainty is ill-defined and difficult to generate. Our approach computes an inner constraint set and an outer constraint set. Optimization is then carried out using these two sets using interval-valued probability approaches. We discuss the requisite associated semantics.*

Keywords—inverval-valued probability measure, random sets, inner constraint set, outer constraint set.

I. Introduction

In order to solve linear programming problems under uncertainty, many researchers convert the problem into a sequence of deterministic problems. The approach in this paper is to work with the constraint set of the problem.

It is difficult (exponentially hard) to find the exact shape of the feasible set of uncertain constraints, as we shall see. Therefore we create two sets (boxes); one represents the set of solutions guaranteed to have a corresponding data to satisfy the constraints while the other is the set that contains all solutions. They are called the *inner set* and the *outer set*, respectively. We use these two sets in our optimization problem.

We present how to obtain these two sets in section III after giving some useful definitions. In section IV, we create an interval-valued probability measure (IVPM) by a pair of possibility and necessity measures on the space between the inner and outer sets.

In section V, we solve an uncertain optimization problem, at each α-level of the uncertain data, on the inner box and give a guaranteed bound on the objective value. We use an IVPM to get the bound on the objective value with some γ degree of occurrence. The final section articulates some conclusions.

II. Background and Definitions

We consider linear programming (LP) problems with uncertain parameters of the form

$$\min_x \breve{c}^{\mathrm{T}} x \quad \text{s.t.} \quad \breve{A}x = \breve{b}. \tag{1}$$

$\breve{A}, \breve{b}$ and $\breve{c}$ may be (or are) uncertain. Without knowing the type of uncertainty, we use $\breve{b}_j$ and $\breve{c}_j$ as the j^{th} component of the uncertain vectors $\breve{b}$ and $\breve{c}$, respectively. Similarly, $\breve{a}_{ij}$ is the (i,j) component of the uncertain matrix $\breve{A}$.

The main types of uncertainties are probabilistic, fuzzy/possibilistic and interval. When we know the type of uncertain component $\breve{a}_{ij}$, we can specify the uncertainty by the following symbols. (The same symbols also apply for components $\breve{b}_j$ and $\breve{c}_j$.)

- $\tilde{a}_{ij} \equiv$ a fuzzy component a_{ij},
- $\hat{a}_{ij} \equiv$ a possibilistic component a_{ij},
- $\breve{a}_{ij} \equiv$ a probabilistic component a_{ij},
- $[a_{ij}] = \left[\underline{a}_{ij}, \overline{a}_{ij}\right] \equiv$ an interval component a_{ij}.

Definition 2.1: Given an LP problem (1), the constraint set $\Omega_{\exists\exists} := \left\{ x_{Ab} \mid \exists A \in \breve{A}, \exists b \in \breve{b} \ : \ Ax_{Ab} = b \right\}$ is said to be the set of all solution x's for the constraints of (1).
The set $\Omega_{\exists\exists}$ may be (usually is) a non-convex set.

Definition 2.2: A finite *random set* on S is a pair $(\mathscr{F}, m)$ where $\mathscr{F}$ is a finite family of distinct non-empty subsets of S and m is a mapping $\mathscr{F} \to [0,1]$ such that $\sum_{A \in \mathscr{F}} m(A) = 1$.

We can also define a random set on S when S is infinite by using the measure space $(S, \mathscr{S}, m)$, where $\mathscr{S}$ is a σ-algebra of S. Therefore, a random set on an infinite set S is a pair $(\mathscr{F}, m)$ where $\mathscr{F}$ is a subset of $\mathscr{S}$ such that $\sum_{A \in \mathscr{F}} m(A) = 1$.

In real applications, we might not know (with certainty) the probability measure for our problems. Lodwick and Jamison [5] use an IVPM, $i_{\breve{m}}(A) = \left[i^-_{\breve{m}}(A), i^+_{\breve{m}}(A)\right]$, to measure a partial representation for an unknown probability measure. The original paper for the idea of IVPM is Weichselberger [12]. We use the following notation and information throughout the paper unless stated otherwise:

- The arithmetic operations applied to intervals are those of interval arithmetic [7] and where tractable, constraint interval arithmetic [3], [6].
- The set of all intervals contained in $[0,1]$ is denoted as

$$\mathrm{Int}_{[0,1]} \equiv \{[a,b] \mid 0 \le a \le b \le 1\}.$$

- $\mathcal{S}$ denotes the universal set and $\mathcal{A}$ is a σ-algebra of $\mathcal{S}$.

Definition 2.3: (Weichselberger [12]) Given measurable space $(\mathcal{S}, \mathcal{A})$, an interval valued function $i_{\breve{m}} : A \subseteq \mathcal{A} \to \mathrm{Int}_{[0,1]}$ is called an *R-probability* if:

- $i_{\breve{m}}(A) = \left[i^-_{\breve{m}}(A), i^+_{\breve{m}}(A)\right] \subseteq [0,1]$,

ISBN: 978-989-95079-6-8

- $\exists$ a probability measure, Pr, on $\mathcal{A}$ such that $\forall A \in \mathcal{A}$, $\Pr(A) \in i_{\check{m}}(A)$.

Definition 2.4: A function $p : \mathcal{S} \to [0,1]$ is called a *regular possibility distribution function* if it is a possibility such that

$$\sup\{p(x) \mid x \in \mathcal{S}\} = 1. \tag{2}$$

Possibility distribution functions (see [11]) define a possibility measure, $Pos : \mathcal{S} \to [0,1]$ where

$$Pos(A) = \sup\{p(x) \mid x \in A\} \tag{3}$$

and its dual necessity measure is

$$Nec(A) = 1 - Pos(A^c), \tag{4}$$

where $\sup\{p(x) \mid x \in \emptyset\} = 0$.
In [4], it is shown that possibility distributions can be constructed which satisfy the following consistency definition.

Definition 2.5: Let $p : \mathcal{S} \to [0,1]$ be a regular possibility distribution function with associated possibility measure Pos and necessity measure Nec. Then p is said to be *consistent* with random variable X if for every measurable set A, $Nec(A) \leq \Pr(X \in A) \leq Pos(A)$. Note that this necessity measure Nec may not be the dual of Pos in this definition.

An R-probability from definition 2.3 is an *interval-valued probability measure* (IVPM) where $i_{\check{m}}^-$ and $i_{\check{m}}^+$ are constructed from a possibility (fuzzy) distribution function. To see this is an IVPM see [4].

The R-probability function $i_{\check{m}}$ in definition 2.3 is used to define IVPMs. A possibility and necessity pair, $i_{\hat{m}}(A) = [Nec(A), Pos(A)]$, constructed by definition 2.5 is able to bound an unknown probability of interest. Therefore it can be used to define an IVPM.

Definition 2.6: The Interval-Valued Probability Measure (IVPM) defined from possibility and necessity measures is $i_{\hat{m}}(A) = [Nec(A), Pos(A)]$.
The reader could find more explanations, examples and a construction of an IVPM in [5] .

III. Generating Two Sets from Equality Constraint Set

In this section, we consider only constraints of (1),

$$\breve{A}x = \breve{b}. \tag{5}$$

Although the set of all solution x's, $\Omega_{\exists\exists}$, for (5) is difficult to construct, it is relatively easy to generate two random sets for each α-level called *inner set*, B_α^I, and *outer set*, B_α^O, such that $B_\alpha^I \subseteq B_\alpha^O$. Each x in B_α^I has a particular matrix $A_x \in A_\alpha$ and vector $b_x \in b_\alpha$, such that $A_x x = b_x$. An outer set, B_α^O, represents a space where x should be. This implies that $\Omega_{\exists\exists} \subseteq B^O$.

A. Interval case

We drop the α subscript when we work with interval version of equation (5).

A system

$$[A]x = [b] \tag{6}$$

is called *solvable* if $Ax = b$ has a solution for each $A \in [A]$ and for each $b \in [b]$, where $[A] = \left[\underline{A}, \overline{A}\right]$ and $[b] = \left[\underline{b}, \overline{b}\right]$.

Gay [2] gives a method for solving interval linear equations when $[A]$ is a regular square interval uncertain matrix, i.e., A is invertible, $\forall A \in [A]$. Rohn [8] provides a powerful theorem for any $m \times n$ interval matrix $[A]$ that guarantees solvability by testing nonnegative solutions of a finite number of linear systems.

The following notations are useful for theorem 3.1 below.

- $A_c = \frac{(\underline{A}+\overline{A})}{2}$, the center matrix.
- $\Delta = \frac{(\overline{A}-\underline{A})}{2}$, the radius matrix.
- $b_c = \frac{(\underline{b}+\overline{b})}{2}$ and $\delta = \frac{(\overline{b}-\underline{b})}{2}$.
- $Y_m = \{y \in \mathrm{R}^m \mid y_j \in \{-1, 1\}, \forall j = 1, 2, \ldots m\}$.
- $T_y = \mathrm{diag}(y_1, y_2, \ldots, y_m)$, for each $y \in Y_m$.
- $\mathrm{Conv}(X) = \{\gamma p_1 + (1-\gamma)p_2 \mid p_1, p_2 \in X,\ \gamma \geq 0\}$.

Theorem 3.1 (see Rohn [8]): A system of linear equations $[A]x = [b]$ is solvable if and only if for each $y \in Y_m$ the system

$$\left.\begin{array}{c}(A_c - T_y\Delta)\,x^1 - (A_c + T_y\Delta)\,x^2 = b_c + T_y\delta, \\ x^1 \geq 0,\ \ x^2 \geq 0\end{array}\right\} \tag{7}$$

has a solution x_y^1, x_y^2. Moreover, if this is the case, then for each $A \in [A], b \in [b]$ the system $Ax = b$ has a solution in the set $\mathbf{C} := \mathrm{Conv}\left\{x_y^1 - x_y^2 \mid y \in Y_m\right\}$.

Theorem 3.1 tells us that there is a solution for the system $Ax = b$ in the set $\mathbf{C}$. However, it does not mean that all the solutions of the system $Ax = b$ are in $\mathbf{C}$, especially when A is not a square matrix.

Corollary 3.1 (see [1]): A system of linear equations $[A]x = [b],\ x \geq 0$ is solvable if and only if for each $y \in Y_m$ the system

$$(A_c - T_y\Delta)\,x = b_c + T_y\delta, \tag{8}$$

has a nonnegative solution x_y. Moreover, if this is the case, then for each $A \in [A], b \in [b]$ the system $Ax = b,\ x \geq 0$ has a solution in the set $\mathrm{Conv}\{x_y \mid y \in Y_m\}$.

Now we are going to generate outer and inner sets given the assumption that our constraint set is solvable.

1) An outer set, B^O: We will consider two cases when $m = n$ and when $m < n$.

- When $m = n$. Suppose that a system of linear equation (6) is solvable by theorem 3.1 and $[A]$ is a regular matrix, we can create a rectangular box that contains all the solution points $x_y^1 - x_y^2$ of the system (7). To see this simply take $\max / \min$ of components. We prefer to use the rectangular box as an outer set over the set $\mathrm{Conv}\left\{x_y^1 - x_y^2;\ y \in Y_m\right\}$ because it is easier to construct. We can use $2^{m^2}2^m$ endpoint equations to generate $\mathbf{C}$, hence creating a rectangular box, B^O.
  We will not consider the case when $[A]$ is not a regular matrix. If there exists a singular matrix $A \in [A]$, then B^O is an unbounded set.
- When $m < n$. Each system $Ax = b$ in the constraint set (6) provides infinitely many solutions in the form of $n - r$ parameters where $\mathrm{rank}(A) = r$. For

ISBN: 978-989-95079-6-8

example, $2x_1 + x_2 = 3$ has solutions as set $S = \{(t, 3-2t) \mid t \in \mathrm{R}\}$. This leads unbounded objective value for the linear programming (1). However, if the constraint set is in the form

$$[A]x = [b];\ x \geq 0, \tag{9}$$

the LP might be bounded (even if it has unbounded feasible set).

Suppose that a system of linear equation (9) is solvable by corollary 3.1. Each system (8) has infinitely many solutions of dimension $n - r$ which are bounded below by $x = 0$. For each $y \in Y_m$, let $S_y = \{x_y \mid (A_c - T_y\Delta)\, x_y = b_c + T_y\delta,\ x_y \geq 0\}$. Then an outer set $B^O = \mathrm{Conv}\{S_y \mid y \in Y_m\}$.

2) An inner set, B^I: To build an inner set, B^I, we also consider two cases when $m = n$ and when $m < n$.

- When $m = n$, we work through the following steps.
  - Step 1 : Choose x_* as one of the point $x_y^1 - x_y^2$ in theorem 3.1 that creates B^O.
  - Step 2 : Find A_* and b_* such that $A_*x_* = b_*$ The matrix A_* and vector b_* is one of the endpoint matrices and vectors created by endpoints of $[A]_{m\times n}$ and $[b]_{n\times 1}$, respectively.
  - Step 3 : An inner set, B^I, is the set $\mathrm{Conv}\{x^* \mid A_*x^* = b^*\}$. There are $\begin{pmatrix} m \\ 0 \end{pmatrix} + \begin{pmatrix} m \\ 1 \end{pmatrix} + \begin{pmatrix} m \\ 2 \end{pmatrix} + \ldots + \begin{pmatrix} m \\ m \end{pmatrix}$ different b^*'s created by changing any component(s) of b_* to the other endpoint(s). For example, if $[b] = \begin{bmatrix} [1,3] \\ [0,2] \end{bmatrix}$ and $b_* = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then the 4 choices of b^* are b_*, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$.
- When $m < n$, we work through the following steps.
  - Step 1 : Choose two elements y^1 and y^2 in Y_m.
  - Step 2 : $B^I = \mathrm{Conv}\left\{S_{y^1}, S_{y^2}\right\}$.

To see that these B^O and B^I are random sets, we define $m(B^I) = 1$ and $m(K) = 0$ for other σ-algebra K's of B^O.

B. General case

When we have mixed uncertain information for (5); for instance, when $\breve{A}$ is an uncertain matrix of 1×2 dimension such that $\breve{a}_{11} := \hat{a}_{11}$ with some possibility distribution and $\breve{a}_{12} := [\underline{a}_{12}, \overline{a}_{12}]$, we need to consider each component separately by using α-levels. The semantics for each α-level of different types of uncertainty is as follow. (Similar idea applies for $\breve{b}_i$).

1) Fuzzy uncertainty: An α-level set, $\left[\underline{a}_{ij_\alpha}, \overline{a}_{ij_\alpha}\right]$, of a fuzzy number $\tilde{a}_{ij}$ represents an interval that every element in it has degree of satisfaction equal to or greater than α.

2) Possibilistic uncertainty: An α-level set, $\left[\underline{a}_{ij_\alpha}, \overline{a}_{ij_\alpha}\right]$, of a possibility distribution $\hat{a}_{ij}$ expresses that every element in the set has degree of occurrence equal to or greater than α.

3) Probabilistic uncertainty: A probabilistic component $\check{a}_{ij}$ with the corresponding density function f. Define $h := \max\{f(a) \mid a \in \check{a}_{ij}\}$, an interval $\left[\underline{a}_{ij_{\alpha h}}, \overline{a}_{ij_{\alpha h}}\right]$ coincides with the α-level set of two previous cases and it means that $\forall a \in \left[\underline{a}_{ij_{\alpha h}}, \overline{a}_{ij_{\alpha h}}\right],\ f(a) \geq \alpha h$.

At each α-level, the uncertainty acts as an interval. However, each interval has a history of the original type of uncertainty, (III-B.1, III-B.2 or III-B.3). So the semantics must be different from the interval case.

We present B^I_α and B^O_α when $\alpha = 0, \frac{1}{k}, \frac{2}{k}, \ldots, 1,\ k \in \mathrm{N}$. We define $[A]_\alpha$ as a matrix uncertainty whose each component is restricted to that particual α-level. (Similar definition applies to $[b]_\alpha$). Moreover, if system $[A]_\alpha x = [b]_\alpha$ is solvable, we write $\mathbf{C}$ in the result of the theorem 3.1 as $\mathbf{C}_\alpha$.

To generate B^O_α, we start with $\alpha = 0$. That means we consider a system $[A]_0 x = [b]_0$. Suppose that this system is solvable, we can create B^O_0 as shown in the interval case III-A.1. Therefore it is not hard to see that $B^O_{\frac{i}{k}} \supseteq B^O_{\frac{j}{k}}, \forall\ i \leq j \leq k,\ i, j, k \in \mathrm{N}$.

In the case of B^I_α, we adopt the construction in III-A.2 as follows.

- When $m = n$:
  - Step 1 : When $\alpha = 0, (i = 0)$, we deal with $[A]_0 x = [b]_0$. We use the method in III-A.2 for finding B^I_0. We also get x_*, A_* and b_* s.t. $A_*x_* = b_*$.
  - Step 2 : Set $c := x_*$ as a reference point.
  - Step 3 : Set $i := i + 1$. Now we deal with $[A]_{\frac{i}{k}} x = [b]_{\frac{i}{k}}$. Choose the point x_* that creates $B^O_{\frac{i}{k}}$ so that $\|x_* - c\| = \min_x \{\|x - c\|\}$, where x is the other point $x_y^1 - x_y^2$ of $B^O_{\frac{i}{k}}$. Follow Step 2 in III-A.2.
  - Step 4 : An inner set $B^I_{\frac{i}{k}} = B^I_{\frac{i-1}{k}} \cap \mathrm{Conv}\{x^* \mid A_*x^* = b^*\}$ where b^*'s are the different endpoints of $[b]_{\frac{i}{k}}$.
  - Step 5 : Go back to Step 2 , or quit when $i = k$.
- When $m < n$: At each i^{th} iteration, we need to choose $y^1, y^2 \in Y_m$ so that $B^I_{\frac{i-1}{k}} \cap \mathrm{Conv}\left\{S_{y^1}, S_{y^2}\right\} \neq \emptyset$. Then $B^I_{\frac{i}{k}} = B^I_{\frac{i-1}{k}} \cap \mathrm{Conv}\left\{S_{y^1}, S_{y^2}\right\}$.

Using this construction, we have $B^I_{\frac{i}{k}} \supseteq B^I_{\frac{j}{k}};\ \forall i \leq j \leq k,\ i, j, k \in \mathrm{N}$.

Therefore the outer and inner sets maintain the following properties.

Theorem 3.2: For all $i \leq j \leq k$ and $i, j, k \in \mathrm{N}$,

$$B^O_{\frac{j}{k}} \subseteq B^O_{\frac{i}{k}} \text{ and } B^I_{\frac{j}{k}} \subseteq B^I_{\frac{i}{k}}.$$

Proof: Follow directly from the construction of the sets. $\diamond$

Using $m(B^I_{\frac{i}{k}}) = \frac{1}{k}, \forall\ i = 1, 2, \ldots, k$ and $m(K) = 0$ for other σ-algebra K's of B^O_0, we create the random set $\left\{B^0_{\frac{i}{k}}, i = 0, 1, \ldots, k\right\}$ and the random set $\left\{B^I_{\frac{i}{k}}, i = 0, 1, \ldots, k\right\}$.

ISBN: 978-989-95079-6-8

IV. Interval-valued probability measure on the set $B_\alpha^O \backslash B_\alpha^I$

In this section we consider each α-level separately. For convenience, we drop the subscript α of B_α^O and B_α^I.

After we have an inner set B^I and an outer set B^O, we consider the space between these two sets, $B^O \backslash B^I$, as our universal space of consideration. Let X be a random variable with an unknown distribution. For any space $B \subseteq B^O \backslash B^I$, there is a probability that $X = x \in \Omega_{\exists\exists}$ is in B. We write this probability as $\Pr(x \in B)$. However, we lack information to justify for sure what the value is.

Each of the figures 1(a) and 1(b) illustrates a geometric interpretation of B^I and B^O in R^2 of a system $[A]x = [b]$ where $[A]$ is regular with $m = n = 2$. Inner sets are the black areas while outer sets are the areas of rectangular boxes (that look like squares in the figures below). It can be difficult to give a geometric interpretation of B^I and B^O, especially in higher dimension.

The points p, q, r and s that create B^O in each figure are $x_y^1 - x_y^2$ where (x_y^1, x_y^2) is a solution of each system (7) in theorem 3.1. As we said in the construction of an outer set that it is easier to get a rectangular box in practice. Therefore, we use the rectangular box in 1(b) to represent B^O instead of Conv $\{p, q, r, s\}$.

When we choose $x_* = p$ in the construction of an inner set, we obtain an inner set as the parallelogram black set in both 1(a) and 1(b) as can be seen.

A point z is more likely to be in the set $\Omega_{\exists\exists}$ than a point y that is further away from B^I. We can create an possibility distribution of a particular nested sets by using this fact.

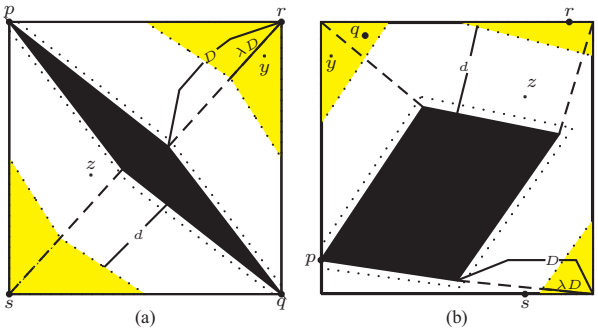


Fig. 1. Inner and outer boxes: inner boxes are the black areas and outer boxes are the rectangular sets.

We choose two corresponding corner points of B^I and B^O that make the longest distance, i.e., the length of one of dashed lines in figure 1(a) (or (b)). We use D to represent this longest distance. Let $0 \leq \lambda \leq 1$ be a scalar. An easy possibility distribution of the nested sets $V(\lambda D)$ could be a function in the term of λ, for instance,

$$\text{Poss}\,(V(\lambda D)) = \lambda, \quad \text{when} \;\; 0 \leq \lambda \leq 1. \tag{10}$$

$V(\lambda D)$ in the domain of the function (10) represents the yellow areas (volume for higher dimensions) in the Figure 1. For $\lambda_1 \leq \lambda_2$, we have $V(\lambda_1 D) \subseteq V(\lambda_2 D)$. When λ is increasing, the possibility that $V(\lambda D)$ contains solutions $x \in \Omega_{\exists\exists}$ is also increasing.

By restricting to the nested domain, we would be able to work in the white space between B^I and B^O. The white space has the width from the edge of the black and yellow areas to be d, as shown in figure 1. The distance d is less than or equal to $(1 - \lambda)D$. We also can use an appropriate necessity distribution, Nec $(V(\lambda D))$.

However, in order to satisfy the definition of an IVPM, we need to make sure that the unknown probability measure is in the bound of our necessity and possibility measures. Therefore, the equation (10) may not be appropriate for most of the systems. We leave this part to the decision maker to come up with a reasonable nestedness necessity and possibility distributions that (s)he thinks they are an lower and upper bounds on the unknown probability.

Thus an IVPM with the restricted nested domain $V(\lambda D)$ is

$$i_{\text{m}}(V(\lambda D)) = [\text{Nec}\,(V(\lambda D)), \text{Poss}\,(V(\lambda D))] \tag{11}$$

where Poss $(V(\lambda D))$ and Nec $(V(\lambda D))$ are defined to cover the unknown probability by the expert's opinion. In the next section, we show the use this IVPM in LP problems.

V. Optimization on boxes using IVPMs

To deal with an optimization problem (1), it would make sense to consider on each α-level of each uncertainty.

As we know for sure that the solutions in B_α^I have their correspondent matrix A and vector b, one approach for this problem (1), restricted on the constraint set B_α^I, can be

$$\min_{x \in B_\alpha^I} \check{c}_\alpha x, \tag{12}$$

where $\check{c}_\alpha = [\underline{c}_\alpha, \overline{c}_\alpha]$. By solving $\min \underline{c}_\alpha x$ and $\min \overline{c}_\alpha x$ on the convex polyhedral B_α^I, we obtain the guaranteed bound on the objective value.

Moreover, since $\Omega_{\exists\exists_\alpha} \supseteq B_\alpha^I$, we use the idea in IV to deal with the solution between inner and outer sets. We use a nestedness possibility distribution in a restricted IVPM obtained form IV to determine the space of consideration. We illustrate through the following examples.

Example 5.1:

$$\left.\begin{array}{ll} \min & [2,4]x_1 + [1,3]x_2 \\ \text{s.t.} & \widehat{3}x_1 + [-2,1]x_2 = [-2,2] \\ & \widetilde{0.5}x_1 + \check{3}x_2 = \widehat{0}, \end{array}\right\} \tag{13}$$

where the uncertain data are presented in the figure 2 above. For this example, we work at the α-level when $\alpha = 0.5$. Therefore, the problem (13) with $\alpha = 0.5$ becomes

$$\left.\begin{array}{ll} \min & [2,4]x_1 + [1,3]x_2 \\ \text{s.t.} & [2,4]x_1 + [-2,1]x_2 = [-2,2] \\ & [-1,2]x_1 + [2,4]x_2 = [-2,2]. \end{array}\right\} \tag{14}$$

Figure 3 shows the inner box and outer box for (14). So the guaranteed bound on the objective value of

$$\min_{x \in B_{0.5}^I} [2,4]x_1 + [1,3]x_2,$$

is $[-25, 25]$. From figure 3, we get $D = 5$ units. By using possibility distribution as in (10) for simplicity, if we want

ISBN: 978-989-95079-6-8

degree of possibility of an IVPM (as constructed in IV) to be 0.6, then

$$\min_{x\in\left(B^O_{0.5}\setminus B^I_{0.5}\right)_{0.6}} [2,4]x_1 + [1,3]x_2$$

would have the guaranteed bound, $[-28, 28]$, on the

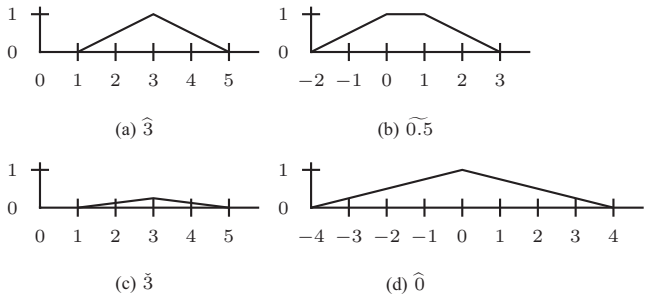


Fig. 2. Uncertain data for example 5.1.

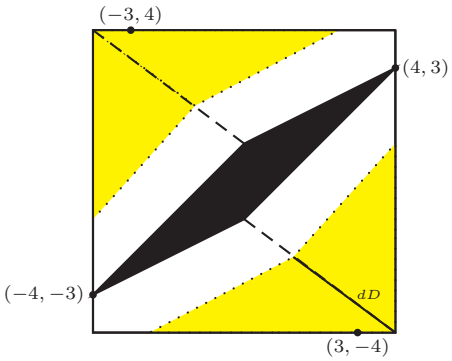


Fig. 3. Inner and outer boxes for the problem (14). Inner box is the black area. Outer box is the square generated by points (-3,4), (4,3), (3,-4) and (-4,-3).

objective value with 0.6 degree of occurrence. Similarly, we can have the guaranteed bound with a certain degree of belief by using an appropriate necessity measure of the IVPM.

Example 5.2: A drug company wants to create two supplement products X and Y that extract mainly from a super food F. This super food F has a fuzzy color 'orange-red-pink'. Some of fully grown F may have the color red to pink or could be more in the orange shade. The company will be totally satisfied if each of the plant F has red color. So the fuzzy color (from the company's point of view) has the membership function as in figure 4(a). There is only one place P that grows F. So P can charge the price that he wants. Although P can not predict his harvest's color, he also prefers F to be red to please his customer. P sets the price of each F relating to its fuzzy color as in figure 4(b).

The company found out that each 1g of F (depend on the color) contains an amount of Omega III and Zinc (in mg) with possibility distribution in figure 4(c) and 4(d), respectively. The company has a secret formula for X and Y. To produce X, the company need to make sure that the basic ingredient has Omega III in the range of [1280,1300]mg. The product Y need to have $\widehat{586}$mg of Zinc, (Fig. 5(a)).

Therefore the company needs to add pure Omega III and Zinc from a substance T if F does not provide enough as required. One gram of T obtains 5mg of Omega III and $\check{1.8}$mg of Zinc , (Fig. 5(b)), which means Zinc in T is not certain. The price of T depends on the market which is given by $\$\check{3}$ per 1g of T, (Fig. 5(c)). We can formulate a

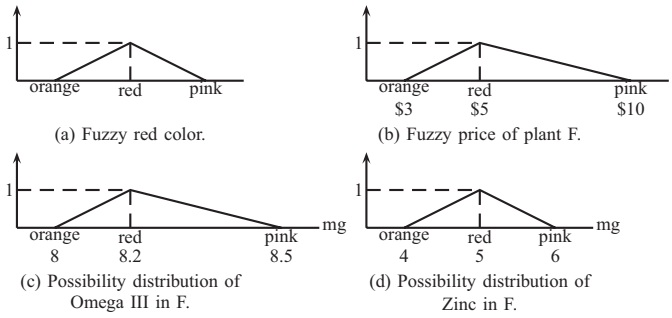


Fig. 4. Membership functions and possibility distributions.

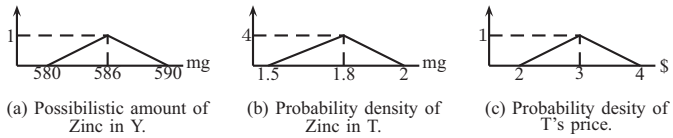


Fig. 5. Possibility distribution and probability density functions.

linear program to minimize the cost to produce these two supplements.

$$\begin{aligned} \min \quad & \widetilde{5}\mathrm{f} + \check{3}\mathrm{t} \\ \text{s.t.} \quad & \widehat{8.2}\mathrm{f} + 5\mathrm{t} = [1280, 1300] \\ & \widehat{5}\mathrm{f} + \check{1.8}\mathrm{t} = \widehat{586} \\ & \mathrm{f}, \mathrm{t} \geq 0, \end{aligned}$$

where f is the amount of grams of plant F and t is the amount of grams of substance T. When $\alpha = 0.5$, the problem becomes

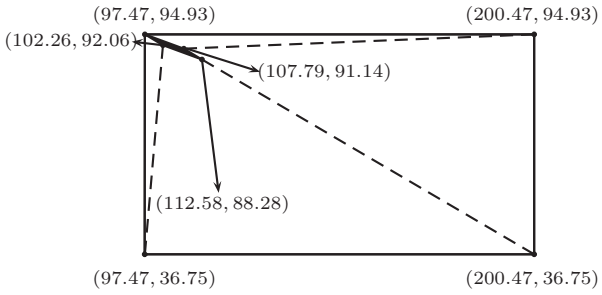


Fig. 6. Inner and outer boxes for the problem (15). Each point represents (t,f).

$$\left.\begin{array}{ll} \min & [4, 7.5]\mathrm{f} + [2.5, 3.5]\mathrm{t} \\ \text{s.t.} & [8.1, 8.35]\mathrm{f} + 5\mathrm{t} = [1280, 1300] \\ & [4.5, 5.5]\mathrm{f} + [1.65, 1.9]\mathrm{t} = [583, 588]. \end{array}\right\} \quad (15)$$

The inner and outer sets of the system (15) is shown in figure 6. The inner set is the black space generated by points (97.47,94.93), (102.26,92.06), (112.58,88.28) and (107.79,91.14). The outer set is the rectangular box. However, points (200.47,94.93) and (97.47,36.75) are not feasible for any $A \in [A]$ and $b \in [B]$.

ISBN: 978-989-95079-6-8

The semantics explanation for this example is as follows. The company sets the degree of satisfaction of F's price to be at least 0.5 together with 75% confidence interval of probabilistic price of T. It also wants 0.5 or more degree of occurrences of Omega III and Zinc in the food F and 93.75% confidence interval of probabilistic amount of Zinc in T. By its secret formula, the company sets at least 0.5 degree of occurrence of Zinc in product Y. Then the guaranteed bound on the objective value with $\alpha = 0.5$ is $\$[623.40, 1060.82]$ by using the corner point (97.47,94.93) and (107.79,91.14) of the black space. We also can find the guaranteed bound for some curtain degree of occurrence in the similar way as presented in the example 5.1.

VI. Conclusion and Further Research

From the construction of B^I_α and B^O_α, we can use B^I_α to find a guaranteed set of solutions and bound on an objective value of an optimization problem. We use the set $B^O_\alpha \setminus B^I_\alpha$ to create an IVPM, then use a possibility to get a set of solutions with some level γ of occurrence. This paper considered merely equality constraints. The next step is to consider the inequality case.

References

[1] M. Fiedler, J. Nedoma, J. Ramík, J. Rohn and K. Zimmermann, "Linear optimization problems with inexact data" *Springer,* 2006.

[2] D. M. Gay, "Solving interval linear equation" *SIAM J. Number Anal.* no. 4, vol.19, pp. 858-870, 1982.

[3] Edited by P. W. Hawkes, "Advances in Imaging and electron physics" *Elsevier Academic Press, NY* vol. 148, pp. 75-192, 2007.

[4] W. A. Lodwick and K. D. Jamison, "The construction of consistent possibility and necessity measures" *Fuzzy Sets and System.* vol. 312, pp. 1-10, 2002.

[5] W. A. Lodwick and K. D. Jamison, "Interval valued probability in the analysis of problems containing a mixture of possibilistic, probabilistic and interval uncertainty" *Fuzzy Information Processing Society, 2006. NAFIPS 2006. Annual meeting of the North American.* pp. 617-620.

[6] W. A. Lodwick, "Constrained interval arithmetic" *CCM Report, 138.*

[7] R. E. Moore, *Methods and Applications of Interval Analysis.* SIAM, Philadelphia, 1979.

[8] J. Rohn, "Solvability of systems of linear interval equations" *SIAM J. Matrix Anal. Appl.* no.1, vol. 25, pp. 237-245, 2003.

[9] P. Thipwiwatpotjana and W. A. Lodwick, "Algorithm for solving optimization problems using interval valued probability measure" *2008 NAFIPS Conference Proceedings*

[10] P. Walley, *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[11] Z. Wang and G. J. Klir, *Fuzzy Measure Theory.* Plenum Press, New York, 1992.

[12] K. Weichselberger, "The theory of interval-probability as a unifying concept for uncertainty," *International Journal of Approximate Reasoning.* no.24, pp. 149-170, 2000.

ISBN: 978-989-95079-6-8

Human perceptions versus computational perceptions in Computational Theory of Perceptions

Gracian Trivino[1] Alejandro Sobrino[2]

1.European Centre for Soft Computing
Asturias, Spain
2.University of Santiago de Compostela
Galicia, Spain
Email: gracian.trivino@softcomputing.es, alejandro.sobrino@usc.es

Abstract— *The development of a computational system based on the Computational Theory of Perceptions is a challenging task that requires taking into account knowledge extended to several scientific disciplines such as Fuzzy Logic, Linguistics and Software Engineering among others. In this paper we contribute to the Computational Theory of Perceptions by providing a set of ideas and definitions which could be useful as components of a future methodology for the development of practical applications of this theory.*

Keywords— Computational Theory of Perceptions, Computing with Words, Fuzzy Logic.

1 Introduction

The Computational Theory of Perceptions (CTP), outlined in the Zadeh's seminal paper [1] and further developed in subsequent papers [2], deals with the automated processing of human knowledge expressed in Natural Language (NL). This theory is inspired in the way human beings use NL to make rational decisions and communicate their experience in an environment of imprecision, uncertainty and partial truth.
The general goal of CTP is to develop computational systems with the capacity of computing with the meaning of NL propositions, i.e. with the capacity of computing with imprecise descriptions of the world in the same way that humans do.
When we face the project of design and development of a computational system we must deal with knowledge extended to several disciplines. Typically the designer must solve problems in the fields of Mechanics, Electricity, Electronics and Software Engineering but in this type of project the multidisciplinary scope extends in different degrees to the disciplines grouped around Cognitive Science, namely, Philosophy, Psychology, Artificial Intelligence, Linguistics, Anthropology, Neurophysiology and Education [3].
Unfortunately it is not easy to obtain a practical perspective of CTP in such a wide context. However ignoring this global perspective could lead to misunderstandings and delays in undertaking the development of this new theory.
This paper tries to contribute with a brief analysis of several concepts of CTP from the point of view of its practical application. We present an ordered set of concepts where the description of the essence of CTP is complemented with ideas taken from Systemic Functional Linguistics and Software Engineering among other disciplines.
We analyze five concepts going from the most general to the most practically applicable.
In section 2 we analyze briefly the computational use of NL from both perspectives, namely, as a tool for describing human experience and as a tool for enacting social processes. Usually, the implications of this analysis are not explicitly accounted in the development of the current computational systems based on NL.
In section 3 we analyze the concept of computational perception versus human perception. We use this concept in our search of the "basic unit of meaning" and the different ways of representing it.
In section 4 we propose a practical hierarchy of levels of Precisiation. The concept of Precisiation by Zadeh is interpreted as a process of aggregating constraints to go from the most general meaning to the most specific.
In section 5 we introduce the practical concept of Linguistic Fuzzy Transducer as a computational module with capacity of translating between two different descriptions of the reality.
Section 6 provides our conclusions

2 Computational use of Natural Language

NL has evolved in parallel with the human brain as a product of natural evolution. NL is not the result of the rational design of human beings as is the case of Mathematics or Music Notation. NL is a dynamic system that can not be explained simply as the sum of its parts [4]. Typically, we do not use a set of formal rules to produce our NL utterances. We could say that a human speaker produces his/her discourse following a chaotic procedure similar to the one used by a tree to build branches and leaves.
Taking the perspective of the developer of CTP applications let us briefly analyze the two complementary functions of language: construing experience and enacting social processes.

2.1 Experience

Regarding the first function, in agreement with Systemic Functional Linguistics (SFL), NL is a resource with which human beings construct the mental maps of their phenomenal world, of their experience of process: what goes on out there and what goes on in the realms of their own consciousness. The experience is a resource of meaning, a potential of understanding, representing and acting on the world. The particulars of our daily life make sense because they are instantiations of this potential. The experience includes descriptions of two layers of reality: First order phenomena related directly with the environment and the second order phenomena formed of

ISBN: 978-989-95079-6-8

the meanings and wordings that perceptions of first order phenomena bring into being [5].
During the last thousands of years, human evolution has been supported not only by genetic transmission but essentially by cultural learning [6]. It is well known that, in this process, the invention of the printing press produced a revolutionary push. During recent decades, computers have acquired an important role as a medium of storing and processing information boosting this process, i.e. computers and the Internet are used as the new encyclopedia.
But we believe computers will have a more important role as contributors to mankind evolution. The next challenge consists of using computers as a tool where NL does not have the general meaning provided in an encyclopedia but the particular meaning that a specific user, with a specific personal experience, requires in a specific context of use.
The achievement of this goal will convert computers into personal assistants, into personalized tools to help us to go far in the construction of our personal experience, i.e. in our personal understanding of the world.
From the perspective of computational systems developers, we will use the term "World Model" for a computational representation of reality, a representation of the context where the computer must perform its function. We can see this World Model as a projection of the programmer's experience into the computer memory.

2.2 Human Computer Interface

Regarding the second function of NL, enacting social processes, the computer can be considered as a new type of partner in the NL communication process typically performed between human beings.
When using a computer with this role, we must consider three modes to communicate meaning with NL [5]:
1. The domain of language of the computational application. We should use the concepts and associated words of the culturally recognized repertory of social practices and concerns. Moreover, we should take into account the personal use of NL, i.e. the personal experience of the specific user at whom the application is aimed.
2. The relationship established between user and computer. When using NL we should use the culturally recognized repertory of role relationships and interactive patterns corresponding with a specific speaker - listener relationship. We should choose and design the role the user expects to be adopted by the computer. That is to say, we must assign the computer the role of "secretary", "weather broadcast speaker", or perhaps a new type of role that still should be carefully designed and explained.
It is important to realize that, when humans communicate, they establish a theory about the mind of the other, about the other's experience and intentions, which is absolutely relevant to interpret of the meaning of the NL utterances [6].
3. The mode of expression. We must consider both, the medium (written, spoken) and the rhetorical function (persuasive, didactical, informative, etc.) of the language used. Additionally, we should consider the possibility of complementing the use of NL with graphics and sounds, and the possibility of expressing emotions with the tone of voice and the face expressions of an avatar.

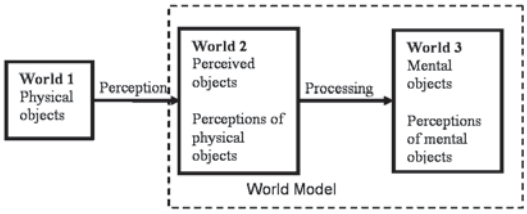


Figure 1: The three worlds by Popper.

3 Computational Perceptions

In agreement with Popper, we can distinguish three worlds: The first is the unknown world of what is going on out there, i.e. the world of the objects in our environment; the second is the world of the perceived objects, i.e. the world of information that we acquire thanks to our senses; and the third is the world of the mental objects that we create in the realm of our consciousness, i.e. they are abstract objects built using NL and that constitute an important part of our Experience (See Fig. 1) [7].
In our approach we can say that we construe our experience using perceptions. Moreover we can say that, remembering or figuring out a mental object is a form of perception. This is because NL is basically a system for describing perceptions [8].
If we define *to perceive* as: "to attain awareness or understanding of elements in the environment" [9] then we can say that *to perceive using a computational system* is: the process performed using a computer to obtain information and to produce a representation useful to the system designer, and therefore for the human operator, to attain awareness or understanding of elements in the environment".
Therefore the programmer and the final user of the computational system are the interpreters of the information provided by the computer. The human uses the computer as a tool to perceive the reality in a comparable way that he/she uses a clock to know about the passage of time or a microscope to detect the presence of bacteria in a drop of water.

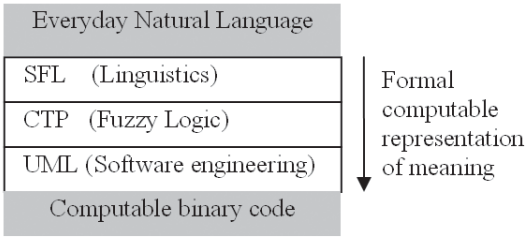


Figure 2: CTP occupies a position between Linguistics and Software Engineering

3.1 Granule

In CTP, granules are used to describe perceptions. The basic unit of meaning is a granule. A granule is a clump of elements which are drawn together by indistinguishability, similarity, proximity or functionality [10].
The perception of an object can be described either only as a granule, or composed by a collection of granules. A granule is a node in a network of relations with other granules. The

ISBN: 978-989-95079-6-8

granularity of the description of objects allows us to deal with real world problems considering different degrees of detail. For example, we summarize the description of an object by hiding the irrelevant granules and remarking on the relevant ones. Indeed, in this sense, the description of an object, i.e. a perception, is always a summary.
Fig. 2 shows three disciplines devoted to NL meaning representation where CTP occupies a relevant position between Linguistics and Software Engineering. Linguistics is needed to face the complexity of a normal conversation, a tale, etc. Unified Modelling Language (UML) is a language developed in Software Engineering during the last decades to develop computational models of reality and it is a type of tool needed when we deal with the development of complex computational applications.
In SFL the basic unit of meaning is a *figure* (figure with the meaning of understanding of the verb figure out). A figure has three types of elements: participants, processes and circumstances [5].
In UML the basic unit of meaning is a *class*. A class has associated attributes and operations [11].
Figures and classes are types of granule and we can use them to computationally describe perceptions.

3.2 Information as constraint

In CTP, granules have fuzzy attributes, i.e. the borders of a granule are fuzzy. Fuzziness of granules, their attributes and their values allow us to model the way in which human concepts are formed, organized and manipulated.
Granules are defined using fuzzy constraints on the possible values of their attributes. In this sense, in CTP information is expressed as a set of constraints [12]. The elasticity of constraints is obtained by using Fuzzy Logic to define them. This elasticity is intended to be a reflection of the same elasticity found in the meaning of words in NL.

3.3 Computational Perception

As mentioned above, a computational perception (perception for short) is a representation of the information obtained about an object with a level of granularity useful for the programmer's purposes.
In CTP there are two different ways of describing a perception: on the one hand using NL and on the other hand using a formalized language which computers can process.
The first description is the natural way of describing the world. NL has evolved with the human being acquiring capacity to understand and communicate complex world descriptions [5], i.e. NL contains all the required resources to describe the objects that we can perceive or imagine, their properties and behaviour.
The second description is a challenge that still lacks of a general solution. It consists of a formalized computable representation with equivalent meaning to the NL propositions. In CTP this language is named Generalized Constraint Language (GCL) [13].
A key element of GCL is the concept of linguistic variable [14]. The values of a linguistic variable are granules. Assigning value to a linguistic variable consists of defining a constraint on the set of its possible values.
In GCL the meaning of a proposition p is expressed as:

$$XisrR$$

where

- *X* is the constrained linguistic variable whose structure of values ranges from the simple linguistic label required to represent a perception associated with the value of a sensor to a complex structure of information, namely, a vector of propositions, a fuzzy graph, a function, etc.
- *R* is the constraining relation
- *isr* is a copula in which *r* is a variable indicating the type of constraint. Primary constraints are formalizations of three basic perceptions: perception of possibility; perception of likelihood; and perception of truth.

4 Precisiation

Granularity allows us to order perceptions by their meaning. At one extreme we have the set of all possible meanings and at the other extreme a simple text with a the meaning of a specific perception, i.e. using constraints we construe a text as an instance of the whole available potential of meaning.
Precisiation consists of aggregating constraints until a description is provided of a perception useful for a specific propose. Precisiation is a form of modelling perceptions by aggregating constraints .
As mentioned above we can see experience as a network of interrelated granules. We describe a perception by defining a sub-network of related granules. Precisiation requires determining a level of granularity, i.e. the composition of this sub-network.
From a practical point of view, one important relation, that we could consider the first in the Precisiation process, is the relation "a type of". This relation organizes the potential of meaning into the hierarchal order of Instantiation - Generalization.
Once determined the position of the perception in this dimension we can continue introducing constraints in other dimensions depending on every specific application such as: "is part of", "causality", "dependency", etc.
In SFL this network of interrelated granules is called a System Network. A System Network is an acyclic directed graph, consisting of elements of meaning partially ordered in hierarchies using relations of meaning [5].

4.1 Levels of Precisiation

Precisiation is a multidimensional process and the ranking is continuous rather than discrete, i.e. there are so many levels of Precisiation as practical uses of NL. From the practical point of view of organizing the design of a computational system based on CTP we have defined the hierarchy of levels represented in Fig. 3.

Domain of Experience/Language
This first level of Precisiation consists of delimiting a Domain of Experience. A Domain of Experience is a closed domain where the process we like to describe takes place. A Domain of Experience defines a specific context, a set of possible situations and corresponds with a Domain of Language. The

ISBN: 978-989-95079-6-8

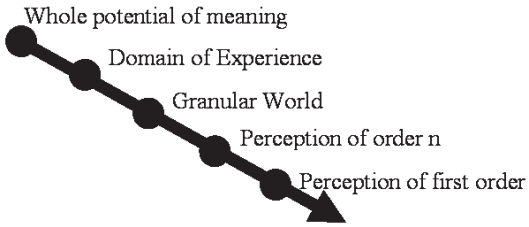


Figure 3: A hierarchy of levels of precisiation

concept of Domain of Language is related with the concept of language-games by Wittgenstein [15].
A Domain of Language contains a subset of possible meanings, each domain has associated a corpus of NL expressions and linguistic patterns. They are at intermediate level in the ordered structure of granularity. For example, the "language of cooking" is an instantiation of the "whole language", the "language of cooking deserts" is a finer level of instantiation, and "the language of cooking apple cakes" is even a more detailed instantiation.

Granular World

A Granular World (GW) contains a subset of meanings belonging to a Domain of Experience. See in [16] for interesting research on the concept of GW.
In our approach this is a quite technical concept taken from Control Theory. We can see GW as an n-dimensional State Space where we must define the needed variables and procedures to create a computational model of the system we want to describe.
A basic element of a GW is the Vector of Linguistic State Variables (SV). Every perception in a GW can be described as a combination of constraints on the possible values of these variables. The application of Fuzzy Logic in Systems Theory was introduced by Zadeh several decades ago [17].
Complex perceptions in a GW are granules built with sets of lower order granules. In a GW the explanation of a perception is a set of more detailed perceptions that describes its meaning with deeper degree of granularity.
The variables belonging to SV provide the most detailed possible explanation of a perception in a GW.

Perception of first order

Perception of first order (p^1) in a computational system is a perception directly associated with the value provided by a sensor at an instant in time.
The basic form of a perception of first order is a proposition that describes the subjective value of a measure, e.g. "The temperature is High" is a subjective description, i. e. for a specific user in a specific use, when the sensor measure reads 45^{o}C.
There are two forms of representing p^1:
The linguistic representation, e.g.:

$$p^1\text{: "The Temperature is High"}$$

And the formal representation:

$$p^1\text{: } X = \mu_{A_i}(x)$$

where:

- X is a linguistic variable (e.g. Temperature).
- A_i is a linguistic term belonging to the set of possible linguistic values of X (e.g. {Low, Medium, High}).
- $\mu_{A_i}(x)$ is the membership function associated with the linguistic term A_i.
- x is a numerical value obtained from the sensor (e.g. 45^{o}C).

We say that the meaning of a perception of first order p^1 can be *explained* using numerical values obtained from sensors and a set of membership functions that covers the domain of these possible values. Because perceptions of the first order work directly upon sensor values, we say these perceptions constitute the most detailed linguistic description of a signal, i.e. it has the finest granularity. Obviously, the linguistic variables associated with first order perceptions are good candidates to belong to the SV of a GW.

Perception of order n

Granularity of perceptions allows us to create a hierarchy describing complex perceptions using sets of more detailed ones.
For example, two first order propositions
p_1^1: "The Temperature is Warm"
and
p_2^1: "The Humidity is Medium"
could be used to explain the meaning of
p^2: "The Room is Comfortable".
Typically this explanation has the form of a rule:
IF p_1^1 AND p_2^1 THEN p^2.
We could extend easily this example by considering other perceptions such as "Acoustic noise" or "Number of persons in the room" to construe the *explanation* of a more complex n-order perception of Comfort.

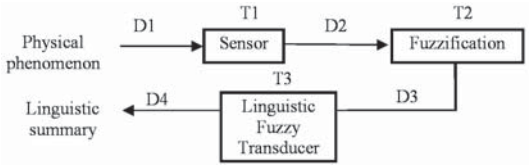


Figure 4: A chain of transducers (T) to translate between several Domains of Language (D).

5 Linguistic Fuzzy Transducer

In our context a Linguistic Fuzzy Transducer is a computational system capable of translating between two different representations of a perception.
Humans are good translators between perceptions described in different domains of language, e.g. when a teacher explains his matter he is translating his perceptions to the domain of language that he expects the students have. A system with the capacity of translating between two natural languages is a transducer. A question answering system could be considered as a transducer with the capacity of providing detailed explanations of a given perception. A system with the capacity of summarize information is a transducer.
Unfortunately to build a computational version of a Linguistic

ISBN: 978-989-95079-6-8

Transducer is still a challenge without a general solution.
Fig. 4 shows a sequence of transducers which convert a perception of the physical environment into representations in different domains of language with different levels of granularity. Indeed, the figure represents a family of computational systems with the capacity of generating a linguistic summary of data provided by sensors. D1 is the domain of language of Physics, D2 is the domain of numerical values provided by sensors, D3 is essentially a domain of first order perceptions obtained after a process of fuzzification and D4 is the domain of language where the final user will use the information.
A Linguistic Fuzzy Transducer has the following basic elements:

- GW1: The Granular World of the input represented by its associated SV1.
- GW2: The Granular World of the Output represented by its SV2.
- Linguistic Fuzzy Model: A model is a simplification of reality constructed using our perceptions. Usually we need to analyze the system at different levels of granularity to understand its functioning. In Fuzzy Logic this type of model is called Linguistic Fuzzy Model that typically can be implemented as a set of fuzzy rules that control the system evolution in time. [18].

The rest of this section describes a simple example of the type of computational system represented in Fig. 4.

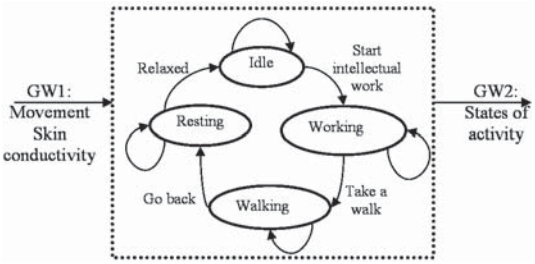


Figure 5: Transducer in the example of application

5.1 Example

This experiment has been performed in the setting of the daily routine of a person dedicated to sedentary work: reading, writing and managing a computer in an office environment. A typical situation that we would like to model is the following:

1. The subject starts his/her day relaxed at the desk.
2. The subject turns on the computer, reads and answers emails, reads papers, writes notes and so on.
3. Initially the subject will remain seated and tension levels rise slowly.
4. After time the tension starts to provoke difficulties in concentration. The subject starts to move in his/her chair and shows behavior such as going to the toilet or moving paperwork around the office.
5. Then the subject decides to take a small walk outside 'to clear his/her mind'.
6. Finally the subject returns to his/her office desk and continues working, again in a less stressful state (similar to the starting point).

Fig. 5 shows a diagram that represents the evolution of the activity and state of the subject during the experiment. Note that we consider the subject able to recover part of the initial 'relaxed state' by doing some physical exercise.
The goal consists of monitoring the physical activity, the levels of stress and mental tiredness of the subject and to create a linguistic report summarizing the temporal evolution of these parameters. This experiment provides the first steps into the development of a computational system with the objective to help the user to increase his/her effectiveness and satisfaction [19].
Let us see briefly how to obtain, in this case, the main components of the last transducer (T3) in Fig. 4:
GW1: The subject were wearing two sensors: A skin conductivity meter fixed attached to the left wrist, with electrodes attached to the index finger and the middle finger. The accelerometer was kept in the chest pocket of the subject's shirt. The subject was instructed to follow the steps of the experiment as described above.
After the fuzzification process we have two fuzzy linguistic variables which constitute the vector: SV1=(Activity, Skin Conductivity).
GW2: is represented by the vector SV2=(Idle, Working, Walking, Relaxing).
LFM: We have built this LFM using a Fuzzy Finite Machine (FFSM) (see [19] for more details) where the states were labeled as: q_1: Idle, q_2: Working, q_3: Walking and q_4: Relaxing. The evolution of this FFSM is controlled by the following set of fuzzy rules:
R_{11}: IF (Q IS q_1) AND (Activity IS Low) AND (Conductivity IS Low) THEN Q IS q_1
R_{12}: IF (Q IS q_1) AND (Activity IS Low) AND (Conductivity IS Medium) THEN Q IS q_2
R_{22}: IF (Q IS q_2) AND (Activity IS Medium) AND (Conductivity IS Medium) THEN Q IS q_2
R_{23}: IF (Q IS q_2) AND (Activity IS Medium) AND (Conductivity IS high) THEN Q IS q_3
R_{33}: IF (Q IS q_3) AND (Activity IS High) THEN Q IS q_3
R_{34}: IF (Q IS q_3) AND (Activity IS Low) THEN Q IS q_4
R_{44}: IF (Q IS q_4) AND (Activity IS Medium) AND (Conductivity IS High) THEN Q IS q_4
R_{41}: IF (Q IS q_4) AND (Activity IS Low) AND (Conductivity IS Medium) THEN Q IS q_1
Where R_{ii} are rules to remain in the state and R_{ij} are rules to change.
The output is the template of a simple report as in the following example:
"The subject started **RELAXED** at the desk at **0 minutes**. Then the subject was working during **15 minutes** without signs of tiredness. Around **25 minutes** the subject was moving and appears uncomfortable. At **60 minutes** the subject decided to take a walk outside and went back at **75 minutes**. Around **80 minutes** the subject was relaxing until to be **RELAXED**. About **90 minutes** the subject started to work again."

ISBN: 978-989-95079-6-8

6 Conclusions

The authors are involved in a long term project aimed to contribute to Zadeh's CTP.
The main contribution of this paper consist in providing an ordered set of definitions obtained from different fields that could help the software engineer to understand and organize the project of a computational system based on CTP. These definitions can be used as the base of a future methodology for developing this type of projects.

Acknowledgment

This work has been partially funded by the Foundation for the Advancement of Soft Computing (Asturias, Spain) and Spanish government (CICYT) under grant: TIN2008-06890-C02-01

References

[1] Lotfi A. Zadeh. From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems*, 1:1–10, 1999.

[2] Lotfi A. Zadeh. Toward human level machine intelligence - is it achievable? the need for a paradigm shift. *IEEE Computational Intelligence Magazine*, 1:1–10, 2008.

[3] J.C. Gardner. *The Minds New Science. A History of the Cognitive Revolution*. Basic Books, New York, 1985.

[4] M.A.K. Halliday and M.I.M Matthiessen. *An Introduction to Functional Grammar*. Oxford University, New York, 2004.

[5] M.A.K. Halliday and M.I.M Matthiessen. *Construing experience through meaning*. Continuum, New York, 1999.

[6] M. Tomasello. *The cultural origins of human cognition*. Harvard, New York, 1999.

[7] K.R. Popper and J.C. Eccles. *The Self and Its Brain*. Springer-Verlag, Berlin, 1977.

[8] Lotfi A. Zadeh. Precisiated natural language (pnl). *AI Magazine*, 25(3):74–91, October 2004.

[9] http://www.merriam webster.com. 2008.

[10] Lotfi A. Zadeh. *Fuzzy sets and information granularity*. In: Gupta M. Ragade R, Yager R (eds) Advances in Fuzzy Set Theory and Applications. North-Holland Publishing CO, 1979.

[11] J. Rumbaugh G. Booch and I. Jacobson. *The Unified Modelling Language user guide*. Addison Wesley, 1999.

[12] Lotfi A. Zadeh. Generalized theory of uncertainty (gtu) - principal concepts and ideas. *Computational Statistics & Data Analysis*, 51:15–46, 2006.

[13] Lotfi A. Zadeh. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4:103–110, 1996.

[14] Lotfi A. Zadeh. The concept of linguistic variable and its application to approximate reasoning. *Information Sciences*, 1975.

[15] L. Wittgenstein. *Philosophical investigations*. Blackwell Publishing, 1953/2001.

[16] A. Bargiela and W.G. Pedrycz. *Granular Computing. An Introduction.* Kluwer Academic Publishers, 2003.

[17] Lotfi A. Zadeh. *Toward a theory of fuzzy systems*. In: Klir G.J., Yuan B. (eds) *Fuzzy sets, Fuzzy logic, and Fuzzy systems: selected papers by Lotfi A. Zadeh*. 1996 - World Scientific Publishing Co., Inc. River Edge, NJ, USA, 1971.

[18] R. R. Yager and D. P. Filev. *Essentials of Fuzzy Modelling and Control*. John Wiley & Sons, Inc, 1994.

[19] G. Trivino and A. van der Heide. Linguistic summarization of the human activity using skin conductivity and accelerometers. *Proceedings of the Conference Information Processing and Management of Uncertainty in Knowledge Based Systems. (IPMU2008), June 22-27, Malaga, Spain*, 2008.

ISBN: 978-989-95079-6-8

A characterization of residual implications derived from uninorms

Isabel Aguiló Jaume Suñer Joan Torrens

Departament de Ciències Matemàtiques i Informàtica
Universitat de les Illes Balears
07122 Palma de Mallorca, Spain
Email: {isabel.aguilo,jaume.sunyer,dmijts0}@uib.es

Abstract*— *In this paper, a set of axioms is given that characterizes those functions $I : [0,1]^2 \to [0,1]$ for which a left-continuous uninorm U exists in such a way that I is the residual implication derived from U. A characterization for the particular case when U is representable is also given.

***Keywords*—** Implication function, left-continuity, residual implication, uninorm.

1 Introduction

Implication functions are probably the most important operations in fuzzy logic, approximate reasoning and fuzzy control, because they are used not only to model fuzzy conditionals, but also to make inferences in any fuzzy rule based system (see for instance [1] or [2]). Moreover, they are useful not only in approximate reasoning and fuzzy control, but also in many other fields like fuzzy relational equations and fuzzy mathematical morphology ([3]), fuzzy DI-subsethood measures and image processing ([4] and [5]), and data mining ([6]). An excellent and very recent book on fuzzy implications is [7].

Among other models (see [3]), residual implications and strong implications are the most usual ones. They are usually derived from t-norms and t-conorms respectively as follows:

- Given a t-norm T, the residual implication or R-implication derived from T is given by

$$I_T(x,y) = \sup\{z \in [0,1] \mid T(x,z) \leq y\} \qquad (1)$$

  for all $x, y \in [0,1]$.

  R-implications come from residuated lattices based on the so-called residuation property that can be written as

$$T(x,y) \leq z \qquad \Longleftrightarrow \qquad I_T(x,z) \geq y.$$

  Note that this property is satisfied if and only if T is left-continuous and then the supremum in (1) can be substituted by maximum.

- Given a t-conorm S and a strong negation N, the strong implication or (S,N)-implication derived from S, N is given by

$$I_{S,N}(x,y) = S(N(x),y) \qquad \text{for all} \quad x, y \in [0,1].$$

  In this case, (S,N)-implications appear as a generalization of the classical boolean implication $p \to q \equiv \neg p \vee q$.

Axiomatic characterizations for R-implications derived from left-continuous t-norms, as well as for (S,N)-implications (even when N is only a continuous fuzzy negation, not necessarily strong), have appeared along the time (see for instance [8], [9], [10], [11], [12], [13], and the recent surveys [3] and [14]).

On the other hand, conjunctive and disjunctive uninorms (see [15]) are a class of associative binary aggregation functions that generalize both t-norms and t-conorms, and that has been successfully used in many application fields where t-norms and t-conorms apply. In particular, R and (S,N)-implications have also been derived from uninorms, obtaining in this way new implication functions with nice and interesting properties (see [16], [17] and [18]). However, an axiomatic characterization for these new classes of implications has been done for the case of (S,N)-implications derived from disjunctive uninorms in [19] (again in the general case when N is not necessarily strong), but not for R-implications.

In this paper, we deal with this problem and we give characterizations for the class of R-implications derived from left-continuous conjunctive uninorms in a similar way as it is done for left-continuous t-norms. Moreover, a characterization for the special case of left-continuous representable uninorms is given. From this case we derive a characterization of a kind of quasi-continuous implications (continuous except at two points) that can be viewed as a generalization of the well known Smets-Magrez Theorem (see the Preliminaries).

2 Preliminaries

We will suppose the reader to be familiar with the theory of t-norms, t-conorms and fuzzy negations (all necessary results and notations can be found in [20]). We recall here only some facts on implications.

Definition 1 *A binary operator $I : [0,1] \times [0,1] \to [0,1]$ is said to be an* implication function*, or an* implication*, if it satisfies:*

(I1) $I(x,z) \geq I(y,z)$ *when* $x \leq y$*, for all* $z \in [0,1]$.

(I2) $I(x,y) \leq I(x,z)$ *when* $y \leq z$*, for all* $x \in [0,1]$.

(I3) $I(0,0) = I(1,1) = 1$ *and* $I(1,0) = 0$.

Note that, from the definition, it follows that $I(0,x) = 1$ and $I(x,1) = 1$ for all $x \in [0,1]$ whereas the symmetrical values $I(x,0)$ and $I(1,x)$ are not derived from the definition.

Special interesting properties for implication functions are:

- The *ordering property*,

$$x \leq y \Longleftrightarrow I(x,y) = 1, \quad \text{for all} \quad x, y \in [0,1]. \qquad (\mathbf{OP})$$

ISBN: 978-989-95079-6-8

- The *exchange principle*,

$$I(x, I(y, z)) = I(y, I(x, z)), \quad \text{for all} \quad x, y, z \in [0, 1]. \tag{EP}$$

With respect to R-implications derived from left-continuous t-norms (see for instance [9] and the survey [14] with the references therein), they satisfy both properties (OP) and (EP). Moreover, we have the following characterization.

Theorem 1 *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is an R-implication derived from a left-continuous t-norm, if and only if, I satisfies $(I2), (OP), (EP)$ and I is right-continuous with respect to the second variable.*

With the assumption of continuity we have a characterization of the following subclass of R-implications, that are also (S, N)-implications, known as the Smets-Magrez Theorem, see [21] and see also [12] for the current version.

Theorem 2 *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is a continuous function satisfying $(OP), (EP)$, if and only if, I is conjugate with the Łukasiewicz implication, that is, there exists an increasing bijection $\varphi : [0,1] \to [0,1]$ such that*

$$I(x, y) = \varphi^{-1}(\min(1, 1 - \varphi(x) + \varphi(y)))$$

for all $x, y \in [0,1]$.

We also suppose that some basic facts on uninorms are known as well as the class of representable ones (see for instance [15]). Let us only recall here the definition.

Definition 2 *A* uninorm *is a two-place function $U : [0,1]^2 \longrightarrow [0,1]$ which is associative, commutative, increasing in each place and such that there exists some element $e \in [0,1]$, called* neutral element*, such that $U(e, x) = x$ for all $x \in [0,1]$.*

It is clear that uninorms generalize both t-norms and t-conorms, since they are retrieved from uninorms just taking $e = 1$ and $e = 0$, respectively. Moreover, for a uninorm U, it is always $U(1,0) \in \{0,1\}$ and U is said to be *conjunctive* when $U(1,0) = 0$, and *disjunctive* when $U(1,0) = 1$.

On the other hand, residual implications derived from uninorms have been also studied (see [16]).

Definition 3 *Let U be a uninorm. The* residual operation *derived from U is the binary operation given by*

$$I_U(x, y) = \sup\{z \in [0,1] \mid U(x, z) \leq y\} \tag{2}$$

for all $x, y \in [0,1]$.

Proposition 1 *([16]) Let U be a uninorm and I_U its residual operation. Then I_U is an R-implication if and only if the following condition holds*

$$U(x, 0) = 0 \quad \textit{for all} \quad x < 1. \tag{3}$$

This includes all conjunctive uninorms but also many disjunctive ones, for instance in the classes of representable and idempotent uninorms (see [16] and [17]). However, when we deal with left-continuous uninorms U we clearly have that U satisfies condition (3) if and only if it is conjunctive.

3 R-implications derived from left-continuous uninorms

There are some properties of R-implications derived from uninorms, that can be deduced directly from the definition, or can be proved in the same way as for those derived from t-norms (see [16]).

Proposition 2 *Let U be a uninorm with neutral element e satisfying condition(3) and I_U its residual implication. Then*

i) *$I_U(e, y) = y$ for all $y \in [0,1]$ counterpart for uninorms of the* neutrality principle*, that will be denoted by* $(\mathbf{NP_U})$.

ii) *$I_U(x, y) \geq e$ for all $x, y \in [0,1]$ such that $x \leq y$.*

iii) *$y \leq I_U(x, U(x, y))$ for all $x, y \in [0,1]$.*

In this section we will deal specially with R-implications derived from left-continuous uninorms. Recall that in this case condition (3) is equivalent to the uninorm U be conjunctive and so, we will refer only to left-continuous conjunctive uninorms. As for t-norms, this case of left-continuity is specially important because then (and only then) I_U satisfies the residuation property. The proof of this fact is similar to the case of t-norms (see [2]) and thus we do not include it.

Proposition 3 *Let U be a conjunctive uninorm and I_U its residual implication. Then U is left-continuous if and only if I_U satisfies the residuation property:*

$$U(x, y) \leq z \iff I_U(x, z) \geq y \tag{RP}$$

for all $x, y, z \in [0,1]$.

Moreover, when U is left-continuous additional properties are satisfied.

Proposition 4 *Let U be a left-continuous conjunctive uninorm with neutral element $e \in (0,1]$. Then I_U satisfies*

i) *Counterpart for uninorms of the ordering property:*

$$x \leq y \iff I_U(x, y) \geq e \quad \textit{for all } x, y \in [0,1]. \tag{OP$_\mathbf{U}$}$$

ii) *Exchange principle,*

$$I_U(x, I_U(y, z)) = I_U(y, I_U(x, z)) \tag{EP}$$

for all $x, y, z \in [0,1]$.

iii) *The* modus ponens property*:*

$$U(x, I_U(x, y)) \leq y \tag{MP}$$

for all $x, y \in [0,1]$.

iv) *$I_U(x, -)$ is right-continuous for all $x \in [0,1]$.*

Proof: Part $i)$ is straightforward from the definition, part $iv)$ follows as for the case of t-norms and the other properties were already proved in [16]. ∎

ISBN: 978-989-95079-6-8

In order to characterize all functions $I : [0,1]^2 \to [0,1]$ that are R-implications derived from left-continuous uninorms, we firstly study those functions with properties (OP_U) and (EP).

Proposition 5 *Let $I : [0,1]^2 \to [0,1]$ be a function satisfying properties (OP_U) and (EP). Then*

i) $I(x,x) \geq e$ for all $x \in [0,1]$.

ii) I satisfies property (I1): $I(x,z) \geq I(y,z)$ when $x \leq y$, for all $z \in [0,1]$.

iii) I satisfies (NP_U). In particular, $I(e,e) = e$.

iv) $I(0,y) = 1$ for all $y \in [0,1]$.

v) If $I(1,e) = 0$ then $N(x) = I(x,e)$ is a fuzzy negation with $N(e) = e$.

Proof:

i) It is obvious from condition (OP_U).

ii) Let us consider $x, y, z \in [0,1]$ with $x \leq y$. Then we have

$$I(y, I(I(y,z),z)) = I(I(y,z), I(y,z)) \geq e$$

$$\Longrightarrow y \leq I(I(y,z),z)$$

by i) and conditions (OP_U) and (EP). Then,

$$x \leq y \leq I(I(y,z),z)$$

and this implies similarly that

$$e \leq I(x, I(I(y,z),z) = I(I(y,z), I(x,z))$$

and thus $I(y,z) \leq I(x,z)$.

iii) First of all,

$$I(x, I(e,x)) = I(e, I(x,x)) \geq e$$

where the last inequality holds because $e \leq I(x,x)$ by i). But the equation above ensures that $x \leq I(e,x)$. To prove the other inequality, it is sufficient to apply i) and conditions (OP_U), (EP) to obtain

$$I(e, I(I(e,x),x)) = I(I(e,x), I(e,x)) \geq e$$

$$\Longrightarrow I(I(e,x),x) \geq e$$

and thus $I(e,x) \leq x$.

iv) We have for all $y \in [0,1]$

$$\begin{array}{rcl} 0 \leq I(1,y) & \Longrightarrow & I(0, I(1,y)) \geq e \\ & \Longrightarrow & I(1, I(0,y)) \geq e \\ & \Longrightarrow & 1 \leq I(0,y) \end{array}$$

and thus $I(0,y) = 1 \; \forall y \in [0,1]$.

v) From ii) we have that N is decreasing and iv) implies that $N(0) = I(0,e) = 1$. Thus, since $N(1) = I(1,e) = 0$, N is a fuzzy negation. Finally, by iii) $N(e) = e$. ∎

Now, we can already give the axiomatic characterization of residual implications derived from left-continuous uninorms. Recall that in this case, since the uninorm must satisfy $U(x,0) = 0$ for all $x < 1$, it must be conjunctive.

Theorem 3 *Let $I : [0,1]^2 \to [0,1]$ be a function. The following statements are equivalent:*

i) *I is an R-implication derived from a left-continuous uninorm U with neutral element $e \in (0,1]$.*

ii) *I satisfies $I2)$, (OP_U), (EP) and $I(x,-)$ is right-continuous for all $x \in [0,1]$.*

Moreover, in this case the uninorm U must be conjunctive and it is given by:

$$U(x,y) = \inf\{z \in [0,1] \mid I(x,z) \geq y\}.$$

Proof: Suppose first that

$$I(x,y) = I_U(x,y) = \sup\{z \in [0,1] \mid U(x,z) \leq y\}$$

for all $x, y \in [0,1]$ where U is a left-continuous conjunctive uninorm with neutral element $e \in (0,1]$. Then we already know that I_U is an implication and consequently satisfies $(I2)$. The other properties follow from Proposition 4 and Proposition 2-iv).

Suppose now that I satisfies the required conditions. We have to prove that the function $U : [0,1]^2 \to [0,1]$ defined by

$$U(x,y) = \inf\{z \in [0,1] \mid I(x,z) \geq y\}$$

for all $x, y \in [0,1]$, is a left-continuous conjunctive uninorm U with neutral element e and also that $I = I_U$.

- $U(x,e) = x \; \forall x \in [0,1]$:
  We have by condition (OP_U) that

$$\begin{array}{rcl} U(x,e) & = & \inf\{z \in [0,1] \mid I(x,z) \geq e\} \\ & = & \inf\{z \in [0,1] \mid z \geq x\} = x. \end{array}$$

- $U(x,y) = U(y,x) \; \forall x, y \in [0,1]$:
  Observe that

$$\begin{array}{rcl} I(y,z) \geq x & \Longleftrightarrow & I(x, I(y,z)) \geq e \\ & \Longleftrightarrow & I(y, I(x,z)) \geq e \\ & \Longleftrightarrow & I(x,z) \geq y \end{array}$$

  by conditions (OP_U) and (EP). Thus

$$\begin{array}{rcl} U(y,x) & = & \inf\{z \in [0,1] \mid I(y,z) \geq x\} \\ & = & \inf\{z \in [0,1] \mid I(x,z) \geq y\} \\ & = & U(x,y) \end{array}$$

- $x_1 \leq x_2 \Longrightarrow U(x_1,y) \leq U(x_2,y)$:
  Let $x_1 \leq x_2$. Now if $I(x_2,z) \geq y$, then $I(x_1,z) \geq I(x_2,z) \geq y$ by property ii) of Proposition 5, and thus $U(x_1,y) \leq U(x_2,y)$.

- $y_1 \leq y_2 \Longrightarrow U(x,y_1) \leq U(x,y_2)$:
  It follows immediately by commutativity and the previous step.

ISBN: 978-989-95079-6-8

- $U(U(x,y),z) = U(x,U(y,z))\ \forall x,y,z \in [0,1]$:

  We have that $U(U(x,y),z) = U(z,U(x,y))$ by commutativity, and thus $U(U(x,y),z)$ is given by

  $$U(U(x,y),z) = \inf\{t \in [0,1] \mid I(z,t) \geq U(x,y)\}.$$

  On the other hand we have that

  $$U(x,U(y,z)) = \inf\{t \in [0,1] \mid I(x,t) \geq U(y,z)\},$$

  and consequently it is sufficient to prove that for all $t \in [0,1]$,

  $$I(z,t) \geq U(x,y) \Longleftrightarrow I(x,t) \geq U(y,z).$$

  Let us prove the left-to-right implication. Since $I(x,-)$ is right-continuous, we have from the expression of $U(x,y)$ that $I(x,U(x,y)) \geq y$. Then

  $$I(z,t) \geq U(x,y) \Longrightarrow I(x,I(z,t)) \geq I(x,U(x,y)) \geq y.$$

  Now, the increasingness with respect to the second component gives that:

  $$\begin{array}{rcl} U(y,z) = U(z,y) & \leq & U(z,I(x,I(z,t))) \\ & = & U(z,I(z,I(x,t))) \\ & \leq & I(x,t) \end{array}$$

  where the last inequality follows from the definition of U. The right-to-left implication follows similarly.

- U is conjunctive and left-continuous:

  Obviously since

  $$U(1,0) = \inf\{z \in [0,1] \mid I(1,z) \geq 0\} = 0,$$

  and left-continuity also follows easily from the definition of U.

- $I(x,y) = I_U(x,y) = \sup\{z \in [0,1] \mid U(x,z) \leq y\}$:

  First of all, observe that the residuation property for I_U allows to write

  $I_U(z, I_U(x,y)) =$

  $$\begin{array}{rcl} & = & \sup\{t \in [0,1] \mid U(z,t) \leq I_U(x,y)\} \\ & = & \sup\{t \in [0,1] \mid U(x,U(z,t)) \leq y\} \\ & = & \sup\{t \in [0,1] \mid U(U(x,z),t) \leq y\} \\ & = & I_U(U(x,z),y) \end{array}$$

  Now, for all $z \in [0,1]$ we have

  $$U(x,z) \leq y \Longleftrightarrow I_U(U(x,z),y) = I_U(z,I_U(x,y)) \geq e$$

  which implies $I_U(x,y) \geq z$. Thus, taking $z = I(x,y)$ we know, by definition of U, that $U(x,I(x,y)) \leq y$ and consequently we will have $I_U(x,y) \geq I(x,y)$.

  On the other hand, right-continuity of $I(x,-)$ implies that $I(x,U(x,z)) \geq z$. Now if we take $z = I_U(x,y)$,

  $$I_U(x,y) \leq I(x,U(x,I_U(x,y))) \leq I(x,y)$$

  again by the residuation property of I_U and the increasingness of I in the second variable. ∎

Remark 1 *Note that when $e = 1$, condition (OP_U) becomes the ordering property (OP) and so, the axiomatic characterization of R-implications derived from left-continuous t-norms (Theorem 1) follows as a particular case.*

The mutual-independence of the properties in the theorem above is an open problem as it already is for the case of t-norms in Theorem 1 (see for instance [7], Remark 2.5.18). However, there are some of these properties that are independent to each other. The same examples for the case of t-norms work here as follows.

Example 1 *i) The Rescher implication given by*

$$I(x,y) = \begin{cases} 1 & if\ x \leq y \\ 0 & if\ x > y \end{cases}$$

satisfies (I2), (OP_U) with respect to any $e \in (0,1]$, and $I(x,-)$ is right-continuous for all $x \in [0,1]$, but it does not satisfy (EP).

ii) The Kleene-Dienes implication given by $I(x,y) = \max(1-x,y)$ satisfies (I2), (EP) and $I(x,-)$ is right-continuous for all $x \in [0,1]$, but there is no $e \in (0,1)$ for which I satisfies (OP_U).

On the other hand, the previous theorem gives the axiomatic characterization of R-implications derived from left-continuous conjunctive uninorms in general. Some particular characterizations can also be given for special classes of left-continuous conjunctive uninorms. Let us study the case of representable uninorms in the next section.

3.1 Residual implications derived from representable uninorms

Recall that a uninorm U with neutral element $e \in (0,1)$ is said to be representable (see [15]) if there exists a strictly increasing and continuous function $h : [0,1] \to [-\infty,+\infty]$ with $h(0) = -\infty$, $h(e) = 0$ and $h(1) = +\infty$ such that U is given by

$$U(x,y) = \begin{cases} h^{-1}(h(x)+h(y)) & \text{if } (x,y) \notin \{(0,1),(1,0)\} \\ 0 \text{ (or } 1) & \text{otherwise.} \end{cases}$$

Function h is called an additive generator of U. Thus it is clear that there are two different representable uninorms with the same generator h, a conjunctive one (which is then left-continuous) and a disjunctive one (which is right-continuous). Representable uninorms are clearly continuous in $[0,1]^2 \setminus \{(1,0),(0,1)\}$ and this is in fact a characterization of such class of uninorms.

Proposition 6 *([22]) A uninorm U with neutral element $e \in (0,1)$ is representable if and only if it is continuous in $[0,1]^2 \setminus \{(1,0),(0,1)\}$*

Note that in both cases, conjunctive and disjunctive, it is $U(0,y) = 0$ for all $y < 1$ and so the residual implication from U can be derived. However, when we consider left-continuous representable uninorms they will be necessarily conjunctive as in the general case.

ISBN: 978-989-95079-6-8

Proposition 7 *([16]) Let U be a representable uninorm with additive generator h. Then its residual implication I_U is given by*

$$I_U(x,y) = \begin{cases} h^{-1}(h(y)-h(x)) & \text{if } (x,y) \notin \{(0,0),(1,1)\} \\ 1 & \text{otherwise.} \end{cases}$$

Let us now prove the characterization theorem for residual implications derived from this kind of uninorms.

Theorem 4 *Let $I : [0,1]^2 \to [0,1]$ be a function. The following statements are equivalent:*

i) *I is an R-implication derived from a left-continuous representable uninorm with neutral element $e \in (0,1)$.*

ii) *I satisfies (OP_U), (EP), $I(1,e) = 0$ and I is continuous except at points $(0,0)$ and $(1,1)$.*

Proof: Suppose first that I is an R-implication derived from a left-continuous representable uninorm with neutral element $e \in (0,1)$. In this case U is conjunctive and by Theorem 3, I satisfies (OP_U) and (EP). Moreover, since U is representable, I must be continuous except at points $(0,0)$ and $(1,1)$ from Proposition 7.

Now suppose that the required conditions for I are satisfied. First consider $N(x) = I(x,e)$ and let us prove that N is a strong negation with fixed point e. Since $I(1,e) = 0$, we already know from Proposition 5 that N is a fuzzy negation with $N(e) = e$. Moreover, since $e \neq 0, 1$, $N(x) = I(x,e)$ is continuous.

On the other hand, again from Proposition 5 we have for all $x \in [0,1]$

$$I(x, I(I(x,e),e)) = I(I(x,e), I(x,e)) \geq e$$

and consequently

$$x \leq I(I(x,e),e) = N(N(x)) = N^2(x) \quad \text{for all} \quad x \in [0,1].$$

Thus, by decreasingness of N we have $N(x) \geq N(N(N(x))) = N^3(x)$. On the other hand,

$$I(I(x,e), I(I(I(x,e),e),e)) =$$

$$= I(I(I(x,e),e), I(I(x,e),e)) \geq e$$

and, thus $N(x) = I(x,e) \leq I(I(I(x,e),e),e) = N^3(x)$. From both inequalities we deduce $N(x) = N^3(x)$ for all $x \in [0,1]$. Now, since $N(x) = I(x,e)$ is continuous in [0,1] by hypothesis, given $x \in [0,1]$, there exists $y \in [0,1]$ such that $N(y) = x$. Then $N(N(x)) = N(N(N(y))) = N(y) = x$, that is, N is involutive and so is a strong negation.

Now, define

$$U(x,y) = N(I(x,N(y))) \quad \text{for all} \quad x,y \in [0,1]$$

and let us prove that U is a uninorm with neutral element e.

- Commutativity comes from condition (EP),

$$\begin{aligned} U(y,x) &= N(I(y,N(x))) \\ &= I(I(y,N(x)),e) \\ &= I(I(y,I(x,e)),e) \\ &= I(I(x,I(y,e)),e) \\ &= N(I(x,N(y))) \\ &= U(x,y) \end{aligned}$$

- From condition (EP) we have

$$\begin{aligned} I(N(y),N(x)) &= I(I(y,e),I(x,e)) \\ &= I(x,I(I(y,e),e)) = I(x,y) \end{aligned}$$

since $I(I(y,e),e) = N^2(y) = y$.

Now, associativity can be proved as folows

$$\begin{aligned} U(x,U(y,z)) &= N(I(x,N(U(y,z)))) \\ &= N(I(x,N(N(I(y,N(z)))))) \\ &= N(I(x,I(y,N(z)))) \\ &= N(I(y,I(x,N(z)))) \\ &= N(I(y,I(z,N(x)))) \\ &= N(I(z,I(y,N(x)))) \\ &= N(I(N(I(y,N(x))),N(z))) \\ &= N(I(N(I(x,N(y))),N(z))) \\ &= N(I(U(x,y),N(z))) \\ &= U(U(x,y),z) \end{aligned}$$

- The increasingness of U comes from Proposition 5 and the decreasingness of N. Given $x_1 \leq x_2$ and $y \in [0,1]$, we have

$$I(x_1,N(y)) \geq I(x_2,N(y)).$$

Thus,

$$N(I(x_1,N(y))) \leq N(I(x_2,N(y)))$$

$$\Longrightarrow U(x_1,y) \leq U(x_2,y)$$

- Finally,

$$\begin{aligned} U(x,e) &= N(I(x,N(e)) = N(I(x,e)) \\ &= N^2(x) = x. \end{aligned}$$

Now, since I is continuous except at points $(0,0)$ and $(1,1)$, U will be continuous except at points $(0,1)$ and $(1,0)$. Moreover, since $I(1,1) \geq e$ we have

$$U(1,0) = N(I(1,1)) \leq e$$

and consequently $U(1,0) = 0$. This proves that U is a conjunctive representable uninorm (in particular it is left-continuous).

Finally, to see that $I = I_U$, it is enough to prove the residuation property:

$$U(x,y) \leq z \Longleftrightarrow I(x,z) \geq y$$

since then $I(x,z)$ will be given by the supremum of $y \in [0,1]$ such that $U(x,y) \leq z$. We have

$$\begin{aligned} U(x,y) \leq z &\iff N(I(x,N(y))) \leq z \\ &\iff I(x,N(y)) \geq N(z) \\ &\iff I(N(z),I(x,N(y))) \geq e \\ &\iff I(x,I(N(z),N(y))) \geq e \\ &\iff I(x,I(y,z)) \geq e \\ &\iff I(y,I(x,z)) \geq e \\ &\iff I(x,z) \geq y \end{aligned}$$

■

ISBN: 978-989-95079-6-8

With respect to the mutual independence of the properties in the theorem above, we already know that (OP_U), and (EP) can not be derived from the others (see the counterexamples below). Now, we are currently working in the independence of the other two conditions.

Example 2 *Consider* $e \in (0,1)$ *and let* $h : [0,1] \to [-\infty, +\infty]$ *be a strictly increasing continuous function with* $h(0) = -\infty, h(e) = 0$ *and* $h(1) = +\infty$. *Then*

- *The function*

$$I(x,y) = \begin{cases} e & if\ x \leq y \\ h^{-1}(h(y) - h(x)) & if\ x > y \end{cases}$$

satisfies all conditions in Theorem 4 except (EP).

- *The function*

$$I(x,y) = \begin{cases} 0 & if\ x = y = 0 \\ & or\ x = y = 1 \\ h^{-1}(h(y) - h(x)) & otherwise \end{cases}$$

satisfies all conditions in Theorem 4 except (OP_U).

Corollary 1 *Let* $I : [0,1]^2 \to [0,1]$ *be a function and* $e \in (0,1)$. *Then there exists a strictly increasing continuous function* $h : [0,1] \to [-\infty, +\infty]$ *with* $h(0) = -\infty, h(e) = 0$ *and* $h(1) = +\infty$ *such that*

$$I(x,y) = \begin{cases} h^{-1}(h(y) - h(x)) & \text{if } (x,y) \notin \{(0,0),(1,1)\} \\ 1 & \text{otherwise.} \end{cases}$$

if and only if I *satisfies* (OP_U), (EP), $I(1,e) = 0$ *and it is continuous except at points* $(0,0)$ *and* $(1,1)$.

This corollary can be viewed as a generalization of the well known Smets-Magrez Theorem (Theorem 2) since it characterizes a kind of implications that are continuous except at two points.

Finally, we want to claim that other characterization theorems can be given for residual implications derived from other classes of left-continuous conjunctive uninorms, and we are currently working in this direction. For instance, from left-continuous conjunctive idempotent uninorms or from left-continuous conjunctive uninorms continuous at the open unit square.

Acknowledgment

This paper has been partially supported by the Spanish grants MTM2006-05540 and MTM2006-08322 (with FEDER support), and the Government of the Balearic Islands grant PCTIB-2005GC1-07.

References

[1] E.E. Kerre, C. Huang, and D. Ruan. *Fuzzy Set Theory and Approximate Reasoning*. Wu Han University Press, Wu Chang, 2004.

[2] S. Gottwald. *A Treatise on Many-Valued Logic*. Research Studies Press, Baldock, 2001.

[3] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.

[4] H. Bustince, V. Mohedano, E. Barrenechea, and M. Pagola. Definition and construction of fuzzy DI-subsethood measures. *Information Sciences*, 176:3190–3231, 2006.

[5] H. Bustince, M. Pagola, and E. Barrenechea. Construction of fuzzy indices from fuzzy DI-subsethood measures: application to the global comparison of images. *Information Sciences*, 177:906–929, 2007.

[6] P. Yan and G. Chen. Discovering a cover set of ARsi with hierarchy from quantitative databases. *Information Sciences*, 173:319–336, 2005.

[7] M. Baczyński and B. Jayaram. *Fuzzy Implications*. Springer, Berlin Heidelberg, 2008.

[8] C. Alsina et al. On the characterization of S and R implications. *Proceedings of* VI^{th} *IFSA World Congress*, pages 317–319, 1995.

[9] J.C. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, Dordrecht, 1994.

[10] J. C. Fodor. Contrapositive symmetry on fuzzy implications. *Fuzzy Sets and Systems*, 69:141–156, 1995.

[11] H. Bustince, P. Burillo, and F. Soria. Automorphisms, negation and implication operators. *Fuzzy Sets and Systems*, 134:209–229, 2003.

[12] M. Baczyński. Residual implications revisited. notes on the smets-magrez theorem. *Fuzzy Sets and Systems*, 145:267–277, 2004.

[13] M. Baczyński and B. Jayaram. On the characterization of (S,N)-implications. *Fuzzy Sets and Systems*, 158:1713–1727, 2007.

[14] M. Baczyński and B. Jayaram. (S,N)- and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems*, 159:1836–1859, 2008.

[15] J. C. Fodor, R. R. Yager, and A. Rybalov. Structure of uninorms. *Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems*, 5:411–127, 1997.

[16] B. De Baets and J. C. Fodor. Residual operators of uninorms. *Soft Computing*, 3:89–100, 1999.

[17] D. Ruiz and J. Torrens. Residual implications and co-implications from idempotent uninorms. *Kybernetika*, 40:21–38, 2004.

[18] D. Ruiz-Aguilera and J. Torrens. S-implications and R-implications from uninorms continuous in $]0,1[^2$. *Fuzzy Sets and Systems, doi: 10.1016/j.fss.2008.05.015.*, 2008.

[19] M. Baczyński and B. Jayaram. (U,N)-implications and their characterizations. *Fuzzy Sets and Systems. doi:10.1016/j.fss.2008.11.001.*, 2008.

[20] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[21] P. Smets and P. Magrez. Implications in fuzzy logic. *International Journal of Approximate Reasoning*, 1:327–347, 1987.

[22] D. Ruiz-Aguilera and J. Torrens. Distributivity and conditional distributivity of a uninorm and a continuous t-conorm. *IEEE Transactions on Fuzzy Systems*, 14(2):180–190, 2006.

ISBN: 978-989-95079-6-8

A logic-based approach for evaluating interpretability of fuzzy rule-based classifiers

Corrado Mencar Ciro Castiello Anna M. Fanelli

Dept. of Informatics, University of Bari
Bari, Italy
Email: {mencar, castiello, fanelli}@di.uniba.it

Abstract— *We describe an automatic approach for evaluating interpretability of fuzzy rule-based classifiers. The approach is based on the logical view of fuzzy rules, which are interpreted as rows in truth tables. These truth tables are subject of a minimization procedure based on a variant of the Quine-McCluskey algorithm. The minimized truth tables are used to build new fuzzy rules, which are compared with the original ones in terms of classification accuracy. If the two sets of rules have similar performances, we deduce that the logical view of rules is applicable to the fuzzy classifier, which is hence considered interpretable. On the other hand, a significant difference in classification ability shows that fuzzy rules cannot be interpreted in logical terms, hence linguistic labelling may not be significant. Two illustrative examples show both the cases.*

Keywords— fuzzy rule-based classifiers, interpretability assessment, logic minimization, Quine McCluskey algorithm

1 Introduction

Interpretability is one of the most important driving forces for the adoption of fuzzy rule-based systems . The reason is clear: several models exist that can perform intelligent tasks (such as prediction, classification, etc.) but fuzzy systems allow for a representation of knowledge that can be easily read and understood by their users. Interpretability, however, is *not* given for granted when fuzzy models are used, especially when they are acquired from data. The main problem is that data-driven design has a great number of degrees of freedom (number of fuzzy sets, their shape, position, etc.) and may end up with fuzzy models that are very accurate but very hard to comprehend. For this reason, interpretability constraints have been defined so as to bind data-driven design in order to derive interpretable fuzzy models [1]. This usually comes to a price that is a lower accuracy with respect to unconstrained design. Furthermore, often interpretability is accounted without taking care of accuracy. This approach has been criticized, since interpretable but inaccurate models are as useless as very accurate but not interpretable ones [2].

Interpretability can be viewed at different levels of a fuzzy model. Interpretability of fuzzy sets involves the possibility of tagging them with linguistic labels. On a higher level, interpretability of partitions concerns the possibility of linguistically describe the domain of data. On the rule-level, which is of our concern in this paper, interpretability is viewed as the ability of explaining relationships between input and output variables in a linguistic form.

An important issue concerns the evaluation of interpretability. Interpretability assessment is important because several fuzzy rule-based models can be derived from the same data. Many of these models could provide for high accuracies but can vary in their interpretability. As a consequence, a tool for interpretability assessment could be helpful in the choice of the final model or, to a greater extent, the design technique. However, interpretability assessment is an ill-posed problem because the definition of interpretability eludes any formal characterization.

In [3], Michalski gives a referential definition of interpretability, the so-called "comprehensibility postulate", reported in the following.

> The results of computer induction should be symbolic descriptions of given entities, semantically and structurally similar to those a human expert might produce observing the same entities. Components of these descriptions should be comprehensible as single "chunks" of information, directly interpretable in natural language, and should relate quantitative and qualitative concepts in an integrated fashion.

This postulate justifies the use of linguistic values in rule-based fuzzy systems, but that is not enough to guarantee interpretability (for a more extensive discussion of the postulate, see [4]). In [5] Zadeh introduces the notion of *co-intension*, a semantic relation between concepts. Roughly speaking, two concepts are co-intensive if they refer to almost the same entities. In fuzzy rule-based systems, rules are defined by composition of linguistic terms, which are related to two different semantics. The first one is defined by the fuzzy model and the second one is implicitly designated by the linguistic label. By merging the notion of co-intension with the comprehensibility postulate, we derive a formulation of interpretability that can be more helpful for designing assessing techniques: *a rule base is interpretable if the two semantics related to each linguistic term are co-intensive*.

On the basis of this definition, we propose an automatic technique for evaluating interpretability. Our approach evaluates interpretability by assessing the co-intension of the semantics of the rule base of a fuzzy model with the intrinsic semantics designated by linguistic labels. The core of the evaluation technique relies on the propositional view of rules and on logical operations. We expect that, for interpretable knowledge bases, logical operations on rules do not change their semantics and, hence, do not affect accuracy. If this is assumption is violated, we deduce a lack of interpretability.

We focus our research on fuzzy rule-based classifiers, described in Section 2. The proposed approach is then described, by first focusing on its rationale and then on the methodology

ISBN: 978-989-95079-6-8

(Sections 3.1 and 3.2, respectively). Two illustrative examples are reported in Section 4, to show how the proposed approach can be useful in detecting lacks of interpretability for two simple knowledge bases. The paper is concluded with some final remarks.

2 Fuzzy rule-based classifiers

We consider a classifier as a system computing a function of type:

$$f : \mathbf{X} \longrightarrow \Lambda, \tag{1}$$

where $\mathbf{X} \subseteq \mathbf{R}^n$ is an n-dimensional input space, and $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_c\}$ is a set of class labels. If a dataset D of pre-classified data is given, i.e.

$$D = \{(\mathbf{x}_i, l_i) | \mathbf{x}_i \in \mathbf{X}, l_i \in \Lambda, i = 1, 2, \ldots, N\}, \tag{2}$$

then the classification error can be computed as:

$$E(f, D) = \frac{1}{N} \sum_{i=1}^{N} (1 - \chi(l_i, f(\mathbf{x}_i))), \tag{3}$$

being $\chi(a, b) = 1$ iff $a = b$ and 0 otherwise.

A fuzzy rule-based classifier (FRBC) is a system that carries out classification (1) through inference on a knowledge base. The knowledge base includes the definition of a linguistic variable for each input. Thus, for each $j = 1, 2, \ldots, n$, linguistic variables[1] are defined as:

$$V_j = (v_j, X_j, Q_j, S_j, I_j), \tag{4}$$

being:

- v_j the name of the variable;
- X_j the domain of the variable (it is assumed that $\mathbf{X} = X_1 \times X_2 \times \cdots \times X_n$);
- $Q_j = \{q_{j1}, q_{j2}, \ldots, q_{jm_j}, \text{ANY}\}$ is a set of labels denoting linguistic values for the variable (e.g. SMALL, MEDIUM, LARGE);
- $S_j = \{s_{j1}, s_{j2}, \ldots, s_{jm_j+1}\}$ is a set of fuzzy sets on X_j, $s_{jk} : X_j \to [0, 1]$;
- I_j associates each linguistic value q_{jk} to a fuzzy set s_{jk}. We will assume that $I_j(q_{jk}) = s_{jk}$.

We assume that each linguistic variable contains the linguistic value "ANY" associated to a special fuzzy set $s \in S_j$ such that $s(x) = 1, \forall x \in X_j$.

The knowledge base of a FRBC is defined by a set of R rules. Each rule can be represented by the schema:

$$\text{IF } v_1 \text{ IS } [\text{NOT}]\ q_1^{(r)} \text{ AND } \cdots \text{ AND } v_n \text{ IS } [\text{NOT}]\ q_n^{(r)} \text{ THEN } \lambda^{(r)} \tag{5}$$

being $q_j^{(r)} \in Q_j$ and $\lambda^{(r)} \in \Lambda$. Symbol NOT is optional for each linguistic value. If for some j, $q_j^{(r)} = \text{ANY}$, then the corresponding atom "v_j IS ANY" can be removed from the representation of the rule.[2]

Inference is carried out as follows. When an input $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is available, the strength of each rule is calculated as:

$$\mu_r(\mathbf{x}) = s_1^{(r)}(x_1) \otimes s_2^{(r)}(x_2) \otimes \cdots \otimes s_n^{(r)}(x_n), \tag{6}$$

being $s_j^{(r)} = \nu_j^{(r)}(I_j(q_j^{(r)}))$, $j = 1, 2, \ldots, n$, $r = 1, 2, \ldots, R$. Function $\nu_j^{(r)}(t)$ is $1 - t$ if NOT occurs before $q_j^{(r)}$, otherwise it is defined as t. The operator $\otimes : [0, 1]^2 \to [0, 1]$ is usually a t-norm, such as minimum or product.

The degree of membership of input $\mathbf{x}$ to class λ_i is computed by considering all the rules of the FRBC as:

$$\mu_{\lambda_i}(\mathbf{x}) = \frac{\sum_{r=1}^{R} \mu_r(\mathbf{x}) \chi(\lambda_i, \lambda^{(r)})}{\sum_{r=1}^{R} \mu_r(\mathbf{x})}. \tag{7}$$

Finally, since just one class label has to be assigned for the input $\mathbf{x}$, the FRBC assigns the class label with highest membership (ties are solved randomly):

$$f_{FRBC}(\mathbf{x}) = \lambda \Rightarrow \mu_\lambda(\mathbf{x}) = \max_{i=1,2,\ldots,c} \mu_{\lambda_i}(\mathbf{x}). \tag{8}$$

3 Interpretability assessment

We assume the availability of an interpretable FRBC, verifying a number of interpretability constraints so that the rule base is described in terms of linguistic values.

Interpretability of FRBC is usually achieved by imposing a number of constraints on the knowledge base. For our purposes, it is necessary that labels denoting linguistic values for each variable can be interpreted as well-distinguished elementary concepts. To achieve this goal, normality, convexity and distinguishability are the basic interpretability constraints to be imposed on the FRBC. Normality is indeed required to represent consistent concepts (i.e. concepts that are fully applicable to at least one object), convexity is necessary to represent elementary concepts (i.e. concepts that cannot be split in two or more sub-concepts) and distinguishability allows the use of different labels to denote different fuzzy sets (for a detailed explanation, the reader is referred to [1]).

3.1 Rationale

The proposed approach for interpretability assessment relies on the formal structure of the FRBC. The rationale behind this approach comes from the observation that the rule base is the linguistic interface of the FRBC to the user. For an interpretable knowledge base, the user should be able to understand the classification rules by simply observing their linguistic representation. All the semantic information (fuzzy sets attached to linguistic values, t-norm used for conjunction, etc.) should be hidden to the user because – this is the key point of interpretability – the semantics of FRBC knowledge should be co-intensive with the user's knowledge, recalled by the linguistic terms.

The proposed approach tries to evaluate how much co-intensive is the semantics of the FRBC knowledge w.r.t. user knowledge. As a further requirement, we want the evaluation

[1] For the sake of clarity, in this paper we use a simplified definition of linguistic variable.

[2] The sequence NOT ANY is not allowed.

ISBN: 978-989-95079-6-8

process to be carried out automatically. To clarify its rationale, we consider the example of two communicating actors, A and B. When A communicates some piece of knowledge to B, he/she uses a linguistic structure that (at least partially) represents his/her knowledge. To be comprehensible, A chooses linguistic terms whose semantics is deemed as co-intensive as possible with B's knowledge. It is not necessary that the semantics of linguistic terms in A's mind perfectly match the semantics in B's mind: a high "overlapping" of semantics is enough for knowledge communication. The actor A can be almost sure of the co-intension with B if they share a similar environment, language, culture, experiences, etc. Thus, co-intension can be achieved if A and B share similar cognitive structures.

To assess interpretability, we exploit the cognitive structures that are shared by users and FRBC. In particular, we observe a strict affinity of a FRBC rule base to logical propositions. Actually, rules are formed so as to resemble propositions, so that they can be understood by users. In consequence of this, FRBC and users share this propositional view of rules. Being like propositions, rules could be transformed by truth-preserving operators without any change of the semantics. This is not completely true since the application of such operators may distort the semantics of rule (which is fuzzy); however, we should expect only small distortions.

The core of our approach is the following: given a rule base of a FRBC, we represent it as a collection of propositions, and then we transform it through a truth-preserving operator, thus obtaining a new set of propositions, that represents a new rule base. We then compare the two rule bases on the basis of their classification ability: if they differ too much, then we conclude that the logical view of rules is wrong. Also, since rules are defined so as to resemble logical propositions, we derive that the semantics of rules (which is responsible of classification) is such that logical view is not possible. This means that the semantics of rules is not co-intensive with user knowledge, since users expect that truth-preserving transformations of propositions do not change (too much) their semantics.

A simple example may further clarify the rationale of our approach. Suppose to have a trivial FRBC with the following two rules:

```
IF cell_size is Large AND
   cell_shape is Irregular THEN malign
IF cell_size is Small THEN benign
```

Any user reading this rule base understands that the size of a cell is enough to discriminate between malign and benign classes. As a consequence, a truth-preserving transformation of the rule base could lead to the following set of rules:

```
IF cell_size is Large THEN malign
IF cell_size is Not Large THEN benign
```

This understanding is legitimate under the Closed World Assumption (CWA), according to which any input satisfies the condition of at least one rule. CWA is generally valid in most fuzzy systems and FRBC in particular, hence it is taken for grant in this work. Open World Assumption (OWA), which leads to non-monotonic reasoning, is out of the scope of this paper.

The two rule bases can be compared in terms of classification errors. If the two classification errors are very similar, we derive that the semantics of the rule base is compatible with the applied transformation and, hence, to the logical view of rules. In this sense, we state that the knowledge base of the FRBC is co-intensive with user knowledge. On the other hand, if the classification errors differ too much, then the semantics of the knowledge base cannot be represented with a set of propositions, hence, it is not co-intensive with user mind. In other words, it is not interpretable.

We cannot expect that classification errors are always identical, even for very interpretable FRBC, because the law of excluded middle ($A\bar{A} = 0$) is not valid in fuzzy logic but it holds in Boolean logic. On the other hand, we should expect that the violation of excluded middle does not influence too much the inference process, otherwise we should conclude that the FRBC is heavily based on inconsistent knowledge. This latter situation clashes with the property of interpretability.

An issue arises in the choice of the truth-preserving transformation. Actually, several transformations can be considered, but we choose to apply a transformation that minimizes the number of linguistic terms used. This choice has a twofold advantage. First, by eliminating as many terms as possible, we test if logical view of rules is preserved with the minimum required information. Second, if assessment leads to positive results, we could retain the simplified rule base because it is easier to read than the original.

3.2 Methodology

The proposed approach for interpretability assessment is based on a four-stage strategy. It is similar to the approach proposed in [8], however our strategy is able to deal with negative information to achieve more compact representations of the rule bases.

3.2.1 Definition of truth tables

Each rule of the FRBC is seen as a proposition, i.e. a combination of propositional variables that is considered true for a class.

For each class label $\lambda_i \in \Lambda$ and for each rule r, a truth function π_i is defined on the propositional variables defined for the FRBC as:

$$\pi_i(\chi_{11}^{(r)}, \ldots, \chi_{1m_1}^{(r)}, \chi_{n1}^{(r)}, \ldots, \chi_{nm_n}^{(r)}) = \chi(\lambda_i, \lambda^{(r)}), \quad (9)$$

being $\chi_{jk}^{(r)} = \chi(q_j^{(r)}, q_{jk})$. Inputs $\chi_{jk}^{(r)}$ assume value DC ("don't care") if the corresponding linguistic value is ANY or, in other words, the linguistic variable V_j does not occur in the r-th rule. For any other combination of inputs, the output of π_i is undefined, i.e. any truth value is possible (again, this condition is usually referred as "don't care").

Each truth function π_i can be represented as a truth table, which enumerates any combination of assignments to the propositional variables of the FRBC and associates the value of π_i to each combination. Combinations associated to undefined values of π_i are not included in the table. The number of rows of each truth table matches the number of rules of the FRBC. This prevents the combinatorial explosion of rows that

ISBN: 978-989-95079-6-8

Small	Large	Regular	Irregular	Malign	Benign
0	1	0	1	1	0
1	0	DC	DC	0	1

Table 1: The truth tables of the simple rulebase

Small	Large	Regular	Irregular	Malign	Benign
DC	1	DC	DC	1	0
DC	0	DC	DC	0	1

Table 2: The minimized truth tables of the simple rulebase

would be expected in the general case of truth function representation. Table 1 shows the truth tables for classes "malign" and "benign" of the previous example.

3.2.2 Minimization

Once each truth table has been built, it can be processed so as to be minimized. The minimization procedure produces a new truth table without modifying the truth function (where it is defined). The new truth table has a number of rows not greater than the original truth table. It also has a number of DC values in its inputs not smaller than in the original truth table. Furthermore, minimization guarantees that any further simplification (in terms of rows and inputs) provides for a truth function different from the original.

The Quine-McCluskey (QMC) algorithm represents an effective mechanism for minimization of truth tables [7]. It is mainly based on the distributive property, which simplifies propositions according to the law: $ABC + A\bar{B}C \equiv AC$.

The QMC algorithm works in two stages:

1. Merge rows with output 1 or DC that differ in only one input;
2. Find the minimum number of merged rows that cover all rows of the original truth table.

Although computationally expensive, the QMC algorithm can be implemented by an efficient procedure that exploits the peculiar structure of truth tables derived from FRBC rules to perform minimization quickly. A specific implementation of the first stage of QMC avoids the generation of all input combinations, thus saving time for merging. This can be achieved because the number of rows of the truth tables representing rule bases is often very small. Also, the second stage can be optimized by using heuristic procedures that drive the minimization process without expensive search. Table 2 shows the minimized truth table for the example FRBC.

3.2.3 Reconstruction

After minimization, a new FRBC is built from the rows of the minimized truth table. For each class label $\lambda_i \in \Lambda$ we consider the minimized table associated to the truth function π_i. A rule is built for each row with output equal to 1. It is easy to show that for each j there is at most one k such that $\chi_{jk}^{(r)} \neq DC$. Therefore, the antecedent of the rule can be defined by atoms v_j IS [NOT] $q_j^{(r)}$ where:

- $q_j^{(r)} = q_{jk}$ if $\chi_{jk}^{(r)} \neq DC$;
- NOT occurs if $\chi_{jk}^{(r)} = 0$ and it does not occur if $\chi_{jk}^{(r)} = 1$;
- $q_j^{(r)} =$ ANY if $\forall k : \chi_{jk}^{(r)} = DC$;

Atoms with ANY are removed to improve the readability of the rule. The consequent of the rule is λ_i.

At the end of reconstruction, each rule is formed by a sequence of atoms tied together by conjunction (AND). In this paper we do not deal with more complex representations (such as conjunctions of disjunctions of atoms) that could provide for more compact representations of rules but deserve further investigation that will be object of future research.

3.2.4 Comparison

The two FRBCs, the first with the original rule base and the second with the minimized version, are compared in terms of classification error on the same data.

If the two errors differ too much, we conclude that the original FRBC lacks of interpretability. Its accuracy is mainly due to the semantic definition of linguistic values, which do not correspond to the propositional view of rules. The FRBC can be used for classification as a "grey box", but its labelling is arbitrary and not co-intensive with user knowledge. Attaching natural language terms to such FRBC is useless and potentially misleading.

If the two errors are very similar, we conclude that the original FRBC is interpretable. The semantic definition of linguistic values is coherent with the logic operators used in minimization. In this sense, the semantic of linguistic values is co-intensive with user knowledge. We could retain the simplified FRBC because its interpretability is greater than the original (due to its higher simplicity) while its accuracy is almost the same of the original.

There is no threshold to decide if a FRBC is interpretable or not, but rather a continuous spectrum of possibilities. Interpretability – as expected – is a matter of degree, and the degree of interpretability is, in our approach, inversely proportional to the difference of accuracies. Even with small variations of accuracy, some considerations can be drawn on the interpretability of the FRBC, as shown in the next Section.

4 Illustrative examples

In this Section we evaluate two different FRBC derived from the same data set with two different methods. The first method is oriented toward interpretability while the second is more inclined to the design of accurate models, albeit taking into account some kind of readability of the knowledge base.

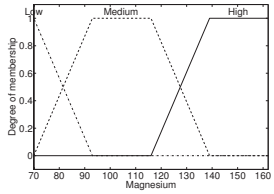
4.1 Interpretable FRBC

We consider a FRBC obtained from the application of HILK [6], a tool for building interpretable fuzzy rule-based systems. HILK allows for the definition of fuzzy rule-based systems from empirical learning, expert knowledge, or both. The resulting systems are considered highly interpretable because fuzzy sets defined for each input satisfy a number of interpretability constraints. Furthermore, the number of fuzzy sets per input, the number of inputs and the number of rules are kept as small as possible since, in principle, the simplest is the knowledge representation, the easier is to read and understand.

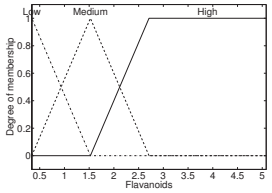
ISBN: 978-989-95079-6-8

If Flavonoids is Low THEN Class 1
If Flavonoids is Medium THEN Class 2
If Flavonoids is High AND Proline is Low THEN Class 2
If Flavonoids is High AND Proline is High THEN Class 3
If Magnesium is Medium AND Flavonoids is High AND Proline is Medium THEN Class 3

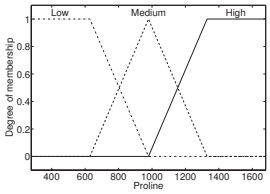
Figure 1: The rule base of the FRBC considered in the first example



(a) Magnesium



(b) Flavonoids



(c) Proline

Figure 2: The linguistic variables used in the first example

In our experimentation, the FRBC obtained from HILK was acquired from data, in order to classify Wine data, a well-known benchmark dataset, freely available from UCI repository [10]. The knowledge base of the FRBC is reported in fig. 1, while the linguistic variables are shown in fig. 2. The FRBC provided 10.67% of classification error on the entire dataset.

We transformed the rule base of the FRBC into three truth tables – one for each class – minimizing them with QMC algorithm; then we rebuilt the FRBC obtaining the simplified rule base reported in fig. 3.

We observe that the number of rules has been reduced to four, and the linguistic variable "Magnesium" has been removed. From the logical viewpoint, the two rule bases are equivalent since they compute the same truth functions. However, after applying the minimized rule base to the dataset, we obtained 11.24% of classification error, i.e. an increase of $+0.57\%$, corresponding to one more misclassified pattern over 178, which is defined as (Magnesium=162, Flavonoids=2.27, Proline=937). For the misclassified pattern we observe that, according to the definitions of the linguistic variables, Magnesium is high, Proline is medium, and Flavonoids is both medium and high to a significant degree. However, in the original rule base, there is not any classification rule activated when Magnesium is high, Proline is medium and Flavonoids is high. From the logical viewpoint, the truth functions are undefined for this condition. As a consequence, the simplified rule base subsumes this condition in the fourth rule, assigning the pattern to class 3, while the original FRBC arbitrarily assigns the pattern to class 2. The correct classification of the pattern performed by the original FRBC

If Flavonoids is Low THEN Class 1
If Flavonoids is Medium THEN Class 2
If Flavonoids is High AND Proline is Low THEN Class 2
If Flavonoids is High AND Proline is NOT Low THEN Class 3

Figure 3: The rule base of the FRBC in the first example after minimization

If Alc. is High And Fla. is High And A.Ash is Low And OD2 is High And Pro. is High Then class 1
If Alc. is Low And Pro. is Low Then class 2
If Alc is Medium And A.Ash is High And Fla. is Low And OD2 is Low And Pro. is Low Then class 3

Figure 4: The original rule base of 2nd example (Alc. = Alcohol, Fla. = Flavonoids, A. Ash = Alcalinity of Ash, OD2 = OD280/315, Pro. = Proline)

is due to the semantic specification of linguistic values, which does not emerge from the propositional view of rules. In this sense, the original FRBC lacks of interpretability for a specific case. However, due to the reduced increase of classification error, we conclude that the original FRBC is highly interpretable and we can retain the simplified version, which offers a further comprehensible knowledge base.

4.2 Non-interpretable FRBC

As a second example, we consider the FRBC obtained through the approach proposed in [9] aimed at optimizing accuracy by taking into account interpretability. In brief, the approach consists of an iterative procedure for developing the FRBC. The initial model is derived from the data and, subsequently, feature selection and rule-base simplification are applied to reduce the model, while a genetic algorithm is used for parameter optimization.

This approach has been applied to Wine data showing very good classification accuracy (1.1% of classification error[3]). One of the best rule bases derived from data is reported in fig. 4. We observe from fig. 5 that in some cases interpretability constraints are violated. As an example, we observe in fig. 5(b) that the rightmost fuzzy set is highly overlapping with the other fuzzy sets of the variable. Notwithstanding, the three fuzzy sets have different labels, thus recalling distinct concepts.

After simplification, we obtain the very small rule base reported in fig. 6. The reconstructed FRBC shows an unacceptable classification error of about 32%. In consequence of this considerable decay of classification performance, we deduce that classification accuracy of this FRBC is mainly based on the parameters of the fuzzy sets that cannot be expressed in linguistic terms. In consequence of this, we state that the logical view of rules is not applicable to the original rule base. In this sense, we conclude that the FRBC is not interpretable.

5 Conclusion

Assessment of interpretability is not an easy task. Difficulty is mainly due to the blurry definition of interpretability, which requires co-intension with human knowledge. Evaluating interpretability is a subjective task, which could be tedious and even ill-posed, therefore automatic techniques for

[3] We use the results reported in the original paper

ISBN: 978-989-95079-6-8

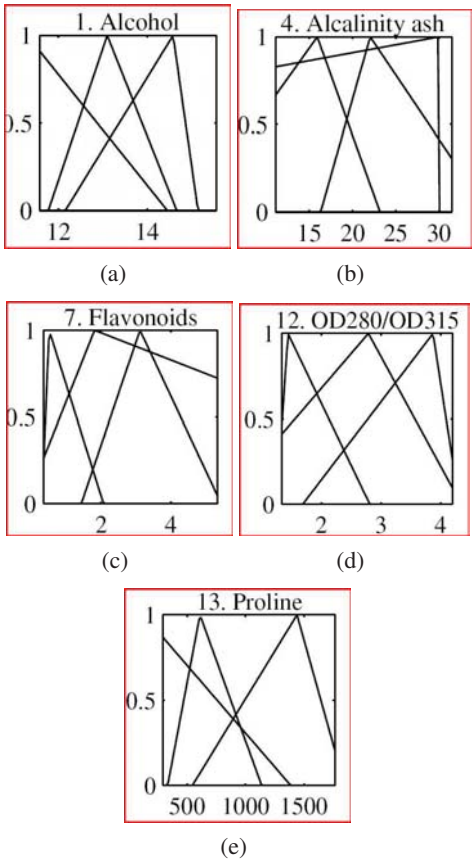


Figure 5: The fuzzy sets used in the second example (for each variable, the leftmost fuzzy set is labelled “Low”, the middle one “Medium” and the rightmost “High”)

If Alcohol is High Then class 1
If Alcohol is Low Then class 2
If Alcohol is Medium Then class 3

Figure 6: The minimized rule base of 2nd example

interpretability assessment are useful, but they should embody some information on semantic co-intension. In this paper, we have described an approach for automatically evaluating interpretability of rule-based fuzzy classifiers, exploiting the propositional view of rules as a mean to define co-intension. On the basis of this hypothesis, a procedure has been devised, so as to evaluate how much the semantics of fuzzy rules is coherent with their logical view.

The illustrative examples show the effects of the proposed approach for a simple knowledge base. Actually, we were able positively assess the interpretability for a FRBC that verifies several interpretability constraints. Also, we have verified that a FRBC is not interpretable even though it is described in linguistic terms. By analyzing the relative performances of the two FRBCs we immediately observe that interpretability preservation comes to the cost of accuracy reduction and, on the other hand, an accurate description of data (preserving interpretability constraints) would require a great number of rules. This, however, would clash with the Comprehensibility Postulate.

Research on this topic is in progress, especially in the direction of enhancing the efficiency of the minimization algorithm. That will allow an extensive experimentation over a wider class of knowledge bases, in order to promote deeper insights of the proposed technique. Further enhancements could spring from the use of additional information to refine the definition of co-intension in computational terms.

References

[1] C. Mencar. *Theory of Fuzzy Information Granulation: Contributions to Interpretability Issues.* PhD thesis, University of Bari, Italy, 2005

[2] A. Riid, E. Rustern. Interpretability of Fuzzy Systems and Its Application to Process Control. *In: Proc. of IEEE International Fuzzy Systems Conference FUZZ-IEEE 2007*, pp. 1–6, 2007

[3] R.S. Michalski. A theory and methodology of inductive learning. *Technical Report UIUCDCS-R-83-1122*, Department of Computer Science, University of Illinois, Urbana, 1983

[4] C. Mencar, G. Castellano, A. M. Fanelli. On the role of interpretability in fuzzy data mining. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS)* 15(5):521 – 537, 2007

[5] L.A. Zadeh. Is there a need for fuzzy logic? *Information Sciences* 178:2751–2779, 2008

[6] J.M. Alonso, L. Magdalena, S. Guillaume. HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems* 23:761–794, 2008

[7] E.J. McCluskey. Minimization of Boolean functions. *Bell Labs Technical Journal* 35(5):1417–1444,1956

[8] R. Rovatti, R. Guerrieri, G. Baccarani. An enhanced two-level Boolean synthesis methodology for fuzzy rules minimization, *Fuzzy Systems, IEEE Transactions on* 3(3):288–299, 1995

[9] J.A. Roubos, M. Setnes, J. Abonyi. Learning fuzzy classification rules from labeled data. *Information Sciences* 150:77–93,2003

[10] *UCI Machine Learning Repository Content Summary:* http://www.ics.uci.edu/ ˜mlearn/MLSummary.html

ISBN: 978-989-95079-6-8

Cloud based design optimization

Martin Fuchs[12]

1. University of Vienna, Faculty of Mathematics, Vienna, Austria
2. CERFACS, Toulouse, France
Email: martin.fuchs81@gmail.com
www.martin-fuchs.net

Abstract— *Cloud based design optimization (CBDO) is an approach to significantly improve robustness and optimality of solutions sought in engineering design. One of the main features is the possibility to capture and model high-dimensional uncertainty information, even in the case that the information available is incomplete or unformalized.*

Continuing our past studies we present the graphical user interface for CBDO in this paper. Also we mention the latest improvements of our methods, we give an illustrative example demonstrating how unformalized knowledge can be captured, and we highlight relations to different uncertainty models, such as p-boxes, Dempster-Shafer structures, and α-level optimization for fuzzy sets.

Keywords— confidence regions, design optimization, higher dimensions, incomplete information, potential clouds

1 Introduction

Design optimization is frequently affected by uncertainties originating from several different sources. Being already a complicated task in absence of uncertainties, design optimization under uncertainty imposes an additional class of difficulties. We have developed a framework dividing design optimization under uncertainty in its two inherent components, i.e., uncertainty modeling and optimization.

The most critical problems in real-life uncertainty modeling are caused by the well-known curse of dimensionality (cf., e.g., [1]), and by lack of information. While in lower dimensions, lack of information can be handled with several tools (e.g., p-boxes [2], Dempster-Shafer structures [3]), in higher dimensions (say, greater than 10) there exist only very few. Often simulation techniques are used which, however, fail to be reliable in many cases, see, e.g., [4]. The clouds formalism [5] is one possibility to deal with both incomplete and higher dimensional information in a reliable and computationally tractable fashion.

The design optimization phase (cf., e.g., [6]) is the second major subject in our framework, loosely linked with the uncertainty modeling. One typically faces problems like strongly nonlinear, discontinuous, or black box objective functions, or mixed integer design variables. We have developed heuristics to solve these problems, e.g., using separable underestimation [7], or convex relaxation based splitting [8].

Since our approach can be considered as design optimization based on uncertainty modeling with clouds, we call the software *cloud based design optimization* (CBDO). We have implemented an interface for our methods that will be presented later in this paper. The implementation was motivated by the need of expert engineers of an easy-to-use tool, a framework respecting their working habits, and demonstrating usefulness in capturing and modeling incomplete, unformalized knowledge. Current research is focussed on improving both optimization and uncertainty modeling phase, and on capturing more types of information virtually, e.g., linguistic expressions. Of course, we are constantly looking for possible real-life applications of the methods. CBDO has already been successfully used in space system design applications, cf. [9, 10].

This paper is organized as follows. We introduce the formal background of CBDO in Section 2 also giving an illustrative example how we capture unformalized knowledge. In Section 3 we summarize relations of the potential clouds formalism to different uncertainty models. Finally, we present our software implementation, a MATLAB package for CBDO, in Section 4.

2 Clouds and robust optimization

Let ε be an n-dimensional random vector. A *potential cloud* is an interval-valued mapping $x \to [\underline{\alpha}(V(x)), \overline{\alpha}(V(x))]$, where the potential function $V : \mathbb{R}^n \to \mathbb{R}$ is bounded below, and $\underline{\alpha}, \overline{\alpha} : V(\mathbb{R}^n) \to [0,1]$ are functions constructed to be a lower and upper bound, respectively, for the cumulative distribution function (CDF) F of $V(\varepsilon)$, $\underline{\alpha}$ continuous from the left and monotone, $\overline{\alpha}$ continuous from the right and monotone. We define $\underline{C}_\alpha := \{x \in \mathbb{R}^n \mid V(x) \le \underline{V}_\alpha\}$ if $\underline{V}_\alpha := \min\{V_\alpha \in \mathbb{R} \mid \overline{\alpha}(V_\alpha) = \alpha\}$ exists, and $\underline{C}_\alpha := \emptyset$ otherwise; analogously $\overline{C}_\alpha := \{x \in \mathbb{R}^n \mid V(x) \le \overline{V}_\alpha\}$ if $\overline{V}_\alpha := \max\{V_\alpha \in \mathbb{R} \mid \underline{\alpha}(V_\alpha) = \alpha\}$ exists, and $\overline{C}_\alpha := \mathbb{R}^n$ otherwise. Thus we find a nested collection of lower and upper confidence regions in the sense that $\Pr(\varepsilon \in \underline{C}_\alpha) \le \alpha$, $\Pr(\varepsilon \in \overline{C}_\alpha) \ge \alpha$, $\underline{C}_\alpha \subseteq \overline{C}_\alpha$.

Note that lower and upper confidence regions $\underline{C}_\alpha, \overline{C}_\alpha$ – also called α-cuts of the cloud – are level sets of V. By choosing the potential function V reasonably one gets an uncertainty representation of high-dimensional, incomplete, and/or unformalized knowledge, cf. [11].

Our framework of cloud based design optimization consists of three essential parts, described in the following sections: uncertainty elicitation, uncertainty modeling, and robust optimization.

2.1 Uncertainty elicitation and modeling

We assume that the initially available uncertainty information consists of both formalized and unformalized knowledge. The formalized knowledge can be given as marginal CDFs, interval bounds on single variables, or real sample data. In real-life situations there is often only interval information, sometimes marginal CDFs without any correlation information, available for the uncertain variables. Moreover, there is typically a significant amount of unformalized knowledge available based

ISBN: 978-989-95079-6-8

on expert experience, e.g., knowledge about the dependence of variables.

Potential clouds enable to capture and formally represent this kind of information. We illustrate this by a simple example: First, we generate a data set from an $N(0,\Sigma)$ distribution with covariance matrix $\Sigma = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}$.

Assume that this data belongs to 2 random variables with a physical meaning, and that the data was given to an expert, without any information about the actual probability distribution of the random variables. Still, the expert may be able to provide vague, unformalized information about the dependence of the variables (opposed to formal knowledge, e.g., correlation information) from his knowledge about the physical relationship between the variables. We model this knowledge by polyhedral constraints on the variables, see, e.g., Fig. 3. We choose the potential function V according to these constraints, i.e., the lower and upper confidence regions $\underline{C}_\alpha, \overline{C}_\alpha$ constructed with clouds become polyhedra. The polyhedra reasonably approximate confidence regions of the true, but unknown distribution linearly, as shown in Fig. 1, although the information was vague and unformalized.

In more than 2 dimensions the polyhedral constraints are provided for projections to 1-dimensional or 2-dimensional subspaces.

It should also be highlighted that this approach for providing unformalized knowledge also allows for information updating, simply by adding further polyhedral constraints.

On the basis of the given information we use the confidence regions constructed by clouds in order to search for worst-case scenarios of certain design points via optimization techniques. The construction of the confidence regions is possible even in case of scarce, high-dimensional data, incomplete information, unformalized knowledge.

For further details on the construction of potential clouds the interested reader is referred to [11]. A comparison of different existing uncertainty models can be found in [12], and Section 3 gives a short summary.

2.2 *Robust optimization*

Assume that we wish to find the design point $\theta = (\theta^1, \theta^2, \ldots, \theta^{n_0})$ with the minimal design objective function value g under uncertainty of the n-dimensional random vector ε. Let $\mathbf{T}$ be the set of possible selections for the design point θ. Assume that the function G models the functional relationship between different design components and the objective function. Also assume that the uncertainty of ε is described by a convex set $\mathcal{C}$, in our case a polyhedral α-cut from the cloud.

We embed the confidence regions constructed above in a problem formulation for robust design optimization as follows:

$$\begin{aligned} \min_{\theta} \quad & \max_{\varepsilon} \; g(x) \\ \text{s.t.} \quad & x = G(\theta, \varepsilon), \\ & \varepsilon \in \mathcal{C}, \\ & \theta \in \mathbf{T}, \end{aligned} \tag{1}$$

where $g : \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^{n_0} \times \mathbb{R}^n \to \mathbb{R}^m$.

The optimization phase minimizes a certain objective function g (e.g., cost, mass of the design) subject to safety constraints $\varepsilon \in \mathcal{C}$ to account for the robustness of the design, and subject to the functional constraints which are represented by the underlying system model G.

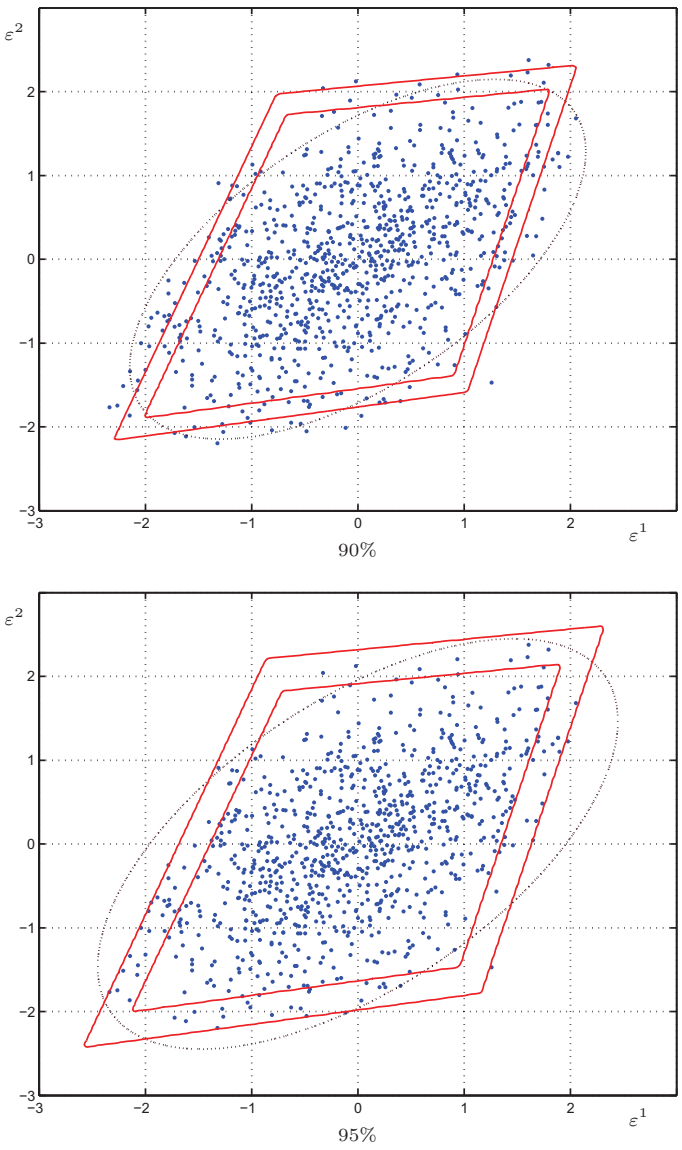


Figure 1: Approximation of confidence regions by 90% and 95% α-cuts, respectively: The polyhedral cloud results in confidence regions that reasonably approximate confidence regions of the true $N(0,\Sigma)$ distribution although the information was given unformalized.

The main difficulties arising from (1) are imposed by the bilevel structure in the objective function, by the mixed integer formulation (since θ^i can be either a discrete or continuous variable), and by the fact that G may comprise strong nonlinearities, or discontinuities, or may be given as a black box.

We have developed multiple techniques to tackle these difficulties and find a solution of (1). For details on approaches to solve such problems of design optimization under uncertainty the interested reader is referred to [7]. The latest improvements of the methods can be found in [8].

3 Relations to different uncertainty models

This section illustrates relations and differences of the potential clouds formalism to three other existing uncertainty mod-

ISBN: 978-989-95079-6-8

els: p-boxes, Dempster-Shafer structures, and α-level optimization for fuzzy sets.

3.1 Relation to p-boxes

A p-box – or p-bound, or probability bound – is a rigorous enclosure of the CDF F of a univariate random variable X, $F_l \leq F \leq F_u$, in case of partial ignorance about specifications of F. Such an enclosure enables, e.g., to compute lower and upper bounds on expectation values or failure probabilities.

There are different ways to construct a p-box depending on the available information about X, cf. [13]. Moreover, it is possible to construct p-boxes from different uncertainty models like Dempster-Shafer structures (cf. Section 3.2). The studies on p-boxes have already lead to successful software implementations, cf. [14, 2].

Higher order moment information on X (e.g., correlation bounds) cannot be handled or processed yet. This is a current research field, cf., e.g., [15]. In higher dimensions, the definition of p-boxes can be generalized similar to the definition of higher dimensional CDFs, cf. [16].

The problem of rigorously quantifying probabilities given incomplete information – as done with p-boxes – is highly complex, even for simple problems, e.g., [17]. Applications of the methods are rather restricted to lower dimensions and non-complex system models G. Black box functions G cannot be handled as one requires knowledge about the involved arithmetic operations. All in all, the methods often appear not to be reasonably applicable in many real-life situations.

The relation to potential clouds becomes obvious, regarding $V(\varepsilon)$ as a 1-dimensional random variable and the functions $\underline{\alpha}$, $\overline{\alpha}$ as a p-box for $V(\varepsilon)$. Thus the potential clouds approach extends the p-box concept to the case of multidimensional ε, without the exponential growth of work in the conventional p-box approach.

3.2 Relation to Dempster-Shafer structures

Dempster-Shafer theory [3] enables to process incomplete and even conflicting uncertainty information. Let $\varepsilon : \Omega \to \mathbb{R}^n$ be an n-dimensional random vector. One formalizes the available information by a so-called *basic probability assignment* $m : 2^\Omega \to [0,1]$ on a finite set $\mathcal{A} \subseteq 2^\Omega$ of non-empty subsets A of Ω, such that

$$m(A) \begin{cases} > 0 & \text{if } A \in \mathcal{A}, \\ = 0 & \text{otherwise,} \end{cases} \tag{2}$$

and the normalization condition $\sum_{A\in\mathcal{A}} m(A) = 1$ holds.

The basic probability assignment m is interpreted as the exact belief focussed on A, and not in any strict subset of A. The sets $A \in \mathcal{A}$ are called *focal sets*. The structure $(m, \mathcal{A})$, i.e., a basic probability assignment together with the related set of focal sets, is called a *Dempster-Shafer structure* (DS structure).

Given a DS structure $(m, \mathcal{A})$ one constructs two fuzzy measures Bel and Pl by

$$\mathrm{Bel}(B) = \sum_{\{A\in\mathcal{A}|A\subseteq B\}} m(A), \tag{3}$$

$$\mathrm{Pl}(B) = \sum_{\{A\in\mathcal{A}|A\cap B\neq\emptyset\}} m(A), \tag{4}$$

for $B \in 2^\Omega$. The fuzzy measures Bel and Pl have the property $\mathrm{Bel} \leq \mathrm{Pr} \leq \mathrm{Pl}$ by construction, where Pr is the probability measure that is unknown due to lack of information.

DS structures can be obtained from expert knowledge or in lower dimensions from histograms, or from the Chebyshev inequality given expectation value μ and variance σ^2 of a univariate random variable X, see, e.g., [18].

To combine different, possibly conflicting DS structures $(m_1, \mathcal{A}_1)$, $(m_2, \mathcal{A}_2)$ (in case of multiple bodies of evidence, e.g., several different expert opinions) to a new basic probability assignment m_{new} one uses Dempster's rule of combination [19].

The complexity of the rule is strongly increasing in higher dimensions, and in many cases requires independence assumptions for simplicity reasons avoiding problems with interacting variables. It is not yet understood how the dimensionality issue can be solved. Working towards more efficient computational implementations of evidence theory it can be attempted to decompose the high-dimensional case in lower dimensional components which leads to so-called compositional models, cf., e.g., [20].

The extension of a function $G(\varepsilon)$ is based on the joint DS structure $(m, \mathcal{A})$ for ε. The new focal sets of the extension are $B_i = G(A_i)$, $A_i \in \mathcal{A}$, the new basic probability assignment is $m_{\mathrm{new}}(B_i) = \sum_{\{A_i\in\mathcal{A}|G(A_i)=B_i\}} m(A_i)$.

To embed DS theory in design optimization one formulates a constraint on the upper bound of the failure probability p_f which should be smaller than an admissible failure probability p_a, i.e., $\mathrm{Pl}(\mathbb{F}) \leq p_a$, for a failure set $\mathbb{F}$. This is similar to the safety constraint in (1). It can be studied in more detail in [21] as evidence based design optimization (EBDO).

It is possible to generate a DS structure that approximates a given potential cloud discretely. Fix some confidence levels $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_N = 1$ of the potential cloud, then generate focal sets and the associated basic probability assignment by

$$A_i := \overline{C}_{\alpha_i} \backslash \underline{C}_{\alpha_i}, \tag{5}$$

$$m(A_1) = \alpha_1, m(A_i) = \alpha_i - \alpha_{i-1}, i = 2, \ldots, N. \tag{6}$$

Thus the focal sets are determined by the level sets of V. An analogous recipe works for approximating p-boxes by DS structures, cf. [13]. Note that focal sets A_i in this construction are not nested, so the fuzzy measures Bel and Pl belonging to $(m, \mathcal{A})$ are not equivalent to possibility and necessity measures.

Conversely, assume that one has a DS structure and the associated fuzzy measures Bel and Pl for the random variable $X := V(\varepsilon)$. Then

$$\underline{\alpha}(t) := \mathrm{Bel}(\{X \leq t\}), \tag{7}$$

$$\overline{\alpha}(t) := \mathrm{Pl}(\{X \leq t\}) \tag{8}$$

give bounds on the CDF of $V(\varepsilon)$ and thus construct a potential cloud.

3.3 Relation to fuzzy sets and α-level optimization

To see an interpretation of potential clouds in terms of fuzzy sets one may consider $\underline{C}_\alpha$, $\overline{C}_\alpha$ as α-cuts of a multidimensional interval valued membership function defined by $\underline{\alpha}$ and $\overline{\alpha}$. The major difference is given by the fact that clouds allow

ISBN: 978-989-95079-6-8

for probabilistic statements, i.e., one cannot go back in the other direction and construct a cloud from a multidimensional interval valued membership function because of the lack of the probabilistic properties mentioned in Section 2. If the interval valued membership function does have these probabilistic properties, it corresponds to consistent possibility and necessity measures [22] which are related to interval probabilities [23].

However, the interpretation of a potential cloud as a fuzzy set with such a membership function shows strong links to α-level optimization for fuzzy sets [24].

The α-level optimization method combines the extension principle and the α-cut representation of a membership function μ of an uncertain variable ε, i.e.,

$$\mu(x) = \sup_{\alpha} \min(\alpha, 1_{C_\alpha}(x)), \tag{9}$$

where 1_A denotes the characteristic function of the set A, $C_\alpha := \{x \mid \mu(x) \geq \alpha\}$ denotes the α-cut of the fuzzy set, in order to determine the membership function μ_f of a function $f(\varepsilon)$, $f : \mathbb{R}^n \to \mathbb{R}$. This is achieved by constructing the α-cuts $C_{f\,\alpha_i}$ belonging to μ_f from the α-cuts C_{α_i} belonging to μ. To this end one solves the optimization problems

$$\min_{\varepsilon \in C_{\alpha_i}} f(\varepsilon), \tag{10}$$

$$\max_{\varepsilon \in C_{\alpha_i}} f(\varepsilon), \tag{11}$$

for different discrete values α_i. Finally from the solution f_{i_*} of (10) and f_i^* of (11) one constructs the α-cuts belonging to μ_f by $C_{f\,\alpha_i} = [f_{i_*}, f_i^*]$.

To simplify the optimization step one assumes sufficiently nice behaving functions f and computationally nice fuzzy sets, i.e., convex fuzzy sets, typically triangular shaped fuzzy numbers.

In n dimensions one optimizes over a hypercube, obtained by the Cartesian product of the α-cuts, i.e., $C_{\alpha_i} = C^1_{\alpha_i} \times C^2_{\alpha_i} \times \cdots \times C^n_{\alpha_i}$, where $C^j_{\alpha_i} := \{\varepsilon^j \mid \mu^j(\varepsilon^j) \geq \alpha_i\}$, $\mu^j(\varepsilon^j) := \sup_{\varepsilon^k, k \neq j} \mu(\varepsilon)$, $\varepsilon = (\varepsilon^1, \varepsilon^2, \ldots, \varepsilon^n)$. Here one has to assume non-interactivity of the uncertain variables $\varepsilon^1, \ldots, \varepsilon^n$.

Using a discretization of the α-levels by a finite choice of α_i the computational effort for this methods becomes tractable. From (9) one gets a step function for μ_f which is usually linearly approximated through the points f_{i_*} and f_i^* to generate a triangular fuzzy number.

Now interpret $\underline{\alpha}$ and $\overline{\alpha}$ from a potential cloud as a multidimensional interval valued membership function and consider a system model $f(\varepsilon) := g(G(\theta, \varepsilon))$ with fixed θ (cf. Section 2.2). Similar to (10,11), optimization of f over $\underline{C}_{\alpha_i}$ for discrete values α_i would give a discretized version of $\underline{\alpha}_f$, i.e., the function belonging to the cloud for $f(\varepsilon)$ given the cloud for ε. Analogously, optimization of f over $\overline{C}_{\alpha_i}$ would give a discretized version of $\overline{\alpha}_f$.

This idea leads to the calculation of functions of clouds which is a current research topic. Also note that in α-level optimization one optimizes over boxes C_{α_i}, that means one assumes that the uncertain variables do not interact. Here a similar idea like interactive polyhedral constraints as described in Section 2.1 could also apply to model unformalized knowledge about interaction of the variables.

4 Cloud based design optimization GUI

We have realized the methods for CBDO in a *graphical user interface* (GUI). To install the software go to the CBDO website [25] and download the CBDO package. A quickstart guide helps through the first steps of the simple installation. A more detailed user manual is also included. How to set up a MATLAB file containing a user defined model is illustrated by an example included in the package.

We have developed the GUI using a sequential, iterative structure. The first and second step represent the uncertainty elicitation. In the first step, the user provides an underlying system model and all formal uncertainty information on the input variables of the model available at the current design stage. In the second step, polyhedral dependence constraints between the variables can be added, cf. Section 2.1. In the third step, the initially available information is processed to generate a cloud that provides a nested collection of confidence regions parameterized by the confidence level α. Thus we produce safety constraints for the optimization (cf. Section 2.2) which is the next step in the GUI. The results of the optimization, i.e., the optimal design point found and the associated worst-case analysis, are returned to the user. In an iterative step the user is eventually given an interactive possibility of adaptively refining the uncertainty information and rerunning the procedure until satisfaction.

4.1 Uncertainty elicitation

After starting the GUI with `cbdogui` from the CBDO folder in MATLAB it asks whether to load the last state to the workspace unless it is run for the first time. In the latter case one should first configure the options to set up the model file and inputs declaration file names, and other user-defined parameters after clicking *Options*/*Edit Options*. The notation – if not self-explanatory – is described in the user manual. Tooltips are given for each option in the GUI to guide the user through the configuration.

Having set up the options one returns to the uncertainty elicitation clicking *Back*. The initially available information can be specified in an inputs declaration file and is modified choosing a variable's name and specifying its associated marginal CDF, or interval bound, respectively, in the first step of the GUI, cf. Fig. 2.

The *Next* button leads to the next step which is scenario exclusion.

4.2 Scenario exclusion

From the information given in the first step the program generates a sample as described in [11]. The second step enables the user to exclude scenarios by polyhedral constraints as shown in Section 2.1, illustrating the great advantage of this approach in modeling unformalized knowledge.

To this end the user selects a 1-dimensional or 2-dimensional projection of the generated sample using the field *Projection* on the right. To add a constraint one hits the *Add constraint* button and defines a linear exclusion by two clicks into the sample projection on the left. All linear constraints can be selected from the *Constraint Selection* box to revise and possibly remove them via the *Remove constraint* button. Fig. 3 shows a possible exclusion in two dimensions.

ISBN: 978-989-95079-6-8

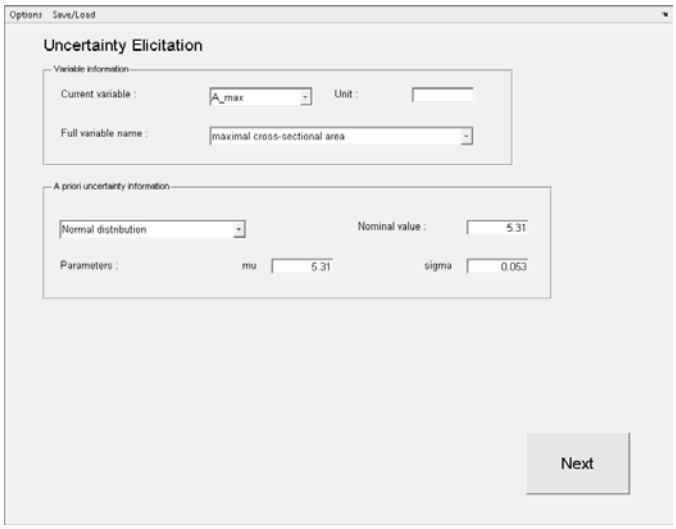


Figure 2: Example for the uncertainty elicitation GUI.

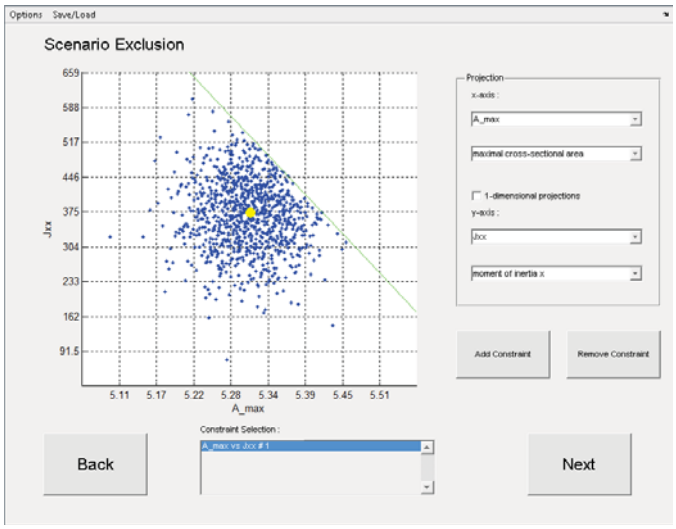


Figure 3: Example for scenario exclusion.

After the exclusion the *Next* button leads to the optimization phase.

4.3 Optimization

The *Start* button initiates two computations: potential cloud generation for (1), and optimization, cf. [7, 8]. As a result one gets the optimal design point found by the program, and the associated objective function value, cf. Fig. 4. It should be remarked that the workspace of the optimization including all results is stored as .mat files in the `cbdo` directory.

The user now has the possibility for the adaptive analysis of the results. Thus the *Next* button leads back to the uncertainty elicitation to be refined.

4.4 Adaptive step

The GUI determining the *a priori* information is not modifiable anymore at this stage of the program. Meanwhile, observe that in the lower part of the GUI a histogram illustrates weighted marginal distributions of the sample.

Hitting the *Next* button makes the scenario exclusion appear again and enables the *a posteriori* adaption of the uncertainty

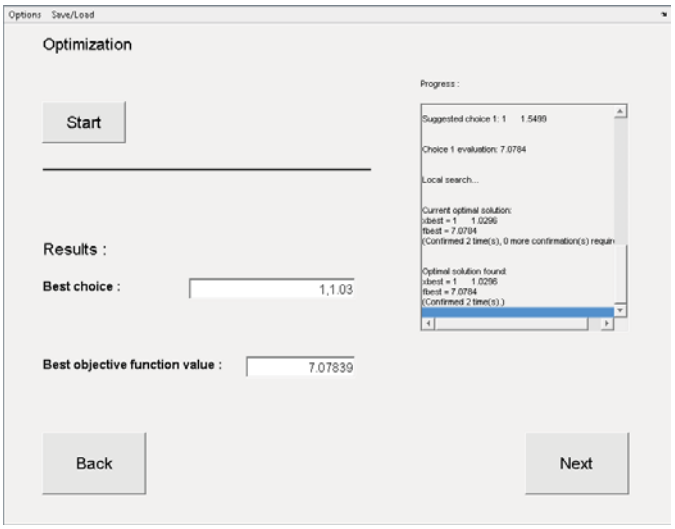


Figure 4: Example for the optimization phase.

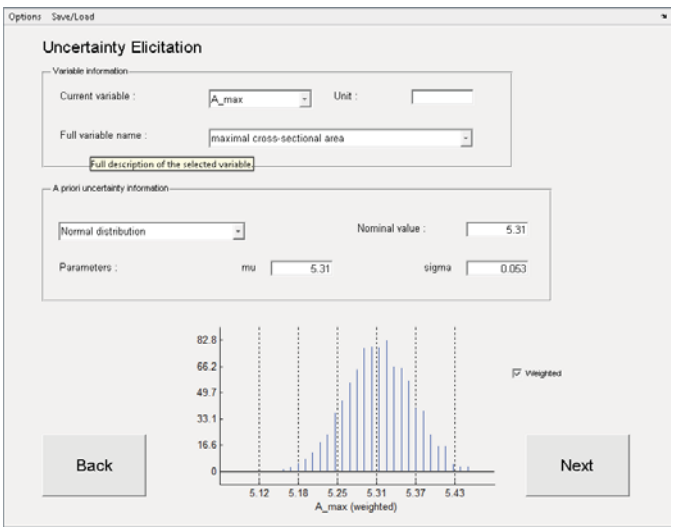


Figure 5: Example for uncertainty elicitation in the adaptive step.

information. For example, the user can consider the worst-case analysis (the worst-case scenario is highlighted with a red dot) to be too pessimistic and exclude it, cf. Fig. 6. Note that this approach is very much imitating real-life working habits of engineers! In early design phases little information is available and safety margins are refined or coarsened iteratively.

The *Next* button leads to the optimization phase again and the user can rerun the procedure until satisfaction.

Acknowledgments

I would like to thank Arnold Neumaier for the valuable discussions during creation of this paper. Also I would like to thank the anonymous reviewers for their constructive comments.

References

[1] P.N. Koch, T.W. Simpson, J.K. Allen, and F. Mistree. Statistical approximations for multidisciplinary optimization: The problem of size. *Special Issue on Multidisciplinary Design Optimization of Journal of Aircraft*, 36(1):275–286, 1999.

ISBN: 978-989-95079-6-8

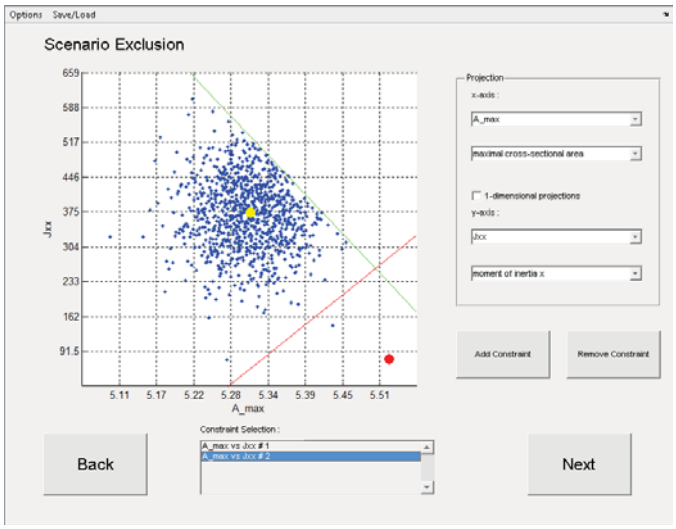


Figure 6: Example for *a posteriori* scenario exclusion.

[2] S. Ferson. *Ramas Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers*. Lewis Publishers, U.S., 2002.

[3] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[4] S. Ferson, L. Ginzburg, and R. Akcakaya. Whereof one cannot speak: When input distributions are unknown. *Risk Analysis*, 1996. In press, available on-line at: `www.ramas.com/whereof.pdf`.

[5] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10(4):249–272, 2004.

[6] N.M. Alexandrov and M.Y. Hussaini. Multidisciplinary design optimization: State of the art. In *Proceedings of the ICASE/NASA Langley Workshop on Multidisciplinary Design Optimization*, Hampton, Virginia, USA, 1997.

[7] M. Fuchs and A. Neumaier. Autonomous robust design optimization with potential clouds. *International Journal of Reliability and Safety, Special Issue on Reliable Engineering Computing*, 2008. Accepted, preprint available on-line at: `www.martin-fuchs.net/publications.php`.

[8] M. Fuchs and A. Neumaier. A splitting technique for discrete search based on convex relaxation. Submitted, 2009. Preprint available on-line at: `www.martin-fuchs.net/publications.php`.

[9] M. Fuchs, D. Girimonte, D. Izzo, and A. Neumaier. *Robust Intelligent Systems*, chapter Robust and automated space system design, pages 251–272. Springer, 2008.

[10] A. Neumaier, M. Fuchs, E. Dolejsi, T. Csendes, J. Dombi, B. Banhelyi, and Z. Gera. Application of clouds for modeling uncertainties in robust space system design. ACT Ariadna Research ACT-RPT-05-5201, European Space Agency, 2007.

[11] M. Fuchs and A. Neumaier. Potential based clouds in robust design optimization. *Journal of Statistical Theory and Practice, Special Issue on Imprecision*, 2008. Accepted, preprint available on-line at: `www.martin-fuchs.net/publications.php`.

[12] M. Fuchs. Clouds, p-boxes, fuzzy sets, and other uncertainty representations in higher dimensions. *Acta Cybernetica, CSCS Special Issue*, 2009. Accepted, preprint available on-line at: `www.martin-fuchs.net/publications.php`.

[13] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Sand Report SAND2002-4015, Sandia National Laboratories, 2003. Available on-line at `www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf`.

[14] D. Berleant and L. Xie. An interval-based tool for verified arithmetic on random variables of unknown dependency. Manuscript, 2005.

[15] S. Ferson, L. Ginzburg, V. Kreinovich, and J. Lopez. Absolute bounds on the mean of sum, product, etc.: A probabilistic extension of interval arithmetic. In *Extended Abstracts of the 2002 SIAM Workshop on Validated Computing*, pages 70–72, Toronto, Canada, 2002.

[16] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations–I: Generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3):649–663, 2008.

[17] V. Kreinovich, S. Ferson, and L. Ginzburg. Exact upper bound on the mean of the product of many random variables with known expectations. *Reliable Computing*, 9(6):441–463, 2003.

[18] M. Oberguggenberger, J. King, and B. Schmelzer. Imprecise probability methods for sensitivity analysis in engineering. In *Proceedings of the 5th International Symposium on Imprecise Probability: Theories and Applications*, pages 317–325, Prague, Czech Republic, 2007.

[19] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967.

[20] R. Jirousek, J. Vejnarova, and M. Daniel. Compositional models of belief functions. In *Proceedings of the 5th International Symposium on Imprecise Probability: Theories and Applications*, pages 243–251, Prague, Czech Republic, 2007.

[21] Z.P. Mourelatos and J. Zhou. A design optimization method using evidence theory. *Journal of Mechanical Design*, 128(4):901–908, 2006.

[22] W.A. Lodwick and K.D. Jamison. Interval-valued probability in the analysis of problems containing a mixture of possibilistic, probabilistic, and interval uncertainty. *Fuzzy Sets and Systems*, 159(21):2845–2858, 2008.

[23] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2–3):149–170, 2000.

[24] B. Möller, W. Graf, and M. Beer. Fuzzy structural analysis using α-level optimization. *Computational Mechanics*, 26(6):547–565, 2000.

[25] CBDO. Cloud Based Design Optimization webpage `www.martin-fuchs.net/cbdo.php`, 2009.

ISBN: 978-989-95079-6-8

Local models for the analysis of spatially varying relationships in a lignite deposit

Bulent Tutmez[1] A. Erhan Tercan[2] Uzay Kaymak[3] Christopher D. Lloyd[4]

1 School of Engineering, Inonu University, Malatya 44280, Turkey
2 School of Engineering, Hacettepe University, Ankara 06532, Turkey
3 Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, the Netherlands
4 School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast, Northern Ireland, UK
Email: btutmez@inonu.edu.tr

Abstract *— Relationships between geographically referenced variables are usually spatially heterogeneous and, to account for such variations, local models are necessary. This paper compares the Geographically Weighted Regression (GWR) model, usually used to integrate and examine the spatial heterogeneity of a relationship, and the Fuzzy Clustering-Based Least Squares (FCBLS) model for the analysis of spatially varying relationships. Both models use the same model parameters and bandwidth values derived from the Akaike Information Criterion. The results show that FCBLS outperforms the GWR model.*

Keywords— *GWR, local model, weighted regression, fuzzy clustering, lignite.*

1 Introduction

Spatial measures contain both attribute and locational information. By their nature, local analyses focus on differences across space whereas global analyses focus on similarities across space. Recently, a variety of useful regression models have been developed to explore the spatial nature of variables [1].

Local modelling has been employed widely in some disciplines for several decades. For example, in image processing local filters have long been used to smooth or sharpen images. However, in some disciplines, like the geosciences in general and geography in particular, a focus on methods that account for local variation has been a comparatively recent development [2]. A variety of localized modelling techniques have been proposed to capture spatial heterogeneity. Geographically Weighted Regression (GWR) is one technique which is being increasingly widely used to explore local spatial variations in relationships, [3]. GWR is a useful and effective technique for locally modelling relationships by calibrating a spatially varying coefficient regression model.

Spatial statistics was developed based on probability and classical statistics. However, many spatial datasets have high levels of uncertainty, and in some cases analyses are dependent on 'soft' data, which may be more qualitative than quantitative in nature [4]. On the other hand, soft approaches like fuzzy computing have desirable features for spatial data analysis [5]: they are based on less restrictive assumptions, and are flexible in modelling non-linearity and non-constant variable structures.

Lignite quality parameters considered in this study have crucial importance in the production of energy in power plants and thus modeling of these parameters is useful in making decisions and in planning future production levels of operating plants [6]. In this study, the spatial relationships between the lignite quality parameters are investigated by local models. To meet this aim, the two local regression methods, GWR and fuzzy clustering based Least Squares, have been applied for local modelling of lignite reserve parameters derived from a deposit. First, both the methods have been used for developing weighted local models and then the performance comparisons have been conducted using a variety of performance indices.

2 Problem and methods

2.1 Description of problem

In general, spatial models and methods of spatial evaluation have been applied at a 'global' level, meaning that one set of results is generated from the models, representing one set of relationships, which is assumed to apply equally across the study area [7]. Although global approaches have proved useful they have the shortcoming that they can mask geographical variations in relationships. Owing to this realization, local regression model approaches have been developed that permit the exploration of spatially varying relationships in datasets [8].

The problem of local estimation considered in this study can be formulated as follows: we have a region and lignite quality values $\{y_\alpha\}, \alpha = 1,...,n,$ at n sampled locations. We aim to estimate the values $\{y_\beta\}, \beta = 1,...,N$ at unsampled locations. To model a system by a spatially weighted method, the identification data and the weights could be arranged as in the following matrices:

ISBN: 978-989-95079-6-8

$$\mathbf{X}=\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix},\ \mathbf{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix},\ \mathbf{W}_i=\begin{bmatrix} w_{i1} & 0 & 0 & 0 \\ 0 & w_{i2} & 0 & 0 \\ 0 & 0 & w_{i3} & 0 \\ 0 & 0 & 0 & w_{iN} \end{bmatrix} \quad (1)$$

2.2 Geographically Weighted Regression (GWR)

An assumption of global regression is that the relationship under study is spatially constant, and thus, the relationships being measured are assumed to be 'stationary' over space. However, in most cases, the relationship varies in space. GWR, as a refinement to traditional regression methods, explicitly deals with the spatial non-stationary of empirical relationships. The method provides a weighting of information that is locally specific, and allows regression model parameters to vary in space [3].

Classical regression equation in matrix form can be given by:

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\varepsilon \quad (2)$$

where the vector of parameters to be estimated, β, is constant over space and is estimated by

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}. \quad (3)$$

The local estimation of the parameters with GWR is as given by (3), but the difference is that the observations used in GWR are weighted in accordance with their distance from the kernel centre. Fig. 1 indicates a spatial kernel in GWR. The parameters for GWR may be estimated by solving

$$\hat{\beta}(u_i,v_i)=\left[\mathbf{X}^T\mathbf{W}(u_i,v_{i,})\mathbf{X}\right]^{-1}\mathbf{X}^T\mathbf{W}(u_i,v_i)\mathbf{Y}, \quad (4)$$

where $\hat{\beta}$ represents an estimate of β, and $\mathbf{W}(u_i,v_i)$ is an n by n matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of each of the n observed data for regression point i [9]. There are many weighting schemes which express w_{ij} as a continuous function of distance between i and j, d_{ij}. In this study, the following Gaussian function has been used.

$$w_{ij}=\exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right] \quad (5)$$

where d_{ij} is the Euclidean distance between the location of measurement i and the centre of the kernel and b is the bandwidth of the kernel.

2.3 Fuzzy Clustering-Based Least Squares

A class of fuzzy clustering algorithms can be used to approximate a set of data by local linear models. Each of these models is represented by a fuzzy subset in the data set available for identification [10]. Most analytical fuzzy clustering algorithms are based on optimization of the basic c-means objective function, or some modification of it.

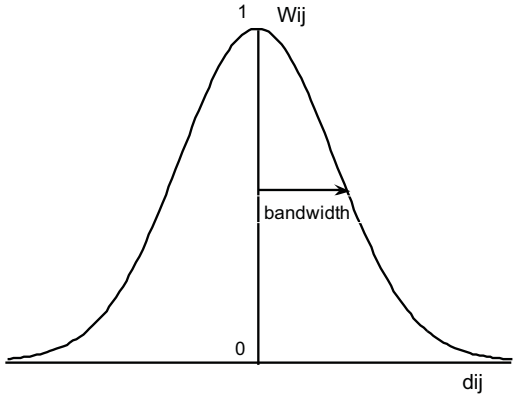


Figure 1: A Gaussian kernel in GWR.

In the present study, we used a general structure which is the fuzzy c-means functional [11] for constructing the weighted least squares model from the fuzzy partitions.

In the FCBLS model, the output parameters for the ith cluster, a_i and b_i are connected by a single parameter vector $\boldsymbol{\theta}_\mathbf{i}$ as follows:

$$\boldsymbol{\theta}_\mathbf{i}=\left[\mathbf{a}_i^T,b_i\right]^T. \quad (6)$$

Appending a unitary column to $\mathbf{X}$ gives the extended regressor matrix $\mathbf{X_e}$:

$$\mathbf{X_e}=[\mathbf{X},\mathbf{1}]. \quad (7)$$

Assuming that each cluster represents a local linear model of the system, the consequent parameter vectors $\boldsymbol{\theta}_\mathbf{i}$, $i=1,2,...,c$, can be estimated independently by the least-squares method. The membership degrees of the fuzzy partition serve as the weights expressing the relevance of the data pair x_k, y_k to that local model [10]. If the columns of $\mathbf{X}_e$ are linearly independent, then

$$\boldsymbol{\theta}_i=\left[\mathbf{X}_e^T\mathbf{W}_i\mathbf{X}_e\right]^{-1}\mathbf{X}_e^T\mathbf{W}_i\mathbf{y} \quad (8)$$

is the least-squares solution of $\mathbf{y}=\mathbf{X_e}\boldsymbol{\theta}+\boldsymbol{\varepsilon}$ where the kth data pair (x_k,y_k) is weighted by w_{ik}. The parameters $\mathbf{a}_i$ and b_i are given by:

$$\mathbf{a}_i=\left\lfloor\theta_1,\theta_2,...,\theta_p\right\rfloor,\ b_i=\theta_{p+1}. \quad (9)$$

3 Implementation

3.1 Data set

The study area, the Sivas-Kalburcayiri field, is one of the most important lignite reserves in Turkey [6]. Lignite seams in this field are used to feed coal to a power plant. In this study, one of the lignite quality parameters, Calorific Value (CV), has been estimated from the independent parameters which are spatial coordinates (x,y), ash and sulphur contents.

ISBN: 978-989-95079-6-8

The locations of the 67 records of the field have been employed in the model. The data set was divided into two subsets randomly: the training set (60%: 40 records) and the validation set (40%:27 records), respectively. For the analyses, data conditioning is necessary. In the present study, scaling was carried out by the local metric (L-metric) rescaling, in which the minimum and maximum values of x_{ij} for each j are respectively mapped onto zero and one respectively [7],

$$x_{ij}^{L} = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})}. \quad (10)$$

3.2. Local Models

GWR model

For the local analysis of the lignite parameters, GWR analyses were undertaken using a fixed spatial kernel with Gaussian function in which the bandwidth was determined by minimization of the Akaike Information Criterion (AIC) [12].

In the literature, various approaches have been used for ascertaining an optimal bandwidth. A method of deriving the bandwidth which provides a trade-off between goodness-of-fit and degrees of freedom is to minimize the AIC [12], used in this study. In [9], the AIC has been modified for GWR as follows:

$$AIC_c = 2n\log_e\left(\hat{\sigma}\right) + n\log_e\left(2\prod\right) + n\left\{\frac{n + tr(\mathbf{S})}{n - 2 - tr(\mathbf{S})}\right\} \quad (11)$$

where n is the sample size, $\hat{\sigma}$ is the estimated standard deviation of the error term, and tr($\mathbf{S}$) denotes the trace of the hat matrix $\mathbf{S}$ $\left(\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}\right)$. The matrix S maps $\hat{\mathbf{y}}$ on to $\mathbf{y}$.

In this application, the GWR model was fitted using a computer software program, *SAM 3.0.* [13]. In addition, the fixed bandwidth value has been determined from AIC using '*spgwr*' package in *R* [14]. Table 1 summarizes the model fitting statistics for training data. Similarly, the test statistics which were obtained from the fixed bandwidth are given in Table 2. In these tables, r^2 relates to observations against their estimates and OLS denotes the Ordinary Least Squares optimization.

Table 1. GWR fitting statistics for training data.

Spatial function:	Gaussian
Fixed Bandwidth (h):	0.191 distance units.
Number of Locations to Fit Model (n):	40
Coefficient of Determination (r^2):	0.811 (OLS: 0.785)

Table 2. GWR fitting statistics for testing data.

Spatial function:	Gaussian
Fixed Bandwidth (h):	0.191 distance units.
Number of Locations to Fit Model (n):	27
Coefficient of Determination (r^2):	0.718 (OLS: 0.655)

FCBLS model

To analyse the spatial relationships by clustering based least-squared optimization, firstly a clustering operation was conducted. The optimal number of clusters was determined experimentally using an index method which has been presented in [6]. Because the data variability is crucial in spatial data modelling, the method aims to reproduce variability of the sample data in the CV of cluster centres with a minimum number of clusters as follows:

$$\text{Minimize } n_c \quad under \quad Std\left[CV(x)\right] \approx Std\left[CV(c)\right] \quad (12)$$

where n_c is the optimal number of clusters, *Std* is the standard deviation of CV values. In this application, the appropriate numbers of clusters was determined as two. The cluster centres are depicted in Fig. 2.

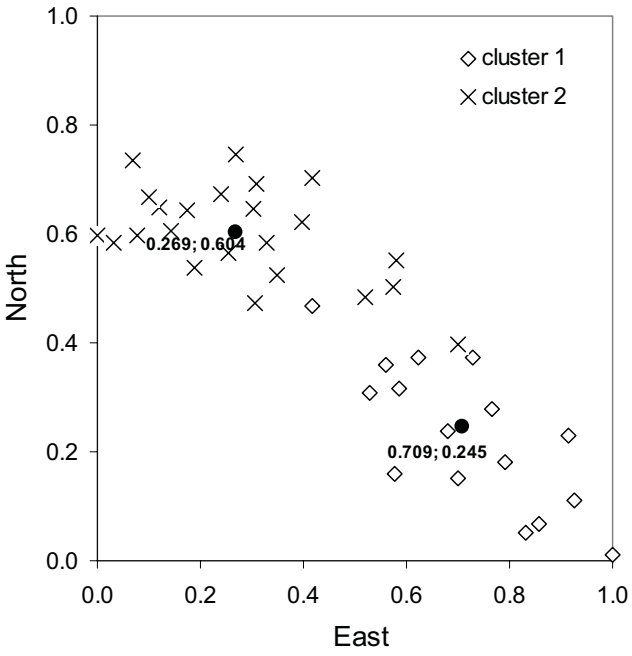


Figure 2: Data and cluster centres.

By using the information taken from the clustering, Gaussian type membership (weighting) functions were adopted. Fig.3 shows the input memberships considered in the model. To optimize the least square systems, weights taken from the Gaussian memberships have been used. Note that the Gaussian functions developed in the model used the same bandwidths derived from AIC. The consequent parameters are summarized in Table 3.

ISBN: 978-989-95079-6-8

Table 3: Consequent parameters for FCBLS model.

Cluster	*Ash*	*Sulphur*	*Constant*
Cluster 1	-0.045	0.957	0.031
Cluster 2	-0.403	0.532	0.0247

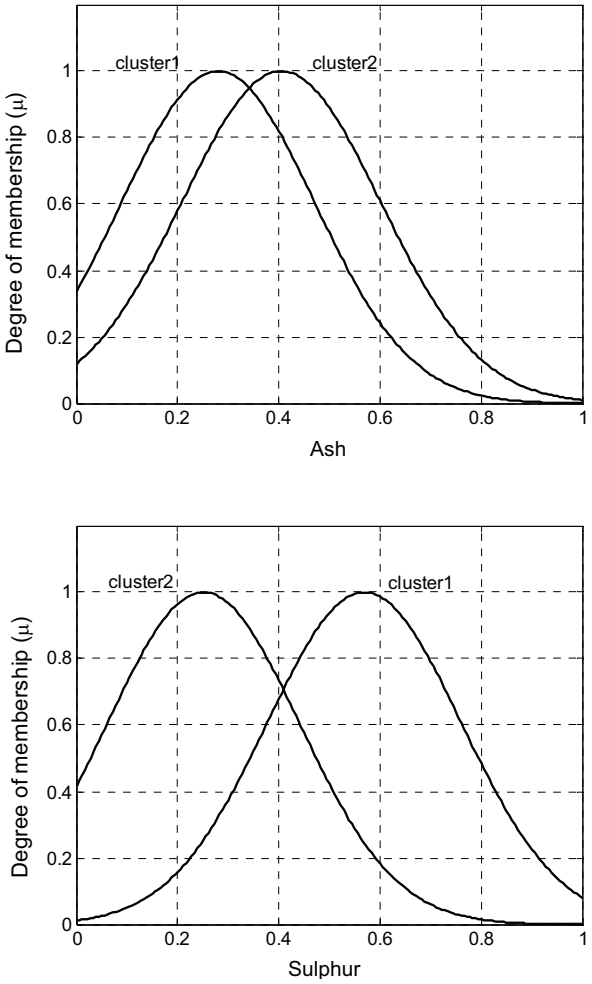


Figure 3: Gaussian memberships.

4 Results and conclusions

To assess the performance of the local models detailed in section 3.2, we have plotted the observed calorific values against the estimated calorific values. Fig. 4 illustrates the GWR model results together with the cross-correlations between estimated and observed data both for the training set and the validation set, respectively. Similarly, the results of the FCBLS model are depicted in Fig. 5.

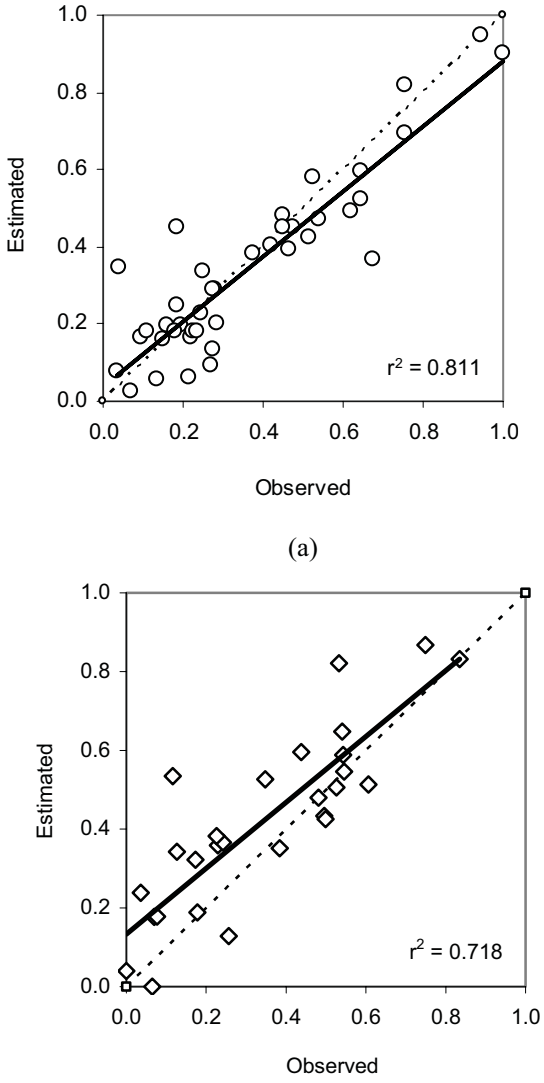
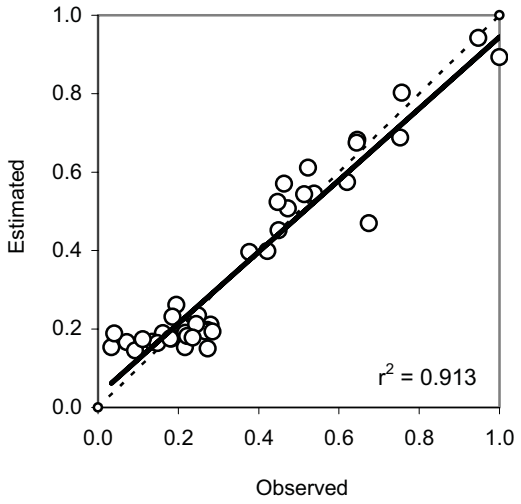


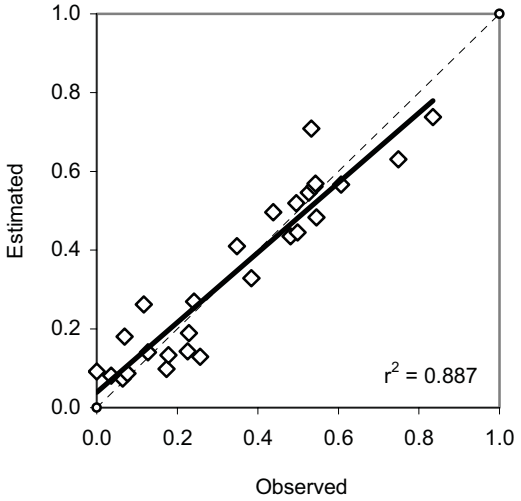
Figure 4: Scatter plot for GWR model: (a) training data and (b) testing data.

The large determination coefficient (r^2) shows that the model has good estimation capability. As observed from the determination coefficient, the FCBLS method outperforms the method based on GWR. In addition to r^2, the performance of the models has been assessed using the performance index, namely, the Variance Accounted For (VAF). In multivariate analysis, the measure of VAF plays a central role. Table 4 gives the VAF values.

ISBN: 978-989-95079-6-8



(a)



(b)

Figure 5: Scatter plot for FCBLS model: (a) training data and (b) testing data.

Table 4: VAF measures for the models.

Model	Training	Testing
GWR	82.06	72.57
FCBLS	91.30	88.73

In addition to performance indices, the estimates can be presented in map form. Figs. 6-8 show the observed values, GWR estimations and FCBLS estimations, respectively. Based on the performance evaluations and the maps, it can be argued that the fuzzy clustering based regression model (FCBLS) outperforms the geographically weighted regression model in this case.

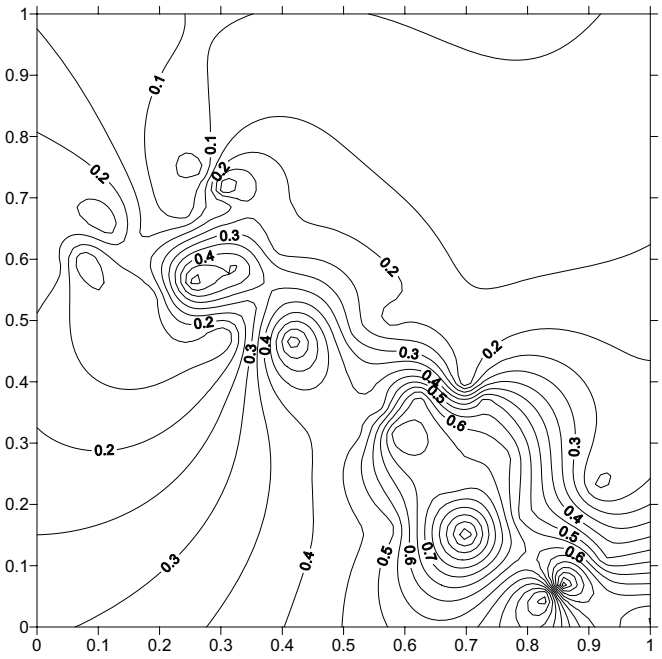


Figure 6: Contour map for observed data.

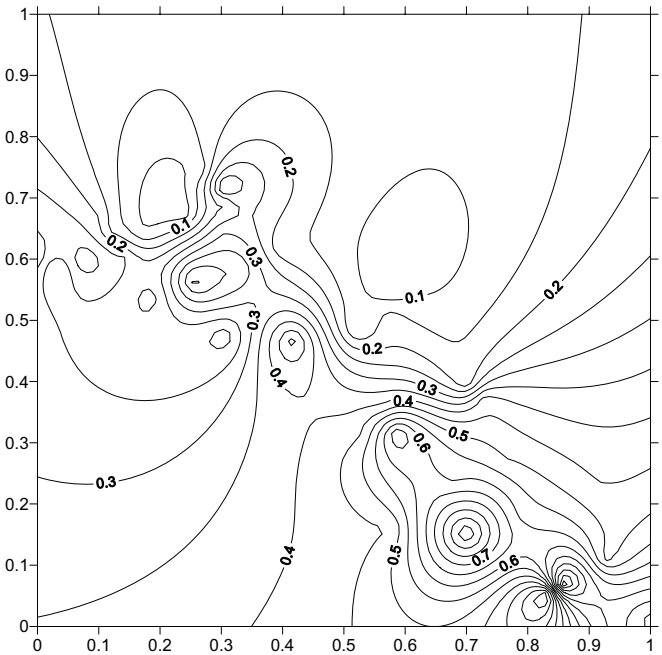


Figure 7: Contour map for the GWR model.

ISBN: 978-989-95079-6-8

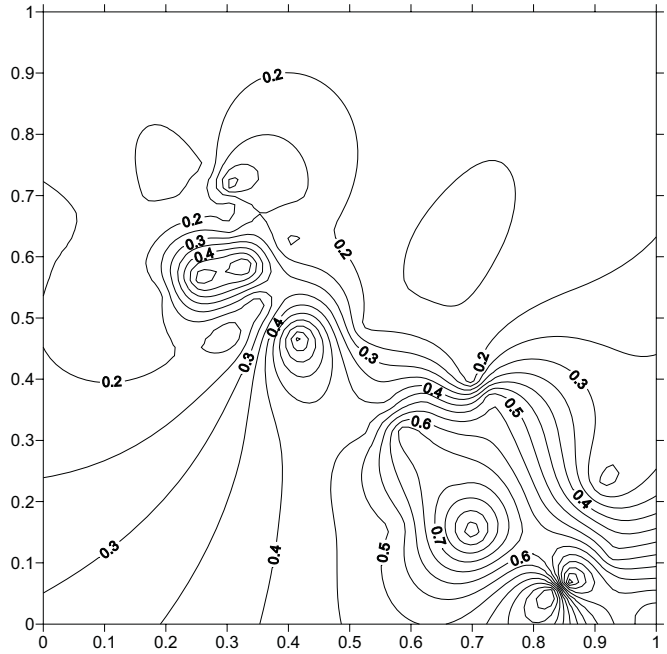


Figure 8: Contour map for FCBLS model.

Acknowledgements

This paper is part of the project (108M393) which is granted by Scientific and Technological Research Council of Turkey (TUBITAK).

References

[1] X. Gao, Y. Asami, CJ. F. Chung. An empirical evaluation of spatial regression models, *Computers &Geosciences*, 32:1040-1051,2006.

[2] C.D. Lloyd, *Local Models for Spatial Analysis*, Boca Raton: CRC Press, 2006.

[3] A.S. Fotheringham, M.E. Charlton, C. Brunsdon, Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis, *Environment and Planning A*, 30: 1905-1927, 1998.

[4] E. S. Lee, Neuro-Fuzzy estimation in spatial statistics, *J Mathematical Analysis and Applications*, 249:221-231, 2000.

[5] P. Wong, F. Aminzadeh, M. Nikravesh, *Soft Computing for Reservoir Characterization and Modeling*, Heidelberg:Physica-Verlag, 2001.

[6] B. Tutmez, A.E. Tercan, U. Kaymak. Fuzzy modeling for reserve estimation based on spatial variability, *Mathematical Geology*, 39(1):87-111, 2007.

[7] A.S. Fotheringham, C. Brunsdon, M.E. Charlton, *Qualitative Geography: Perspectives on Spatial Data Analysis*, London: SAGE Publications, 2000.

[8] C.D. Lloyd and I. Shuttleworth, Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning, *Environment and Planning A*, 37:81-113, 2005.

[9] A.S. Fotheringham, C. Brunsdon, M.E. Charlton, *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*, Chichester:Wiley, 2002.

[10] R. Babuska, *Fuzzy Modeling for Control*, USA:Kluwer Academic, 1998.

[11] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy c-means clustering algorithm, Computers&Geosciences, 10(2-3):191-203, 1984.

[12] H. Akaike, Information theory and an extension of the maximum likelihood principle, In B. Petrov, F. Csaki (eds) *2nd Symposium on Information Theory*, Budapest, 267-281, 1973.

[13] Thiago F. L.V.B. Rangel, Jose A.F. Diniz-Filho, L.M. Bini, Towards and integrated computational tool for spatial analysis in macroecology and biogeography, *Global Ecology and Biogeography*, 15:321-327, 2006.

[14] R.S. Bivand, E.J. Pebesma, V. Gomez-Rubio, Applied Spatial Data Analysis with R, New York: Springer, 2008.

Fuzzy Integrals over Complete Residuated Lattices

Antonín Dvořák, Michal Holčapek

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava,
30. dubna 22, 701 03 Ostrava 1, Czech Republic
Email: {antonin.dvorak, michal.holcapek}@osu.cz

Abstract— *The aim of this paper is to introduce two new types of fuzzy integrals, namely, $\otimes$-fuzzy integral and $\rightarrow$-fuzzy integral, where $\otimes$ and $\rightarrow$ are the multiplication and residuum of a complete residuated lattice, respectively. The first integral is based on a fuzzy measure of* **L***-fuzzy sets and the second one on a complementary fuzzy measure of* **L***-fuzzy sets, where* **L** *is a complete residuated lattice. Some of their properties and a relation to the fuzzy (Sugeno) integral are investigated.*

Keywords— fuzzy measure, fuzzy integral, fuzzy quantifier.

1 Introduction

Let us consider two time series $t_1 = (t_{1k})_{k\in T}$, $t_2 = (t_{2k})_{k\in T}$ displayed on Fig. 1 and suppose that our goal is to compare them and to find the "better" time series, where a greater value of time series at some time point k means a better value. To

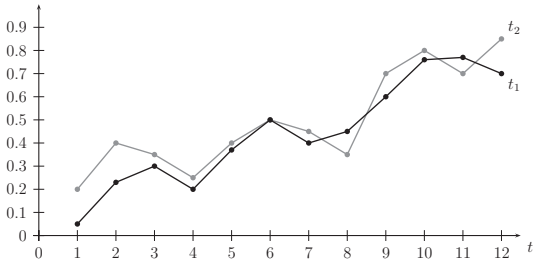


Figure 1: Time series.

solve this task it is reasonable, firstly, to determine degrees of truth saying how much formulas

$$\varphi(i,j,k) := \text{"}\textit{the value } t_{ik} \textit{ is better than the value } t_{jk}\text{"}$$

are true, where, first, $i = 1$ and $j = 2$, and then $i = 2$ and $j = 1$. Obviously, the degrees of truth of formula $\varphi(i,j,k)$ may be modeled by a fuzzy relation $R : [0,1]^2 \to [0,1]$, e.g.

$$R(a,b) = \max(a - b, 0), \tag{1}$$

where $R(a,b)$ determines the degree that a is better than b. Note that $R(a,b) = \neg(a \to b)$, where $a \to b$ is the operation of residuum and $\neg$ is the operation of negation in the Łukasiewicz algebra (see Example 2.1 and consider $\neg a = 1 - a$). If we know how one value of time series is better than the other one for each time k, then we have to solve a task how to aggregate the obtained values to find a degree in which one times series is better than the second one. One of the natural approaches could be to evaluate the following formula of the second order logic

$$\varphi(i,j) := (\exists Y \in \mathcal{P}(T) \setminus \{\emptyset\})(\forall k \in Y)(\varphi(i,j,k) \& \psi(Y)),$$

where $\mathcal{P}(T)$ is the power set of T and

$$\psi(Y) := \textit{"the set } Y \textit{ is a big subset of } T\textit{"}.$$

Note that formula $\varphi(i,j)$ may be defined using the formula $(Qk \in T)\varphi(i,j,k)$ with a fuzzy quantifier Q like *for nearly all* or *many* etc. Some types of fuzzy quantifiers could be determined by the interpretations of the formula ψ (cf. [1] and also see [2]). For example, if we consider the Łukasiewicz algebra as the structure of truth values for our logic, the truth value of the formula $\varphi(i,j,k)$ is defined as $R(t_{ik}, t_{jk})$ from (1), i.e., as the degree that t_{ik} is better than t_{jk}, and the truth value of the formula $\psi(Y)$ is interpreted by the value $\mu(Y)$, where $\mu : \mathcal{P}(T) \to [0,1]$ is a fuzzy measure (see Definition 3.2), then the evaluation of the formula $\varphi(i,j)$ is given by

$$||\varphi(i,j)|| = \bigvee_{Y\in\mathcal{P}(T)\setminus\{\emptyset\}} \bigwedge_{k\in Y} (R(t_{ik}, t_{jk}) \otimes \mu(Y)), \tag{2}$$

where $\otimes$ is the operation of multiplication in the Łukasiewicz algebra (see Example 2.1) interpreting the logical connective $\&$. Finally, we can conclude that, for example, the time series t_1 is better than t_2, if $||\varphi(1,2)|| > ||\varphi(2,1)||$.

Let us define a mapping $I_\mu : \mathcal{F}(T) \to [0,1]$ by

$$I_\mu(A) = \bigvee_{Y\in\mathcal{P}(T)\setminus\{\emptyset\}} \bigwedge_{k\in Y} (A(k) \otimes \mu(Y)), \tag{3}$$

where $\mathcal{F}(T)$ denotes the set of all fuzzy sets over T and $\mu : \mathcal{P}(T) \to [0,1]$ is a fuzzy measure. One could simply verify that if $c \in [0,1]$ is a constant and $A(k) = c$ for any $k \in T$, then $I_\mu(A) = c$, and if $A(k) \le B(k)$, then $I_\mu(A) \le I_\mu(B)$. Hence, I_μ is a fuzzy measure (in the sense of Definition 3.2) which could be understood, according to Mesiar [3], as an example of (fuzzy) integral. Putting $A(k) = R(t_{ik}, t_{jk})$ for any $k \in T$, we can write $||\varphi(i,j)|| = I_\mu(A)$ or also $||(Qk \in T)\varphi(i,j,k)|| = I_\mu(A)$ and, hence, we can see that fuzzy integrals may be used to model some types of fuzzy quantifiers. This idea is not new (see e.g. [4] or [5]), but some disadvantage of proposed approaches is that fuzzy quantifiers are defined as mappings from the set of all measurable (fuzzy) sets over a set M to $[0,1]$, although, the classical definition introduces fuzzy quantifiers as mappings from the set of all (fuzzy) sets over a set M to $[0,1]$ (see e.g. [6] or [7]). It will be clear that this drawback vanishes when fuzzy quantifiers are modeled by fuzzy integrals in a form similar to (3).

The aim of this contribution is to generalize the fuzzy integral defined in (3), namely, to introduce a $\otimes$-fuzzy integral that could be used for modeling fuzzy quantifiers like *all*, *some*,

ISBN: 978-989-95079-6-8

for nearly all, or *many*, etc. and then to introduce a $\rightarrow$-fuzzy integral that could be used for modeling quantifiers like *no*, *not all*, etc. Both types of fuzzy integrals are defined over a complete residuated lattice and it could be shown than, in general, one type of fuzzy integral cannot be expressed by another one. Nevertheless, if the structure of truth values is complete MV-algebra, then it is possible to define the $\rightarrow$-fuzzy integral from the $\otimes$-fuzzy integral using negation (see Theorem 4.14). Moreover, it is surprising that we are able to show that the well-known Sugeno integral [8] is, under certain conditions, a special case of our fuzzy integral (see Theorem 4.7).

2 Preliminaries

2.1 Structures of truth values

Let us suppose that the structure of truth values is a *complete residuated lattice* (see e.g. [9]), i.e., an algebra $\mathbf{L} = \langle L, \wedge, \vee, \rightarrow, \otimes, \bot, \top \rangle$ with four binary operations and two constants such that $\langle L, \wedge, \vee, \bot, \top \rangle$ is a complete lattice, where $\bot$ is the least element and $\top$ is the greatest element of L, respectively, $\langle L, \otimes, \top \rangle$ is a commutative monoid (i.e., $\otimes$ is associative, commutative and the identity $a \otimes \top = a$ holds for any $a \in L$) and the adjointness property is satisfied, i.e.,

$$a \leq b \rightarrow c \quad \text{iff} \quad a \otimes b \leq c \tag{4}$$

holds for each $a, b, c \in L$, where $\leq$ denotes the corresponding lattice ordering. The operations $\otimes$ and $\rightarrow$ are usually called the multiplication and residuum, respectively. A residuated lattice is *divisible*, if $a \otimes (a \rightarrow b) = a \wedge b$ holds for arbitrary $a, b \in L$, and satisfies the *law of double negation*, if $(a \rightarrow \bot) \rightarrow \bot = a$ holds for any $a \in L$. A divisible residuated lattice satisfying the law of double negation is called an *MV-algebra*. For other information about residuated lattices we refer to [9].

Example 2.1. It is easy to prove (see e.g. [10]) that the algebra

$$\mathbf{L}_{\mathrm{T}} = \langle [0,1], \min, \max, T, \rightarrow_T, 0, 1 \rangle,$$

where T is a left continuous t-norm and $a \rightarrow_T b = \bigvee \{c \in [0,1] \mid T(a,c) \leq b\}$ defines the residuum, is a complete residuated lattice. Moreover, if T is the Łukasiewicz t-norm, i.e., $T(a,b) = \max(a+b-1, 0)$ for all $a, b \in [0,1]$, then $\mathbf{L}_{\mathrm{T}}$ is a complete MV-algebra called a *Łukasiewicz algebra* (on $[0,1]$). One checks easily that $a \rightarrow_T b = \max(1-a+b, 0)$ is the residuum in the Łukasiewicz algebra.

Example 2.2. One checks easily that

$$\mathbf{L}_{[0,\infty]} = \langle [0,\infty], \min, \max, \rightarrow, 0, \infty \rangle,$$

where $\otimes = \min$ and

$$a \rightarrow b = \begin{cases} b, & \text{if } b < a, \\ \infty, & \text{otherwise,} \end{cases} \tag{5}$$

is a complete residuated lattice. Note that $\mathbf{L}_{[0,\infty]}$ is a special example of more general residuated lattice called a *Heyting algebra*.[1]

Let us define the following additional operations for all $a, b \in L$:

$$a \leftrightarrow b = (a \rightarrow b) \wedge (b \rightarrow a) \qquad (\textit{biresiduum})$$
$$\neg a = a \rightarrow \bot. \qquad (\textit{negation})$$

[1] A Heyting algebra is a residuated lattice with $\otimes = \wedge$.

2.2 L-fuzzy sets

Let $\mathbf{L}$ be a complete residuated lattice and M be a universe of discourse. A mapping $A : M \rightarrow L$ is called an $\mathbf{L}$-*fuzzy set on* M. A value $A(m)$ is called a *membership degree of* m *in the* $\mathbf{L}$-*fuzzy set* A. The set of all $\mathbf{L}$-fuzzy sets on M is denoted by $\mathcal{F}_{\mathbf{L}}(M)$. An $\mathbf{L}$-fuzzy set A on M is called *crisp*, if there is a subset X of M such that $A = 1_X$, where 1_X denotes the characteristic function of X. Particularly, $1_\emptyset$ denotes the empty $\mathbf{L}$-fuzzy set on M, i.e. $1_\emptyset(m) = \bot$ for any $m \in M$. The set of all crisp $\mathbf{L}$-fuzzy sets on M is denoted by $\mathcal{P}_{\mathbf{L}}(M)$. An $\mathbf{L}$-fuzzy set A is *constant*, if there is $c \in L$ such that $A(m) = c$ for any $m \in M$. For simplicity, a constant $\mathbf{L}$-fuzzy set is denoted by the corresponding element of L, e.g., a, b, c.[2] Let us denote $\mathrm{Supp}(A) = \{m \mid m \in M \;\&\; A(m) > \bot\}$ and $\mathrm{core}(A) = \{m \mid m \in M \;\&\; A(m) = \top\}$ the *support* and *core* of an $\mathbf{L}$-fuzzy set A, respectively. Obviously, $\mathrm{Supp}(1_X) = \mathrm{core}(1_X) = X$ for any crisp $\mathbf{L}$-fuzzy set. An $\mathbf{L}$-fuzzy set A is called *normal*, if $\mathrm{core}(A) \neq \emptyset$.

Let $\{A_i \mid i \in I\}$ be a non-empty family of $\mathbf{L}$-fuzzy sets on M. Then the *union of* A_i is defined by

$$\left(\bigcup_{i \in I} A_i\right)(m) = \bigvee_{i \in I} A_i(m) \tag{6}$$

for any $m \in M$ and the *intersection of* A_i is defined by

$$\left(\bigcap_{i \in I} A_i\right)(m) = \bigwedge_{i \in I} A_i(m) \tag{7}$$

for any $m \in M$. Let A be an $\mathbf{L}$-fuzzy set on M. The *complement* of A is an $\mathbf{L}$-fuzzy set $\overline{A}$ on M defined by $\overline{A}(m) = \neg A(m)$ for any $m \in M$. Finally, an extension of the operations $\otimes$ and $\rightarrow$ on L to the operations on $\mathcal{F}_{\mathbf{L}}(M)$ is given by

$$(A \otimes B)(m) = A(m) \otimes B(m) \tag{8}$$
$$(A \rightarrow B)(m) = A(m) \rightarrow B(m) \tag{9}$$

for any $A, B \in \mathcal{F}_{\mathbf{L}}(M)$ and $m \in M$, respectively. The following theorem shows the well-known relation between the operations of the union and intersection of sets which also holds for $\mathbf{L}$-fuzzy sets, if we restrict ourselves to a special class of complete residuated lattices.

Theorem 2.1. *Let* $\mathbf{L}$ *be a complete residuated lattice satisfying the law of double negation and* $\{A_i \mid i \in I\}$ *be a non-empty family of* $\mathbf{L}$*-fuzzy sets on* M*. Then*

$$\bigcup_{i \in I} A_i = \overline{\bigcap_{i \in I} \overline{A_i}} \quad \textit{and} \quad \bigcap_{i \in I} A_i = \overline{\bigcup_{i \in I} \overline{A_i}}. \tag{10}$$

We say that an $\mathbf{L}$-fuzzy set A is *less than or equal to* an $\mathbf{L}$-fuzzy set B and denote by $A \subseteq B$, if, for any $m \in M$, we have $A(m) \leq B(m)$. Let $f : M \rightarrow M'$ be a mapping. Then $f^{\rightarrow}(A)(m) = \bigvee_{m' \in f^{-1}(m)} A(m')$ defines a mapping $f^{\rightarrow} : \mathcal{F}_{\mathbf{L}}(M) \rightarrow \mathcal{F}_{\mathbf{L}}(M')$. Obviously, if f is a bijective mapping, then $f^{\rightarrow}(A)(f(m)) = A(m)$ for any $m \in M$.

[2] We suppose that the meaning of this symbol will be unmistakable from the context, that is, it should be clear when an element of L is considered and when a constant $\mathbf{L}$-fuzzy set is assumed.

ISBN: 978-989-95079-6-8

3 Fuzzy measures

In this section, we will introduce a notion of fuzzy measure and complementary fuzzy measure of **L**-fuzzy sets. More information about fuzzy measures could be found in [11, 12].

For our considerations we will consider algebras of **L**-fuzzy sets as a base for defining fuzzy measures of **L**-fuzzy sets.

Definition 3.1 ([11])**.** Let M be a non-empty universe of discourse. A subset $\mathcal{M}$ of $\mathcal{F}_{\mathbf{L}}(M)$ is an *algebra of* **L***-fuzzy sets on* M, if the following conditions are satisfied

(i) $1_\emptyset, 1_M \in \mathcal{M}$,

(ii) if $A \in \mathcal{M}$, then $\overline{A} \in \mathcal{M}$,

(iii) if $A, B \in \mathcal{M}$, then $A \cup B \in \mathcal{M}$.

A couple $(M, \mathcal{M})$ is called a *fuzzy measurable space*, if $\mathcal{M}$ is an algebra of **L**-fuzzy sets on M.

Example 3.1. The sets $\{1_\emptyset, 1_M\}$, $\mathcal{P}_{\mathbf{L}}(M)$, σ-algebras on M, or $\mathcal{F}_{\mathbf{L}}(M)$ are algebras of **L**-fuzzy sets on M.

Example 3.2. Let us say that an **L**-fuzzy set A on M is a *simple* **L***-fuzzy set* on M, if there exists a family of sets $\{M_i \mid i = 1, \ldots, n\}$ such that $\bigcup_{i=1}^n M_i = M$, $M_i \neq M_j$ for any $i \neq j$ and $A(m) = A(m')$ holds for each $m, m' \in M_i$, where $i = 1, \ldots, n$. Obviously, the set of all simple **L**-fuzzy sets on M is an algebra of **L**-fuzzy sets on M.

Example 3.3. Let **L** be the Łukasiewicz algebra on $[0, 1]$ (see Example 2.1) and $M = [0, 1]$. Then the set of all continuous mappings $A : [0, 1] \to [0, 1]$ is an algebra of **L**-fuzzy sets in M.[3]

Let us introduce the concepts of fuzzy measure and complementary fuzzy measure as follows. The first definition is a modification of the definition of a normed fuzzy measure with respect to truth values (see e.g. [12, 13]).

Definition 3.2. Let $(M, \mathcal{M})$ be a fuzzy measurable space. A mapping $\mu : \mathcal{M} \to L$ is called a *fuzzy measure* on $(M, \mathcal{M})$, if

(i) $\mu(1_\emptyset) = \bot$ and $\mu(1_M) = \top$,

(ii) if $A, B \in \mathcal{M}$ such that $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

A triplet $(M, \mathcal{M}, \mu)$ is called a *fuzzy measure space*, if $(M, \mathcal{M})$ is a fuzzy measurable space and μ is a fuzzy measure on $(M, \mathcal{M})$.

Definition 3.3. Let $(M, \mathcal{M})$ be a fuzzy measurable space. A mapping $\nu : \mathcal{M} \to L$ is called a *complementary fuzzy measure* on $(M, \mathcal{M})$, if

(i) $\nu(1_\emptyset) = \top$ and $\nu(1_M) = \bot$,

(ii) if $A, B \in \mathcal{M}$ such that $A \leq B$, then $\nu(A) \geq \nu(B)$.

A triplet $(M, \mathcal{M}, \nu)$ is called a *complementary fuzzy measure space*, if $(M, \mathcal{M})$ is a fuzzy measurable space and ν is a complementary fuzzy measure on $(M, \mathcal{M})$.

[3]Note that the set of all continuous mappings need not be an algebra of **L**-fuzzy sets for other residuated lattices determined by left continuous T-norms, because the negation is not a continuous mapping in general.

Example 3.4. Let $(M, \mathcal{M})$ be the fuzzy measurable space of all continuous mappings from Example 3.3. It is easy to see that

$$\mu(A) = \int_0^1 A(m)\, dm,$$

where $\int_0^1 A(m)\, dm$ denotes the Riemann integral, defines a fuzzy measure on $(M, \mathcal{M})$.

Example 3.5. Let **L** be a complete residuated lattice with the support $[0, 1]$ and $\mathbb{N}$ be the set of natural numbers with 0. For any non-empty countable (i.e., finite or denumerable) universe M, injective mapping $f : M \to \mathbb{N}$, $n \in \mathbb{N}$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$, denote

$$A_{f,n}(m) = \begin{cases} A(m), & \text{if } f(m) \leq n; \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Further, for any injective mapping $f : M \to \mathbb{N}$ and $n \in \mathbb{N}$, define $\mu_{f,n} : \mathcal{F}_{\mathbf{L}}(M) \to [0, 1]$ as follows

$$\mu_{f,n}(A) = \frac{\sum_{m \in \mathrm{Supp}(A_{f,n})} A_{f,n}(m)}{|\mathrm{Supp}(1_{M_{f,n}})|} \tag{12}$$

and, finally, define $\underline{\mu}_f, \overline{\mu}_f : \mathcal{F}_{\mathbf{L}}(M) \to [0, 1]$ as follows

$$\underline{\mu}_f = \liminf_{n \to \infty} \mu_{f,n}(A), \tag{13}$$

$$\overline{\mu}_f = \limsup_{n \to \infty} \mu_{f,n}(A). \tag{14}$$

It is easy to see that $\mu_{f,n}$, $\underline{\mu}_f$ and $\overline{\mu}_f$ are fuzzy measures on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by an injective mapping f.[4] If, for example, $M = \mathbb{N}$ and $f = \mathrm{id}$, then $\underline{\mu}_f(A) = \overline{\mu}_f(A) = \bot$ for any **L**-fuzzy set with finite universe. For the set of all even or odd numbers, both fuzzy measures give $\frac{1}{2}$ and, for the set of all prime numbers, we obtain 0.

If M is finite, then $\underline{\mu}_f = \underline{\mu}_g = \overline{\mu}_f = \overline{\mu}_g$ for any injective mappings $f, g : M \to \mathbb{N}$ and

$$\underline{\mu}_f(A) = \overline{\mu}_f(A) = \frac{\sum_{m \in M} A(m)}{|M|}. \tag{15}$$

Hence, it is easy to see that $\mu_f(A) = \mu_f(h^{\rightarrow}(A))$ holds for any non-empty finite universe M, $A \in \mathcal{F}_{\mathbf{L}}(M)$, injective mapping $f : M \to \mathbb{N}$ and bijective mapping $h : M \to M$. Unfortunately, this equality fails for denumerable universes in general. In fact, consider $M = \mathbb{N}$, $f = \mathrm{id}$ and a bijective mapping $h : \mathbb{N} \to \mathbb{N}$ such that the image of all even numbers is the set of prime numbers. Then both fuzzy measures give $\frac{1}{2}$ for the set of all even numbers, however, 0 for the set of all prime numbers.

Example 3.6. Let μ_f be one of the fuzzy measures on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by f defined in (13) and (14). If $h : [0, 1] \to [0, 1]$ is a non-decreasing mapping with $h(0) = 0$ and $h(1) = 1$, then $h \circ \mu_f$ is a fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by μ_f and h. If $h : [0, 1] \to [0, 1]$ is a non-increasing mapping with $h(0) = 1$ and $h(1) = 0$, then $h \circ \mu_f$ is a complementary fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by μ_f and h.

[4]Note that $\underline{\mu}_f$ and $\overline{\mu}_f$ could be understood as a generalization of lower and upper weighted densities well known in the number theory which are examples of so-called lower and upper asymptotic fuzzy measures (see [14]).

ISBN: 978-989-95079-6-8

Theorem 3.1. *Let $(M, \mathcal{M})$ be a fuzzy measurable space. If μ (ν) is a fuzzy measure (a complementary fuzzy measure) on $(M, \mathcal{M})$, then $\nu'(A) = \neg\mu(A)$ ($\mu'(A) = \neg\nu(A)$) defines a complementary fuzzy measure (a fuzzy measure) on $(M, \mathcal{M})$.*

Definition 3.4. Let $(M, \mathcal{M})$ be a fuzzy measurable space and $X \in \mathcal{F}_{\mathbf{L}}(M)$. We say that X is $\mathcal{M}$*-fuzzy measurable*, if $X \in \mathcal{M}$.

Let $(M, \mathcal{M})$ be a fuzzy measurable space and $X \in \mathcal{F}_{\mathbf{L}}(M)$. Denote $\mathcal{M}_X$ the set of all $\mathcal{M}$-fuzzy measurable sets which are contained in X, i.e.,

$$\mathcal{M}_X = \{A \mid A \in \mathcal{M} \text{ and } A \subseteq X\}. \tag{16}$$

Note that $1_\emptyset \in \mathcal{M}_X$ for each $X \in \mathcal{F}_{\mathbf{L}}(M)$ and if X is $\mathcal{M}$-fuzzy measurable set, then also $X \in \mathcal{M}_X$. If $X = M$, then we will write only $\mathcal{M}$ instead of $\mathcal{M}_M$.

Theorem 3.2. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. A mapping $\mu^* : \mathcal{F}_{\mathbf{L}}(M) \to L$ defined by*

$$\mu^*(X) = \bigvee_{A \in \mathcal{M}_X} \mu(A) \tag{17}$$

is a fuzzy measure on the fuzzy measurable space $(M, \mathcal{F}_{\mathbf{L}}(M))$. We say that μ^* is the inner fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by μ.

Example 3.7. Let $(M, \mathcal{P}_{\mathbf{L}}(M), \mu)$ be an arbitrary fuzzy measurable space (recall that $\mathcal{P}_{\mathbf{L}}(M)$ is the power set of M). Then the inner fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ is defined by

$$\mu^*(A) = \begin{cases} \mu(A'), & \text{if } 1_{\mathrm{core}(A)} = A', \\ \bot, & \text{otherwise.} \end{cases} \tag{18}$$

Thus all $\mathbf{L}$-fuzzy sets that are not normal have the inner fuzzy measure equal to $\bot$.

Example 3.8. Let $\mathbf{L}$ be the Łukasiewicz algebra on $[0, 1]$, $(M, \mathcal{M}, \mu)$ be the fuzzy measure space of continuous $\mathbf{L}$-fuzzy sets from Example 3.4. Then, for example, we have $\mu^*(1_{[a,b]}) = b - a$, however, $1_{[a,b]} \notin \mathcal{M}$.

Theorem 3.3. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space. A mapping $\nu^* : \mathcal{F}_{\mathbf{L}}(M) \to L$ defined by*

$$\nu^*(X) = \bigwedge_{A \in \mathcal{M}_X} \nu(A) \tag{19}$$

is a complementary fuzzy measure on the fuzzy measurable space $(M, \mathcal{F}_{\mathbf{L}}(M))$. We say that ν is the inner complementary fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by ν.

In the following part we will define an isomorphism between fuzzy measure spaces and then between complementary fuzzy measure spaces.

Definition 3.5. Let $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ be fuzzy measurable spaces. We say that a mapping $g : \mathcal{M} \to \mathcal{M}'$ is an *isomorphism between $(M, \mathcal{M})$ and $(M', \mathcal{M}')$*, if

(i) g is a bijective mapping with $g(1_\emptyset) = 1_\emptyset$,

(ii) $g(A \cup B) = g(A) \cup g(B)$ and $g(\overline{A}) = \overline{g(A)}$ hold for any $A, B \in \mathcal{M}$,

(iii) there exists a bijective mapping $f : M \to M'$ with $A(m) = g(A)(f(m))$ for any $A \in \mathcal{M}$ and $m \in M$.

Theorem 3.4. *Let $(M, \mathcal{M})$, $(M', \mathcal{M}')$ be fuzzy measurable spaces and $g : \mathcal{M} \to \mathcal{M}'$ be a surjective mapping. Then g is an isomorphism between $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ if and only if there exists a bijective mapping $f : M \to M'$ such that $g = f^{\to}$.*

Definition 3.6. Let $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ be fuzzy measurable spaces. We say that a mapping $g : \mathcal{M} \to \mathcal{M}'$ is an *isomorphism between $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$* (or *between $(M, \mathcal{M}, \nu)$ and $(M', \mathcal{M}', \nu')$*), if

(i) g is an isomorphism between $(M, \mathcal{M})$ and $(M', \mathcal{M}')$,

(ii) $\mu(A) = \mu'(g(A))$ (or $\nu(A) = \nu'(g(A))$) for any $A \in \mathcal{M}$.

If g is an isomorphism between fuzzy measure spaces $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$ or between complementary fuzzy measure spaces $(M, \mathcal{M}, \nu)$ and $(M', \mathcal{M}', \nu')$, then we write $g(M, \mathcal{M}, \mu) = (M', \mathcal{M}', \mu')$ or $g(M, \mathcal{M}, \nu) = (M', \mathcal{M}', \nu')$, respectively.

Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. If $f : M \to M'$ is a bijective mapping, then $(M', f^{\to}(\mathcal{M}), \mu_{f^{\to}})$, where

$$\mu_{f^{\to}}(f^{\to}(A)) = \mu(A) \tag{20}$$

holds for any $A \in \mathcal{M}$, is a fuzzy measure space isomorphic to $(M, \mathcal{M}, \mu)$. A simple consequence of Theorem 3.4 is the fact that each fuzzy measure space $(M', \mathcal{M}', \mu')$ isomorphic to $(M, \mathcal{M}, \mu)$ has the form $(M', f^{\to}(\mathcal{M}), \mu_{f^{\to}})$ for a suitable bijective mapping $f : M \to M'$. Analogously, to every couple of isomorphic complementary fuzzy measure spaces $(M, \mathcal{M}, \nu)$ and $(M', \mathcal{M}', \nu')$ there is a bijective mapping $f : M \to M'$ such that $(M', f^{\to}(\mathcal{M}), \nu_{f^{\to}}) = (M', \mathcal{M}', \nu')$.

4 Fuzzy integrals

In this section, we will introduce two types of fuzzy integrals. The first one is a generalization of the formula (3) derived in Introduction. For more information about fuzzy integrals we refer to [11, 12].

4.1 $\otimes$-fuzzy integral

In this part, we will introduce a type of fuzzy integral that can be defined on an arbitrary fuzzy measure space $(M, \mathcal{M}, \mu)$. The form of this integral is motivated by our need to describe a class of models of $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1 \rangle$. In [2], we show that this class of models is bounded by the models of determiners *all* and *some*. Note that models of *all* and *some* are the same as the interpretations of quantifiers $\forall$ and $\exists$, respectively, in fuzzy logic (see e.g. [6, 7, 15, 4]).

Definition 4.1. Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and X be a $\mathcal{M}$-fuzzy measurable $\mathbf{L}$-fuzzy set. The $\otimes$*-fuzzy integral of A on X* is given by

$$\int_X^{\otimes} A \, d\mu = \bigvee_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in \mathrm{Supp}(Y)} (A(m) \otimes \mu(Y)). \tag{21}$$

If $X = 1_M$, then we write $\int^{\otimes} A \, d\mu$.

ISBN: 978-989-95079-6-8

Remark 4.1. It is easy to see that $\int^{\otimes}_{1_\emptyset} A\, d\mu = \bigvee \emptyset = \bot$ for any $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $\int^{\otimes}_X A\, d\mu \leq \int^{\otimes}_Y A\, d\mu$, whenever $X \subseteq Y$. Since $\int^{\otimes}_{1_M} A\, d\mu \neq \top$ in general, $\mu_A(X) = \int^{\otimes}_X A\, d\mu$ does not define a fuzzy measure on $(M, \mathcal{M})$ in the sense of Definition 3.2.

Remark 4.2. One can also define a $\wedge$-fuzzy integral of A on X in such way that $\otimes$ is replaced by $\wedge$ in (21). Since $\otimes$ and $\wedge$ have many common properties, both types of fuzzy integral will have similar properties. Nevertheless, we prefer the $\otimes$-fuzzy integral in this paper, because it is closely related (due to the adjointness property) to $\rightarrow$-fuzzy integral that will be introduced in the following subsection.

Theorem 4.1. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. Then $\mu' : \mathcal{F}_{\mathbf{L}}(M) \rightarrow L$ defined by*

$$\mu'(A) = \int^{\otimes} A\, d\mu \tag{22}$$

is a fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$.

Theorem 4.2. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. Then*

(i) $\int^{\otimes}_X (A \cap B)\, d\mu \leq \int^{\otimes}_X A\, d\mu \wedge \int^{\otimes}_X B\, d\mu$,

(ii) $\int^{\otimes}_X (A \cup B)\, d\mu \geq \int^{\otimes}_X A\, d\mu \vee \int^{\otimes}_X B\, d\mu$,

(iii) $\int^{\otimes}_X (c \otimes A)\, d\mu \geq c \otimes \int^{\otimes}_X A\, d\mu$,

(iv) $\int^{\otimes}_X (c \rightarrow A)\, d\mu \leq c \rightarrow \int^{\otimes}_X A\, d\mu$,

hold for any $X \in \mathcal{M}$, $A, B \in \mathcal{F}_{\mathbf{L}}(M)$ and $c \in L$.

Theorem 4.3. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space and $c \in L$. Then we have*

(i) $\int^{\otimes} (c \otimes 1_X)\, d\mu = c \otimes \mu^*(1_X)$ *for any $X \subseteq M$,*

(ii) $\int^{\otimes} (c \otimes 1_X)\, d\mu = c \otimes \mu(1_X)$ *for any $X \subseteq M$ such that $1_X \in \mathcal{M}$,*

(iii) $\int^{\otimes} 1_X\, d\mu = \mu(1_X)$ *for any $X \subseteq M$ such that $1_X \in \mathcal{M}$,*

(iv) $\int^{\otimes} c\, d\mu = c$.

Theorem 4.4. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. If $X \in \mathcal{M}$ is such that $1_{\mathrm{Supp}(Y)} \in \mathcal{M}_X$ for any $Y \in \mathcal{M}_X$, then, for any $A \in \mathcal{F}_{\mathbf{L}}(M)$, we have*

$$\int^{\otimes}_X A\, d\mu = \bigvee_{1_Y \in \mathcal{P}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in Y} (A(m) \otimes \mu(1_Y)), \tag{23}$$

where $\mathcal{P}_X = \{1_{\mathrm{Supp}(Z)} \mid Z \in \mathcal{M}_X\}$.

Theorem 4.5. *Let $\mathbf{L}$ be a complete MV-algebra, $(M, \mathcal{M}, \mu)$ be a fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $X \in \mathcal{M}$. Then*

$$\int^{\otimes}_X A\, d\mu = \bigvee_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \big(\mu(Y) \otimes \bigwedge_{m \in \mathrm{Supp}(Y)} A(m)\big). \tag{24}$$

Moreover,

$$\int^{\otimes}_X (c \otimes A)\, d\mu = c \otimes \int^{\otimes}_X A\, d\mu \tag{25}$$

for any $c \in L$.

Theorem 4.6. *Let g be an isomorphism between fuzzy measure spaces $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$ and $X \in \mathcal{M}$. Then we have*

$$\int^{\otimes}_X A\, d\mu = \int^{\otimes}_{g(X)} g(A)\, d\mu' \tag{26}$$

for any $A \in \mathcal{F}_{\mathbf{L}}(M)$.

In the end of this part, we will show that the Sugeno integral is a special case of our proposed integral. For this purpose we will use a slight modification of the usual Sugeno integral definition with respect to the fuzzy measurable spaces over complete residuated lattices.

Let $\mathbf{L}$ be a complete residuated lattice and $(M, \mathcal{M})$ be a fuzzy measurable space such that $A \cap B \in \mathcal{M}$ for any $A, B \in \mathcal{M}$.[5] Denote $A_a = \{m \mid m \in M \;\&\; A(m) \geq a\}$. We say that an $\mathbf{L}$-fuzzy set A is *$\mathcal{M}$-Sugeno measurable*, if $1_{A_a} \in \mathcal{M}$ for any $a \in L$. The *Sugeno integral* is given, for any fuzzy measure space $(M, \mathcal{M}, \mu)$ with $B \cap C \in \mathcal{M}$ for any $B, C \in \mathcal{M}$, for any $\mathcal{M}$-Sugeno measurable $\mathbf{L}$-fuzzy set A and for any $X \in \mathcal{M}$, by

$$\int_X A\, d\mu = \bigvee_{a \in L} (a \wedge \mu(1_{A_a} \cap X)). \tag{27}$$

Theorem 4.7. *Let $\mathbf{L}$ be a complete Heyting algebra, $(M, \mathcal{M}, \mu)$ be a fuzzy measure space with $B \cap C \in \mathcal{M}$ for any $B, C \in \mathcal{M}$, A be a $\mathcal{M}$-Sugeno measurable $\mathbf{L}$-fuzzy set and $X \in \mathcal{M}$. Then $\int_X A\, d\mu = \int^{\otimes}_X A\, d\mu$.*

4.2 $\rightarrow$-fuzzy integral

In this part, we will introduce another type of fuzzy integral that can be defined on an arbitrary complementary fuzzy measure space $(M, \mathcal{M}, \nu)$. The form of this integral is motivated by our need to describe another class of models of $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1 \rangle$ which are kind of negations of the previous ones.

Definition 4.2. Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and X be a $\mathcal{M}$-fuzzy measurable $\mathbf{L}$-fuzzy set. The *$\rightarrow$-fuzzy integral of A on X* is given by

$$\int^{\rightarrow}_X A\, d\nu = \bigwedge_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \bigvee_{m \in \mathrm{Supp}(Y)} (A(m) \rightarrow \nu(Y)). \tag{28}$$

If $X = 1_M$, then we write $\int^{\rightarrow} A\, d\nu$.

Remark 4.3. It is easy to see that $\int^{\rightarrow}_{1_\emptyset} A\, d\nu = \bigwedge \emptyset = \top$ for any $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $\int^{\rightarrow}_X A\, d\nu \leq \int^{\rightarrow}_Y A\, d\nu$, whenever $Y \subseteq X$. Since $\int^{\rightarrow}_{1_M} A\, d\nu \neq \bot$ in general, $\nu_A(X) = \int^{\rightarrow}_X A\, d\nu$ does not define a complementary fuzzy measure on $(M, \mathcal{M})$ in the sense of Definition 3.3.

[5] Note that, according to Theorem 2.1, p. 2, each complete residuated lattice satisfying the law of double negation has this property. Nevertheless, there are fuzzy measurable spaces which keep this property, but $\mathbf{L}$ does not satisfy the law of double negation. A simple example is a fuzzy measurable space $(M, \mathcal{M})$ such that $\mathcal{M} \subseteq \mathcal{P}_{\mathbf{L}}(M)$ and $\mathbf{L}$ is an arbitrary complete residuated lattice (e.g. $L_{[0,\infty]}$ from Example 2.2).

ISBN: 978-989-95079-6-8

Theorem 4.8. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space. Then $\nu' : \mathcal{F}_{\mathbf{L}}(M) \to L$ defined by*

$$\nu'(A) = \int^{\to} A \, d\nu \tag{29}$$

is a complementary fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$.

Theorem 4.9. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space. Then*

(i) $\int_X^{\to}(A \cap B)\, d\nu \geq \int_X^{\to} A \, d\nu \vee \int_X^{\to} B \, d\nu$,

(ii) $\int_X^{\to}(A \cup B)\, d\nu \leq \int_X^{\to} A \, d\nu \wedge \int_X^{\to} B \, d\nu$,

(iii) $\int_X^{\to}(c \otimes A)\, d\nu \leq c \to \int_X^{\to} A \, d\nu$,

(iv) $\int_X^{\to}(c \to A)\, d\nu \geq c \otimes \int_X^{\to} A \, d\nu$

hold for any $X \in \mathcal{M}$, $A, B \in \mathcal{F}_{\mathbf{L}}(M)$ and $c \in L$.

Theorem 4.10. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space and $c \in L$. Then we have*

(i) $\int^{\to}(c \otimes 1_X)\, d\nu = c \to \nu^*(1_X)$ *for any $X \subseteq M$,*

(ii) $\int^{\to}(c \otimes 1_X)\, d\nu = c \to \nu(1_X)$ *for any $X \subseteq M$ such that $1_X \in \mathcal{M}$,*

(iii) $\int^{\to} 1_X \, d\nu = \nu(1_X)$ *for any $X \subseteq M$ such that $1_X \in \mathcal{M}$,*

(iv) $\int^{\to} c \, d\nu = \neg c$.

Theorem 4.11. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space. If $X \in \mathcal{M}$ is such that $1_{\mathrm{Supp}(A)} \in \mathcal{M}_X$ for any $A \in \mathcal{M}_X$, then, for any $A \in \mathcal{F}_{\mathbf{L}}(M)$, we have*

$$\int_X^{\to} A \, d\nu = \bigvee_{1_Y \in \mathcal{P}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in Y} (A(m) \to \nu(1_Y)), \tag{30}$$

where $\mathcal{P}_X = \{1_{\mathrm{Supp}(A)} \mid A \in \mathcal{M}_X\}$.

Theorem 4.12. *Let $\mathbf{L}$ be a complete MV-algebra, $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $X \in \mathcal{M}$. Then*

$$\int_X^{\to} A \, d\nu = \bigwedge_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \big((\bigwedge_{m \in \mathrm{Supp}(Y)} A(m)) \to \nu(Y) \big). \tag{31}$$

Moreover,

$$\int_X^{\to} (c \otimes A)\, d\nu = c \to \int_X^{\to} A \, d\nu \tag{32}$$

for any $c \in L$.

Theorem 4.13. *Let g be an isomorphism between complementary fuzzy measure spaces $(M, \mathcal{M}, \nu)$ and $(M', \mathcal{M}', \nu')$ and $X \in \mathcal{M}$. Then we have*

$$\int_X^{\to} A \, d\nu = \int_{g(X)}^{\to} g(A) \, d\nu' \tag{33}$$

for any $A \in \mathcal{F}_{\mathbf{L}}(M)$.

The following statement shows that if we consider a complete MV-algebra, then we can restrict ourselves, for example, to $\otimes$-fuzzy integrals, since each $\to$-fuzzy integral is uniquely determined by the negation of a suitable $\otimes$-fuzzy integral.

Theorem 4.14. *Let $\mathbf{L}$ be a complete MV-algebra and $(M, \mathcal{M})$ be a fuzzy measurable space. Then*

$$\int_X^{\to} A \, d\nu' = \neg \int_X^{\otimes} A \, d\mu, \tag{34}$$

$$\int_X^{\otimes} A \, d\mu' = \neg \int_X^{\to} A \, d\nu \tag{35}$$

hold for any fuzzy measure μ and complementary fuzzy measure ν, where $\nu' = \neg\mu$ and $\mu' = \neg\nu$.

5 Conclusions

In this contribution, new types of fuzzy integrals, that are quite useful for modeling fuzzy quantifiers (of the type $\langle 1 \rangle$), are introduced and some of their properties are studied. The definitions of fuzzy quantifiers using these types of fuzzy integrals and some of their semantical properties could be found in [2].

Acknowledgment

This paper has been supported by the Grant IAA108270901 of the GA AV ČR.

References

[1] V. Novák. A formal theory of intermediate quantifiers. *Fuzzy Sets and Systems*, 159(10):1229 1246, 2008.

[2] A. Dvořák and M. Holčapek. Basic properties of **L**-fuzzy quantifiers of the type $\langle 1 \rangle$ determined by fuzzy measures. In *IFSA/EUSFLAT 2009*, Lisboa, 2009.

[3] R. Mesiar. Fuzzy measures and integrals. *Fuzzy Sets and Systems*, 156:365–370, 2005.

[4] M. Ying. Linguistic quantifiers modeled by Sugeno integrals. *Artificial Intelligence*, 170:581–606, 2006.

[5] W. San-min and Z. Bin. Prenex normal form in linguistic quantifiers modeled by Sugeno integrals. *Fuzzy Sets and Systems*, 159:1719–1723, 2008.

[6] I. Glöckner. *Fuzzy Quantifiers: A Computational Theory*. Springer-Verlag, Berlin, 2006.

[7] M. Holčapek. Monadic **L**-fuzzy quantifiers of the type $\langle 1^n, 1 \rangle$. *Fuzzy Sets and Systems*, 159:1811–1835, 2008.

[8] M. Sugeno. *Theory of Fuzzy Integrals and its Applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[9] R. Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer Academic Publisher, New York, 2002.

[10] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht, 1998.

[11] Z. Wang and G.J. Klir. *Fuzzy measure theory*. Plenum Press, New York, 1992.

[12] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy Measures and Integrals. Theory and Applications*. Studies in Fuzziness and Soft Computing. Physica Verlag, Heidelberg, 2000.

[13] E.P. Klement, R. Mesiar, and E. Pap. A universal integral based on measures of level sets. *IEEE Transactions on Fuzzy Systems*, to appear.

[14] L. Mišík and J. Tóth. On asymptotic behaviour of universal fuzzy measure. *Kybernetika*, 42(3):379–388, 2006.

[15] S. Peters and D. Westerståhl. *Quantifiers in Language and Logic*. Oxford University Press, New York, 2006.

ISBN: 978-989-95079-6-8

A link between the 2-additive Choquet Integral and Belief functions

Brice Mayag[1,2] Michel Grabisch[1] Christophe Labreuche[2]

1. University of Paris I, Centre d'Economie de la Sorbonne
106-112 Boulevard de l'Hôpital, 75013 Paris, France
2. Thales Research & Technology
Campus de Polytechnique, 1 avenue Augustin Fresnel
91767 Palaiseau Cedex- France
{brice.mayag, christophe.labreuche}@thalesgroup.com, michel.grabisch@univ-paris1.fr

Abstract— In the context of decision under uncertainty, we characterize the 2-additive Choquet integral on the set of fictitious acts called binary alternatives or binary actions. This characterization is based on a fundamental property called MOPI which permits us to relate belief functions and the 2-additive Choquet integral.

Keywords— Capacity, Möbius transform, Choquet integral, k-monotone function, Belief function, Decision under uncertainty

1 Introduction

Decision under uncertainty is a part of decision making where each act has several possible consequences, depending on the state of nature whose probability of occurrence is unknown. As shown by the well-known Ellsberg's paradox [1, 2], the use of the expected utility model [3] in decision under uncertainty is limited. Therefore some non-additive models like Choquet expected utility [4] have been proposed in order to overcome the limitations of the expected utility model.

The Choquet integral is defined w.r.t. a capacity (or non-additive monotonic measure, or fuzzy measure), and can be thought of as a generalization of the expected value, the capacity playing the role of a probability measure. In this paper we focus on the 2-additive Choquet integral [5, 6], a particular Choquet integral where interaction between two states of nature can be represented, but not more complex interaction. This model is in practice already sufficiently flexible. In many situations, it is important for the Decision-Maker (DM) to construct a preference relation over the set of all acts X. Because it is not an easy task (the cardinality of X may be very large), we ask him to give, using pairwise comparisons, an ordinal information (a preferential information containing only a strict preference and an indifference relations) on a particular reference subset $\mathcal{B} \subseteq X$. The set $\mathcal{B}$ we use is the set of binary acts or binary actions. A binary action is a fictitious act which takes only two values denoted $\mathbf{1}$ and $\mathbf{0}$ belonging to the set of consequences, such that $\mathbf{1}$ is strictly preferred to $\mathbf{0}$. We present necessary and sufficient conditions on the ordinal information for the existence of a 2-additive capacity such that the Choquet integral w.r.t. this capacity represents the preference of the decision maker. We introduce the new fundamental property MOPI, a kind of monotonicity coming from the definition of a 2-additive capacity, in order to have this characterization. We found through our MOPI property the following link between the 2-additive Choquet integral and belief functions (Shafer [7]): *Any ordinal information representable by a belief function is representable by a Choquet integral w.r.t. a 2-additive capacity*.

Because a belief function is a capacity, we show another characterization of the representation of any ordinal information by a belief function. The new fundamental property defined in this case is called the 2-MOPI property. This property and the MOPI property in the previous paragraph are related by the following statement: *if the 2-MOPI property is satisfied then the MOPI property is satisfied*.

The article is organized as follows. The next section presents the basic concepts we need. Section 3 concerns a representation of ordinal information by the 2-additive Choquet integral. In the last section, after some results on the case of the k-monotone functions, we study the representation of an ordinal information by a belief function.

2 Preliminaries

Let us denote by $N = \{1, \ldots, n\}$ the set of n states of nature and by 2^N the set of all subsets of N. The set of possible consequences (also called "outcomes") is denoted by C. An act x is identified to an element of $X = C^n$ with $x = (x_1, ..., x_n)$. We introduce the following convenient notation: for two acts $x, y \in X$ and a subset $A \subseteq N$, the compound act $z = (x_A, y_{N-A})$ is defined by $z_i := x_i$ if $i \in A$, and $z_i := y_i$ otherwise. For all i, j in N, the element $i \vee j$ denotes one of the elements i, j.

We want to construct a preference relation over X, but this is not easy because X may contain infinitely many acts. In practice [8] one can only ask to the DM pairwise comparisons of acts on a finite subset X' of X. Hence we get a preference relation $\succsim_{X'}$ on X'. The question is then: how to construct $\succsim_X$ from $\succsim_{X'}$? To this end, people usually suppose that $\succsim_X$ is representable by an overall utility function:

$$x \succsim_X y \Leftrightarrow F(U(x)) \geq F(U(y)) \qquad (1)$$

where $U(x) = (u(x_1), \ldots, u(x_n))$, $u : C \to \mathbb{R}$ is called a utility function, and $F : \mathbb{R}^n \to \mathbb{R}$ is an aggregation

ISBN: 978-989-95079-6-8

function. Usually, we consider a family of aggregation functions characterized by a parameter vector θ (e.g., a probability distribution over the states of nature). The parameter vector θ can be deduced from the knowledge of $\succsim_{X'}$, that is, we determine the possible values of θ for which (1) is fulfilled over X'. We study the case where F is the Choquet integral, the parameters are the 2-additive capacity and X' is the set of binary actions. The aim of this paper is to give necessary and sufficient conditions on $\succsim_{X'}$ to be represented by a 2-additive Choquet integral. The model obtained in X' will be automatically extended to X.

2.1 The 2-additive Choquet Integral

The 2-additive Choquet integral [6] is a particular case of the Choquet integral [9, 10]. This integral generalizes the arithmetic mean and takes into account interactions between the states of nature. A 2-additive Choquet integral is based on a 2-additive capacity defined below and its Möbius transform [11, 12]:

Definition 2.1.

1. A *capacity* on N is a set function $\mu : 2^N \to [0,1]$ such that:
   (a) $\mu(\emptyset) = 0$
   (b) $\mu(N) = 1$
   (c) $\forall A, B \in 2^N$, $[A \subseteq B \Rightarrow \mu(A) \leq \mu(B)]$ (monotonicity).
2. The *Möbius transform* of a capacity μ on N is a function $m : 2^N \to \mathbb{R}$ defined by:

$$m(T) := \sum_{K \subseteq T} (-1)^{|T \setminus K|} \mu(K), \forall T \in 2^N \qquad (2)$$

When m is given, it is possible to recover the original μ by the following expression:

$$\mu(T) := \sum_{K \subseteq T} m(K), \forall T \in 2^N \qquad (3)$$

Definition 2.2. A capacity μ on N is said to be *2-additive* if

- For all subset T of N such that $|T| > 2$, $m(T) = 0$;
- There exists a subset B of N such that $|B| = 2$ and $m(B) \neq 0$.

Notations We simplify our notations by using for a capacity μ and its Möbius transform m: $\mu_i := \mu(\{i\})$, $\mu_{ij} := \mu(\{i,j\})$, $m_i := m(\{i\})$, $m_{ij} := m(\{i,j\})$, for all $i, j \in N$, $i \neq j$. Whenever we use i and j together, it always means that they are different.

The following important Lemma shows that a 2-additive capacity is entirely determined by the value of the capacity on the singletons $\{i\}$ and pairs $\{i,j\}$ of 2^N:

Lemma 1.

1. *Let μ be a 2-additive capacity on N. We have*

$$\mu(K) = \sum_{\{i,j\} \subseteq K} \mu_{ij} - (|K|-2) \sum_{i \in K} \mu_i, \ \forall K \subseteq N, \ |K| \geq 2. \qquad (4)$$

2. *If the coefficients μ_i and μ_{ij} are given for all $i, j \in N$, then the necessary and sufficient conditions that μ is a 2-additive capacity are:*

$$\sum_{\{i,j\} \subseteq N} \mu_{ij} - (n-2) \sum_{i \in N} \mu_i = 1 \qquad (5)$$

$$\mu_i \geq 0, \ \forall i \in N \qquad (6)$$

$$\sum_{i \in A \setminus \{k\}} (\mu_{ik} - \mu_i) \geq (|A| - 2)\mu_k, \ \forall A \subseteq N, |A| \geq 2, \forall k \in A. \qquad (7)$$

Proof. See [6] □

For an act $x := (x_1, ..., x_n) \in X$, the expression of the Choquet integral w.r.t a capacity μ is given by:

$$C_\mu((u(x_1), ..., u(x_n))) := u(x_{\tau(1)})\mu(N) + \sum_{i=2}^{n} (u(x_{\tau(i)}) - u(x_{\tau(i-1)}))\mu(\{\tau(i), ..., \tau(n)\})$$

where τ is a permutation on N such that $u(x_{\tau(1)}) \leq u(x_{\tau(2)}) \leq ... \leq u(x_{\tau(n-1)}) \leq u(x_{\tau(n)})$.

A Choquet integral with a 2-additive capacity μ is called a 2-additive Choquet integral. Given an act $x := (x_1, ..., x_n) \in X$, the 2-additive Choquet integral can be written also as follows:

$$C_\mu((u(x_1), \ldots, u(x_n))) = \sum_{i=1}^{n} v_i u(x_i) - \frac{1}{2} \sum_{\{i,j\} \subseteq N} I_{ij} |u(x_i) - u(x_j)| \qquad (8)$$

where $v_i = \sum_{K \subseteq N \setminus i} \frac{(n - |K| - 1)!|K|!}{n!} (\mu(K \cup i) - \mu(K))$ represents the importance of the state of nature i and corresponds to the Shapley value [13]; $I_{ij} = \mu_{ij} - \mu_i - \mu_j$ is the interaction index between the two states of nature i and j.

The above development suggests that the Choquet integral w.r.t. a 2-additive capacity seems to be of particular interest, and offers a good compromise between flexibility of the model and complexity. Therefore, we focus in this paper on the 2-additive model.

2.2 Binary actions

We assume in this paper that the DM is able to identify in C two consequences denoted $\mathbf{1}$ and $\mathbf{0}$ such that he strictly prefers $\mathbf{1}$ to $\mathbf{0}$. In the sequel, we call $\mathbf{0}$ the "neutral level" (even if this is not the neutral level understood in bipolar model).

We call a *binary action or binary act*, an element of the set $\mathcal{B} = \{\mathbf{0}_N, \ (\mathbf{1}_i, \mathbf{0}_{N-i}), \ (\mathbf{1}_{ij}, \mathbf{0}_{N-ij}), \ i, j \in N, i \neq j\} \subseteq X$ where

- $\mathbf{0}_N = (\mathbf{1}_\emptyset, \mathbf{0}_N) =: a_0$ is an act which has a consequence $\mathbf{0}$ on all states of nature.

ISBN: 978-989-95079-6-8

- $(\mathbf{1}_i, \mathbf{0}_{N-i}) =: a_i$ is an act which has a consequence $\mathbf{1}$ on state of nature i and a consequence $\mathbf{0}$ on the other states of nature.

- $(\mathbf{1}_{ij}, \mathbf{0}_{N-ij}) =: a_{ij}$ is an act which has a consequence $\mathbf{1}$ on states of nature i and j and a consequence $\mathbf{0}$ on the other states of nature.

By convention we set $u(\mathbf{0}) = 0$ and $u(\mathbf{1}) = 1$. The above convention have the following consequences:

Remark 1.

1. The Choquet integral satisfies the following property [14, 10]: if μ is a capacity then

$$C_\mu(U(\mathbf{1}_A, \mathbf{0}_{N-A})) = \mu(A),\ \forall A \subseteq N. \qquad (9)$$

2. Let μ be a 2-additive capacity. We have

$C_\mu(U(a_0)) = 0$;

$C_\mu(U(a_i)) = \mu_i = v_i - \frac{1}{2} \sum_{k\in N,\ k\neq i} I_{ik}$;

$C_\mu(U(a_{ij})) = \mu_{ij} = v_i + v_j - \frac{1}{2} \sum_{k\in N,\ k\notin\{i,j\}} (I_{ik} + I_{jk})$

Generally the DM knows how to compare some acts using his knowledge of the problem, his experience, etc. These acts form a set of reference acts and allows to determine the parameters of a model (utility functions, subjective probabilities, weights,...) in the decision process (see [8] for more details). As shown by the previous Remark 1 and Lemma 1, it should be sufficient to get some preferential information from the DM only on binary acts. To entirely determine the 2-additive capacity this information is expressed by the following relations:
$P = \{(x, y) \in \mathcal{B} \times \mathcal{B} :$ DM strictly prefers x to $y\}$, $I = \{(x, y) \in \mathcal{B} \times \mathcal{B} :$ DM is indifferent between x and $y\}$.

Definition 2.3. The *ordinal information on* $\mathcal{B}$ is the structure $\{P,\ I\}$.

Now we will suppose P nonempty for any ordinal information $\{P,\ I\}$ ("non triviality axiom"). Before we end this section, let us introduce another relation M which completes the ordinal information $\{P,\ I\}$ given by the DM and models the natural relations of monotonicity between binary actions. For $(x, y) \in \{(a_i, a_0), i \in N\} \cup \{(a_{ij}, a_i), i, j \in N, i \neq j\}$,

$$x\ M\ y \text{ if not}(x\ (P \cup I)\ y).$$

The relation M models the monotonicity conditions $\mu(\{i\}) \geq 0$ and $\mu(\{i, j\}) \geq \mu(\{i\})$ for a capacity μ.

Example 1. *If we consider*
$N = \{1, 2, 3\}$, $\mathcal{B} = \{a_o, a_1, a_2, a_3, a_{12}, a_{13}, a_{23}\}$, $P = \{(a_{13}, a_3), (a_2, a_3), (a_{23}, 0)\}$, $I = \{(a_{12}, a_1)\}$, *then the relation* M *is* $M = \{(a_{12}, a_0), (a_{13}, a_0), (a_1, a_0), (a_2, a_0), (a_3, a_0), (a_{12}, a_2), (a_{13}, a_1), (a_{23}, a_2), (a_{23}, a_3)\}$.

3 The representation of the ordinal information by the Choquet integral

An ordinal information $\{P, I\}$ is said to be *representable by a 2-additive Choquet integral* if there exists a 2-additive capacity μ such that:

1. $\forall x, y \in \mathcal{B},\ x\ P\ y \Rightarrow C_\mu(U(x)) > C_\mu(U(y))$
2. $\forall x, y \in \mathcal{B},\ x\ I\ y \Rightarrow C_\mu(U(x)) = C_\mu(U(y))$.

Given an ordinal information $\{P, I\}$, we look for the necessary and sufficient conditions on $\mathcal{B}$ for which $\{P, I\}$ is representable by a 2-additive Choquet integral. To do it, we need to define first the notion of strict cycle of the relation $(P \cup I \cup M)$.

3.1 Cycle of $(P \cup I \cup M)$

For a binary relation $\mathcal{R}$ on $\mathcal{B}$ and x, y elements of $\mathcal{B}$, $\{x_1, x_2, \cdots, x_p\} \subseteq \mathcal{B}$ is a *path* of $\mathcal{R}$ from x to y if $x = x_1\ \mathcal{R}\ x_2\ \mathcal{R} \cdots \mathcal{R}\ x_{p-1}\ \mathcal{R}\ x_p = y$. A path of $\mathcal{R}$ from x to x is called a *cycle* of $\mathcal{R}$.

- A path $\{x_1, x_2, ..., x_p\}$ of $(P \cup I \cup M)$ is said to be a *strict path* from x to y if there exists i in $\{1, ..., p-1\}$ such that $x_i\ P\ x_{i+1}$. In this case, we will write $x\ TC_P\ y$.

- A cycle $(x_1, x_2, ..., x_p)$ of $(P \cup I \cup M)$ is a *nonstrict cycle* if it is not strict.

- We note $x \sim y$ if there exists a nonstrict cycle of $(P \cup I \cup M)$ containing x and y.

Contrarily to the strict cycle which is a classic concept used in graph theory [15, 16], we need to define a new fundamental property called *MOPI*.

3.2 MOPI property and theorem of Characterization

Before defining the property *MOPI*, let us discover this new condition through a simple example:

Example 2. *Suppose that the DM says :* $a_{12}\ I\ a_3$, $a_{13}\ I\ a_2$ *and* $a_1\ P\ a_0$. *Using the relation* M, *we have* $a_{12}\ M\ a_2\ I\ a_{13}\ M\ a_3\ I\ a_{12}$. *So,* $(a_{12}, a_2, a_{13}, a_3, a_{12})$ *forms a nonstrict cycle of* $(P \cup I \cup M)$. *If* $\{P, I\}$ *is representable by a 2-additive Choquet integral* C_μ, *this implies* $\mu_{12} = \mu_{13} = \mu_2 = \mu_3$ *and* $\mu_1 > 0$. *However, we get a contradiction with the monotonicity constraint* $\mu_{12} + \mu_{13} \geq \mu_1 + \mu_2 + \mu_3$ *of a 2-additive capacity with the subset* $A = \{1, 2, 3\}$, $k = 1$ *(see Equation* (7) *in Lemma 1).*

This type of inconsistency is defined by:

Definition 3.1 (MOPI property)**.** Let $i, j, k \in N$, i fixed.

1. We call *Monotonicity of Preferential Information in* $\{i, j, k\}$ *w.r.t.* i the following property (denoted by **(**$\{i, j, k\}$**,***i***)-MOPI**):

$$\left.\begin{array}{l} a_{ij} \sim a_{i\vee j} \\ a_{ik} \sim a_{i\vee k} \\ i \vee j \in \{i, j\} \\ i \vee k \in \{i, k\} \\ i \vee j \neq i \vee k \end{array}\right\} \Rightarrow \begin{array}{l} [not(a_l\ TC_P\ a_0) \\ l \in \{i, j, k\} \setminus \{i \vee k, i \vee j\}] \end{array}$$

ISBN: 978-989-95079-6-8

If the property **($\{i,j,k\}$,i)-MOPI** is satisfied then the element $a_l, l \in \{i,j,k\} \setminus \{i \vee k, i \vee j\}$ is called the *neutral binary action of* $\{i,j,k\}$ *w.r.t.* i.

2. We say that $\{i,j,k\}$ satisfies the property *Monotonicity of Preferential Information* (MOPI) if $\forall l \in \{i,j,k\}$, ($\{i,j,k\}$,l)-MOPI is satisfied.

Example 3. *Let* $N = \{1,2,3,4\}$ *and* $i = 1$ *fixed. The property* $(\{1,2,3\},1)$*- MOPI reads as follows:*

- $\left\{ \begin{array}{l} a_{12} \sim a_2 \\ a_{13} \sim a_1 \end{array} \right. \Rightarrow not(a_3\ TC_P\ a_0)$
- $\left\{ \begin{array}{l} a_{12} \sim a_1 \\ a_{13} \sim a_3 \end{array} \right. \Rightarrow not(a_2\ TC_P\ a_0)$
- $\left\{ \begin{array}{l} a_{12} \sim a_2 \\ a_{13} \sim a_3 \end{array} \right. \Rightarrow not(a_1\ TC_P\ a_0)$

The MOPI condition given in this paper is equivalent to the MOPI property presented in [5]. We give below our theorem of characterization of consistent ordinal information $\{P,I\}$ representable by a 2-additive Choquet integral:

Theorem 1. *An ordinal information* $\{P,I\}$ *is representable by a 2-additive Choquet integral on* $\mathcal{B}$ *if and only if the following conditions are satisfied:*

1. $(P \cup I \cup M)$ *contains no strict cycle;*
2. *Any subset* K *of* N *such that* $|K| = 3$ *satisfies the MOPI property.*

Is it possible to represent an ordinal information by another operator instead of the 2-additive Choquet integral? If the answer is yes, can we give a similar characterization like in Theorem 1? In the next section, we will show that it is possible by using for instance belief functions.

4 The representation of ordinal information by belief functions

4.1 General definitions

Beliefs functions are one of the fundamental concepts used in the theory of evidence of Shafer [7]. They are defined by the belief function mass m as follows:

Definition 4.1. A function $m : 2^N \rightarrow [0,1]$ is called a *mass distribution* or a *basic belief assignment* if m satisfies the following two properties:

1. $m(\emptyset) = 0$;
2. $\sum_{A \subseteq N} m(A) = 1$.

The quantity $m(A)$ expresses the total amount of belief that supports the proposition: " the actual state of nature is in A", and does not support any more specific subset of N because of lack of information.

Based on this concept, we define the belief function Bel by:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad \forall A \subseteq N.$$

Remark 2.

- Bel is a capacity;
- The sets A such that $m(A) > 0$ are called the focal elements;
- If all focal elements are singletons then a mass distribution can be considered as a probability distribution;
- The mass distribution m corresponds to the Möbius transform of Bel. So we have $\forall T \in 2^N$,

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B).$$

Thus, we can have a definition of the representation of ordinal information by a belief function which is similar to the same representation by a Choquet integral (see Section 3).

Definition 4.2. An ordinal information $\{P,I\}$ is said to be *representable by a belief function* if there exists a belief function Bel such that

1. $\forall x,y \in \mathcal{B},\ x\ P\ y \Rightarrow C_{Bel}(U(x)) > C_{Bel}(U(y))$
2. $\forall x,y \in \mathcal{B},\ x\ I\ y \Rightarrow C_{Bel}(U(x)) = C_{Bel}(U(y))$.

By using Definition 2.2, a 2-additive belief function has a mass distribution m characterized by:

1. $\exists i,j \in N$ tel que $m(\{i,j\}) \neq 0$;
2. $\forall K \in 2^N$ tel que $|K| \geq 3$, $m(K) = 0$.

Theorem 2 below provides a relation between a k-monotone function [6, 12] and a belief function, and a relation between k-monotone functions and the previous MOPI property.

4.2 k-monotone functions and belief functions

Given an integer $k \geq 2$, a function $\mu : 2^N \rightarrow [0,1]$ is k-monotone (shorthand for: monotone of order k) if for each family $\{A_1, A_2, \ldots, A_k\} \subseteq 2^N$, we have

$$\mu(\bigcup_{i=1}^{k} A_i) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \mu(\bigcap_{i \in I} A_i). \qquad (10)$$

A simpler characterization of k-monotone functions by their Möbius inversion is given by the following proposition:

Proposition 1. *Let* $\mu : 2^N \rightarrow [0,1]$ *and* m *its Möbius transform.* μ *is* k*-monotone (*k *integer,* $k \geq 2$*) if and only if*

$$\sum_{A \subseteq L \subseteq B} m(L) \geq 0 \quad \forall A,B \subseteq N, A \subseteq B \text{ and } 2 \leq |A| \leq k. \qquad (11)$$

ISBN: 978-989-95079-6-8

Proof. See [12] □

It is well-known that $\mu : 2^N \to [0,1]$ is a belief function if and only if μ is a k-monotone capacity for all $k \geq 2$. The following result gives another sufficient condition to obtain a belief function from a k-monotone and 2-additive capacity, and relates belief function with the MOPI condition translated in terms of capacity.

Theorem 2.
Let $\mu : 2^N \to [0,1]$ be a function and k be an integer such that $k \geq 2$.

1. *If μ is monotone, k-monotone and 2-additive then μ is a belief function (precisely a 2-additive belief function);*
2. *If μ is monotone and k-monotone then μ satisfies the following property: for all $i, j, k \in N$, i fixed*

$$\left.\begin{array}{l}\mu_{ij} = \mu_{i\vee j}\\ \mu_{ik} = \mu_{i\vee k}\\ i\vee j \in \{i,j\}\\ i\vee k \in \{i,k\}\\ i\vee j \neq i\vee k\end{array}\right\} \Rightarrow \begin{array}{l}[\mu_l = 0,\\ l \in \{i,j,k\}\setminus\{i\vee k, i\vee j\}]\end{array}$$

We end the paper by a characterization of ordinal information by belief functions.

4.3 A link between Belief functions and the 2-additive Choquet integral

In this section, we give through the MOPI property (see Section 3) a link between beliefs functions and the 2-additive Choquet integral.

Proposition 2. *Let $\{P,\ I\}$ be an ordinal information on $\mathcal{B}$.*
If there exist $i, j, k \in N$, i fixed such that the property $(\{i,j,k\}, i)$-MOPI is violated, then there is no belief function Bel which represents $\{P,\ I\}$.

Corollary 1. *Every ordinal information $\{P,\ I\}$ on $\mathcal{B}$ representable by a belief function $Bel : 2^N \to [0,1]$ is representable by a 2-additive Choquet integral.*

The inverse of Corollary 1 is false. If we suppose $P = \{(a_2, a_0)\}$, $I = \{(a_{12}, a_1)\}$ and μ a 2-additive capacity, we will have $\{P,\ I\}$ representable by a 2-additive Choquet integral and $I_{12} = m_{12} = \mu_{12} - \mu_1 - \mu_2 < 0$. So no belief function can represent $\{P,\ I\}$ in this case. Then it is interesting to look for the class of 2-additive capacities which are belief functions. In order to characterize them, we introduce a new fundamental property called 2-MOPI property:

Definition 4.3. An ordinal information $\{P,\ I\}$ satisfies the 2-*MOPI property* if

$$\forall i, j \in N, i \neq j, [a_{ij} \sim a_i \Rightarrow not(a_j\ TC_P\ a_0)]. \qquad (12)$$

The relation between the 2-MOPI property and the MOPI property is given by the following proposition:

Proposition 3. *Let $\{P,\ I\}$ an ordinal information on $\mathcal{B}$.*

$$\{P,\ I\} \text{ satisfies the 2-MOPI property}$$
$$\Downarrow$$
$$\forall i, j, k \in N,\ \{i,j,k\} \text{ satisfies the MOPI property}$$

Now we have the main result of this section:

Theorem 3.
$\{P,\ I\}$ is representable by a 2-additive belief function if and only if the two following conditions are satisfied:

1. *$(P \cup I \cup M)$ contains no strict cycle;*
2. *$\{P,\ I\}$ satisfies the 2-MOPI property.*

4.4 Interpretation of 2-MOPI and MOPI properties

We try to give an interpretation in terms of decision behavior of the two main conditions introduced in this paper. We assume here for clarity that consequence **1** is a good consequence for the DM, while consequence **0** is neither bad nor good (statu quo).

Facing a situation where for two states of nature i and j the DM is indifferent between the two acts a_{ij} and a_i, the 2-MOPI property says that act a_j is equivalent to act a_0 (statu quo for every state of nature). Hence in such a situation, the DM thinks that state of nature j is unlikely to occur. This is a strong condition, since it suffices that one such state i exists to infer the "nullity" of state j. This condition can be related to the notion of null set in generalized measure theory (see, e.g., [17]): a set $A \subseteq N$ is said to be *null* for capacity μ if $\mu(B \cup A) = \mu(B)$, $\forall B \subseteq N \setminus A$. Taking $A = \{j\}$ and $B = \{i\}$ gives our condition 2-MOPI. Observe that for the nullity condition, $\{j\}$ would be null if for *all* subsets B not containing j we would have $\mu(B \cup j) = \mu(B)$, but the 2-MOPI condition asks to find only one singleton satisfying this equality.

The MOPI property is a weakening of the above one, and can be interpreted in a similar way. Let us consider now three states of nature i, j and k. The MOPI condition can be translated as follows (see Example **??**, with $i = 1$, $j = 2$, and $k = 3$). Suppose that a_{ij} and a_j are indifferent. As above, this would suggest that i is unlikely to occur for the DM, but this is *relatively to the occurence of* j, or put differently, i is much less likely than j. Suppose in addition that a_{ik} is indifferent to a_i. Again, this suggests that k much less likely to occur than i. Since i is much less likely than j, the conclusion is that k is very unlikely to occur, hence a_k is indifferent to a_0. This explains the first case in the MOPI condition. The second case (indifference between a_{ik} and a_k, and between a_{ij} and a_i) works exactly the same way. The third case says that a_{ij} and a_j are indifferent (i is much less likely than j) as well as a_{ik} and a_k (i is much less likely than k). Since i is much less likely than both j and k, the conclusion is that i is very unlikely, so that a_i is indifferent with a_0.

For $N = \{1,2\}$, the 2-MOPI property can be also viewed as uncertainty aversion[1] (see [18]). Indeed,

[1] Uncertainty aversion, as presented in [18], is defined as follows: For three acts x, y, z, if y and z are comonotonic then:

$$x \sim y \Rightarrow x + z \gtrsim y + z.$$

Comonotonicity between two acts y, z means that there are no $i, j \in N$ such that $u(y_i) > u(y_j)$ and $u(z_i) < u(z_j)$.

ISBN: 978-989-95079-6-8

given the three acts $a_{12} = (\mathbf{1}, \mathbf{1})$, $a_1 = (\mathbf{1}, \mathbf{0})$, $z = (-\mathbf{1}, \mathbf{0})$ and using the property of uncertainty aversion, we have:

$$(\mathbf{1}, \mathbf{0}) \sim (\mathbf{1}, \mathbf{1}) \Rightarrow (\mathbf{0}, \mathbf{0}) \gtrsim (\mathbf{0}, \mathbf{1})$$

which corresponds to the 2-MOPI property in this case. However, this interpretation does not work any more for the MOPI condition.

References

[1] D. Ellsberg. Risk, ambiguity, and the Savage axioms. *Quart. J. Econom.*, 75:643–669, 1961.

[2] A. Chateauneuf. Ellsberg paradox intuition and Choquet expected utility. In *Mathematical models for handling partial knowledge in artificial intelligence*. Plenum Press, 1995.

[3] J. Von Neuman and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[4] D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.

[5] B. Mayag, M. Grabisch, and C. Labreuche. A characterization of the 2-additive Choquet integral. In *Proceedings of IPMU'08 (CD)*, Malaga, Spain, June 2008.

[6] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.

[7] G. Shafer. *A Mathematical Theory of Evidence*. Princeton Univ. Press, 1976.

[8] Th. Marchant. Towards a theory of MCDM: Stepping away from social choice theory. *Mathematical Social Sciences*, 45(3):343–363, 2003.

[9] M. Grabisch and Ch. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *4OR*, 6:1–44, 2008.

[10] M. Grabisch and C. Labreuche. Fuzzy measures and integrals in MCDA. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis*, pages 563–608. Kluwer Academic Publishers, 2004.

[11] M. Grabisch. The Möbius transform on symmetric ordered structures and its application to capacities on finite sets. *Discrete Mathematics*, 287 (1-3):17–34, 2004.

[12] A. Chateauneuf and J.Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.

[13] L.S. Shapley. A value for n-person games. *Contributions to the Theory of games*, pages 307–317, 1953.

[14] J.L. Marichal. An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Tr. on Fuzzy systems*, 8(6):800–807, 2000.

[15] M. Gondran and M. Minoux. *Graphes et algorithmes*. Eyrolles, 1986.

[16] M. Pirlot and Ph. Vincke. *Semiorders*. Kluwer academic publishers, 1997.

[17] E. Pap. *Null-Additive Set Functions*. Kluwer Academic, 1995.

[18] A. Chateauneuf. Modeling attitudes towards uncertainty and risk through the use of Choquet integral. *Annals of Operations Research*, 52:3–20, 1994.

ISBN: 978-989-95079-6-8

Agent-based Fuzzy Constraint-directed Negotiation Mechanism for Planning and Scheduling in Supply Chain

Bo-Ruei Kao,[1] K. Robert Lai,[1] Menq-Wen Lin[2]

1.Department of Computer Science & Engineering, Yuan Ze University,
Chung-Li, Taiwan
2.Department of Information Management, Ching Yun University,
Chung-Li, Taiwan
Email: s909402@mail.yzu.edu.tw, krlai@cs.yzu.edu.tw, linmw@cyu.edu.tw

Abstract— *This paper presents an agent-based fuzzy constraint-directed negotiation mechanism for planning and scheduling in supply chain. The supply chain scheduling problem is modeled as a set of fuzzy constraint satisfaction problems (FCSPs), interlinked together by inter-agent constraints. For converging each distinct firm's interests, the conflicts among the set of FCSPs are resolved through negotiation by iteratively exchanging offers/counters with limited sharing of their perspectives and preferences. During the negotiation, proposing offers/counter-offers takes not only firm's self-interest and preferences but also opponents' perspectives into consideration. By sharing perspectives between agents to gradually uncovering the intent of opponents, consensus can be obtained and the quality of consensus can be guaranteed at satisfactory level. Experimental results suggest that the proposed approach obtains a superior solution for supply chain scheduling than other negotiation models in a fully distributed processing.*

Keywords— Supply Chain Planning and Scheduling, Fuzzy Constraint-directed Negotiation, Fuzzy Constraint Satisfaction Problem, Multi-Agent System.

1 Introduction

Facing with global economic downturn and intense competition, firms have to rapidly respond to the variations in market situation. Business entities (suppliers, manufacturers, distributors, retailers and customers) are urged to integrate their operations into multi-layer supply chain for product/service provision [1]. However, supply chain usually handles multiple projects concurrently with shared components, facilities, and capacities governed by distinct firms. Collaborating with supply chain partners in planning and scheduling in a cost effective manner is the key to survive in today's fierce market competition.

Since the supply chain environment in nature is distributed, autonomous, and heterogeneous, agent-based approaches which are characterized by decentralization of computation and information processing are particularly attractive for supply chain modeling and problem solving [2, 3]. To ensure the global performance with limited interaction among agents, a mediator or a third party agent is used in some systems for coordination. Negotiation processes can be facilitated by a mediator which helps the parties understand their needs, suggests possible agreements and/or supports them in the implementation of the agreement. But, agents may have to share sensitive strategic information that should not be revealed to competitors or even to a third party agent, to the mediator for coordination.

So, instead of coordinating with a third party agent, coordination/negotiation in a fully distributed environment for privacy concern has been addressed in some agent-based approaches. Those papers address the dynamic nature of supply chain and emphasize the flexibility and responsiveness. Two popular and fully distributed negotiation models, contract net protocol (CNP) and market-based protocol, have been proposed for these purposes [4]. The CNP can rapidly produce a feasible solution to overcome the frequently changes in supply chain. With simplified negotiation protocol, however, agents with very limited interaction and information sharing can only make their decisions independently and optimize their local objectives in a myopic way. Accordingly, more sophisticated negotiation mechanism is needed and thus bringing us to a market-based approach. Market-based approaches employ a bargaining or auction process characterized by iterative bidding among agents. During the bidding process, agents will adjust bids according to the direction of surplus and deficit of demand. The bids which imply the degree of competition can be used to resolve the conflicts caused by contention and improve the system performance. But the bid indicates only the demand that is in high or low contention, not where the contention can be resolved. Thus, the process might oscillate and not achieve the convergence. It also affects the quality of solution.

Accordingly, facilitating the convergence and guaranteeing the system performance in supply chain are the critical challenges and they are highly influenced by the level of information sharing [5, 6, 7, 8] as well. To cope with these problems, this paper proposes an agent-based fuzzy constraint-directed negotiation mechanism (AFCN) in a fully distributed environment.

In AFCN, a supply chain problem is modeled as a set of fuzzy constraint satisfaction problems (FCSPs), interlinked together by inter-agent constraints. Each FCSP represents the firm's perspectives and issues in supply chain. Constraints which are characterized by non directional, declarative, and intuitive properties are closely related to the real-world problem descriptions. Besides, the subjective, imprecision and qualitative knowledge (e.g., human cognition, preferences, or even opponent's perspectives) which are frequently used for business decision making in supply chain can easily be coped

ISBN: 978-989-95079-6-8

by fuzzy constraints with the levels of consistency. Fuzzy constraints also can be used to rank the solutions by specifying the possibilities prescribing to what extent the solutions are suitable. Additionally, it even provides a measure of similarity between solutions and opponents' counteroffers and a basis for the selection of multi-objective decision from a set of feasible alternatives. Thus, AFCN can be a practical and effective methodology for the supply chain problem modeling.

To coordinate supply chain parties, the conflicts among set of FCSPs are resolved through AFCN negotiation protocol by iteratively exchanging offers/counters with limited sharing of their perspectives. For converging each distinct firm's interests in supply chain, proposing an offer/counteroffer takes not only firm's self-interest and preferences but also opponents' perspectives into consideration. In a proposed offer/counteroffer, it indicates not only the region of acceptable solutions and preference degrees but the possibility of conflicts in the region. For each FCSP, incremental propagation eliminates the redundant decision values and infeasible combination of solutions. According to the ranking of solutions by fuzzy constraints, the set of feasible solutions can be further restricted in a preferred region with an acceptable/controlled threshold. It supports agent to quick response the changes in the environment and promises the proposed offers/counteroffers to be focused within the interest/attention area.

By sharing limited information among agents to incrementally uncovering the intention of opponents, consensus can be obtained and the quality of solution can be guaranteed at a satisfaction level as well. Moreover, the framework of AFCN also provides flexibility to incorporate negotiation strategies, such as self-interested, cooperative, and win-win, for various global performance measures.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical basis for modeling supply chain planning and scheduling problem as a distributed fuzzy constraint satisfaction problem (DFCSP). Section 3 then presents an agent-based fuzzy constraint-directed negotiation mechanism in detail. Then, Section 4 provides the experimental results to show the effectiveness of the proposed approach followed by conclusions in Section 5.

2 Modeling Supply Chain Planning and Scheduling as DFCSP

During the process of production/service provision from acquiring material to deliver to the end customers, it passes through the stages of raw material supply, intermediate supply, manufacturing, distribution, and retail. This paper focuses on the stage of manufacturing in supply chain. The supply chain project scheduling problem, as shown in Figure 1, is to schedule multiple projects in a network of manufacturers and suppliers (or contractors). Each project involves a set of tasks or operations with complex precedence relationships in which the task can be performed by a set of alternative suppliers. Suppliers differ from each other in terms of resource capacity, processing times, and costs of performing tasks. The process of project planning and scheduling specifies how the suppliers are selected from alternatives and tasks are scheduled. The problem is multi–objective in nature. For example, manufacturers wish to optimize the scheduling performance as well as to minimize the operation cost. On the other hand, the suppliers are more inclined to maximize the utilization and profit.

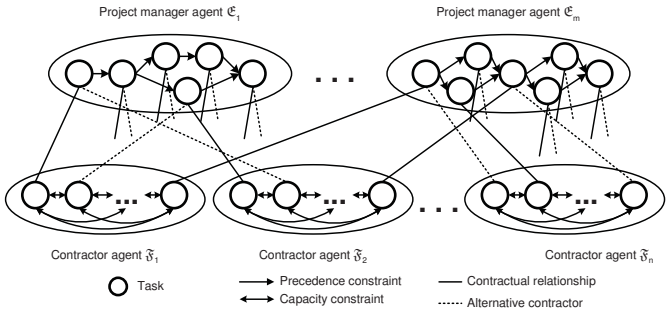


Figure 1: Supply chain project scheduling with m project agents and n contractor agents.

The problem can be represented as a triple $(\mathfrak{E}, \mathfrak{F}, \mathfrak{I})$, where $\mathfrak{E}$ is a set of m project manager agents (PAs) for the manufacturers, $\mathfrak{F}$ is a set of n contractor agents (CAs) for the suppliers, and $\mathfrak{I}$ is a set of inter-agent constraints between two class of agents. Each PA consists of a chain of tasks which are specified further by precedence constraints, including the processing time on the set of alterative suppliers for each task, due date, arrival (release) date, and tardiness cost. Each CA consists of capacity constraints and processing cost. Thus, the problem can be modeled as a distributed constraint satisfaction problem (DFCSP) in that coming up to a mutually acceptable solution between two classes of agents is the same as uncovering a consistent solution satisfying all the constraints in a distributed fuzzy constraint network specifying the fuzzy relationships inside each agent and between agents. Adapted from [9], a distributed fuzzy constraint network (DFCN) can be defined as below.

Definition 1
A distributed fuzzy constraint network $(\mathcal{U}, \mathbf{X}, \mathbf{C})$ *for a supply chain project scheduling problem* $(\mathfrak{E}, \mathfrak{F}, \mathfrak{I})$ *can be defined as a set of* $m+n$ *fuzzy constraint networks* $\{\mathfrak{N}_1, \ldots, \mathfrak{N}_{m+n}\}$, $\mathfrak{N}_l$ *representing an agent* $l \in \{\mathfrak{E}, \mathfrak{F}\}$, *where*

- $\mathcal{U}_l$ *is a universe of discourse for an FCN* $\mathfrak{N}_l$*;*
- $\mathbf{X}_l$ *is a tuple of non-recurring objects of agent* l*;*
- $\mathbf{C}_l$ *is a set of fuzzy constraints which involves a set of internal fuzzy constraints existing among objects in* $\mathbf{X}_l$*, and a set of external fuzzy constraints* $\mathfrak{I}_l$ *between agent* l *and opponent agents;*
- $\mathfrak{N}_l$ *is connected to other FCNs by a set of external fuzzy constraints* $\mathfrak{I}_l$*;*
- $\mathcal{U}$ *is a universe of discourse;*
- $\mathbf{X} = \left(\cup_{l=1}^{m+n} \mathbf{X}_l\right)$ *is a tuple of all non-recurring objects;*
- $\mathbf{C} = \left(\cup_{l=1}^{m+n} \mathbf{C}_l\right)$ *is a set of all fuzzy constraints.*

In Definition 1, a set of $\mathbf{X}_l$ of agent l is corresponded to its beliefs, its knowledge of the environment (tasks, resources, etc.), and any other attitudes (desires, intentions, opponents' response, etc.). A set of fuzzy constraints $\mathbf{C}_i$ for project agent $\mathfrak{E}_i$ is corresponding to a set of restrictions (e.g. precedence

ISBN: 978-989-95079-6-8

constraints), objectives (e.g. flowtime, operating cost) and inter-agent constraints between project agent $\mathfrak{E}_i$ and related contractor agents; $\mathbf{C}_k$ for contractor agent $\mathfrak{F}_k$ is corresponding to a set of restrictions (e.g. capacity constraints), objectives (e.g. utilization) and inter-agent constraints between contractor agent $\mathfrak{F}_k$ and related project agents.

By Definition 1, a consistent solution of a DFCSP for scheduling is an instantiation of all time allocation of the tasks such that all the constraints of the agents are satisfied at a degree of membership function. No agent knows about other agents' feasible solutions and possible agreements a priori. An agent negotiation mechanism is to explore potential agreements and then to move the negotiation toward a globally beneficial solution.

3 Agent-based fuzzy constraint-directed negotiation mechanism

Fuzzy constraint-directed approach has been demonstrated in [10, 11, 12, 9, 13] as an effective framework for agent negotiation. During the negotiation, each agent plans the solution by solving its own FCSP, and exchanges the proposals among the agents for resolving the inconsistency of time allocation of activities. As a proposal cannot be accepted by its opponents, a counter-proposal will be made according to the negotiation strategies which consider alternative solutions at the same constraint satisfaction level or offer a solution with a lower constraint satisfaction level. Exchanging proposals and counter-proposals will continue until termination conditions (e.g., achieving a consensus or a failure) are met.

In the proposed negotiation mechanism, the process of proposing a offer/counter-offer in an agent is regarded as the inference process for solving its own FCSP. The process includes the following steps: opponent responsive state evaluation, internal state update, behavioral state determination, set of feasible solutions generation, prospective solution selection and offer/counter-offer generation. The process of proposing an offer $\mathbf{A}^*_{i,j}$ with the FCN $\mathfrak{N}_{i,j} = (\mathcal{U}_{i,j}, \mathbf{X}_{i,j}, \mathbf{C}_{i,j})$ for the task j in project manager agent $\mathfrak{E}_i$ in a negotiation round is described as follows.

While receiving the counter-offer $\mathbf{B}_{i,j}$ over negotiation issues $\mathbf{I}_{i,j} \in \mathbf{X}_{i,j}$ from opponent agents, the opponent responsive state $\sigma_{i,j}$, which indicates the difference between current self-interest and opponents' perspectives, is obtained by

$$\sigma_{i,j} = 1 - ((\widehat{D}_{i,j} - D_{i,j})/(\widehat{D}_{i,j})), \tag{1}$$

where $D_{i,j}$ is the distance between the offer $\mathbf{A}'_{i,j}$ generated in last round and the latest counter-offer $\mathbf{B}_{i,j}$ and $\widehat{D}_{i,j}$ is the basis of distance for normalizing which is obtained at the first negotiation round.

The distance D between an offer $\mathbf{A}$ and a counter-offer $\mathbf{B}$ over set of issues $\mathbf{I}$ is obtained from

$$D = \frac{1}{N_I}\sqrt{\sum\nolimits_{k=1}^{N_I} \mathcal{L}(A^k, B^k)^2}, \tag{2}$$

where $\mathcal{L}$ is a distance measure for two fuzzy set, A^k and B^k is the fuzzy set for the offer $\mathbf{A}$ and counter-offer over issue $I^k \in \mathbf{I}$, respectively.

As none of the counter-offers can be accepted by the project manager agent $\mathfrak{E}_i$, agent $\mathfrak{E}_i$ has to decide the behavioral state to proposal alternative solutions in the current preferred region or offer a solution with a lower constraint satisfaction level. The internal states of agent and the opponent responsive state are involving to determine the behavioral state. In this manner, the internal states include the satisfaction level $\rho_{i,j}$ for task j and degree of tightness in solution space $\delta_{i,j}$. The aggregated satisfaction value $\rho_{i,j}$ for task j indicates the satisfaction level for the current prospective solution $\mathbf{S}'_{i,j}$ and is obtained by

$$\rho_{i,j} = \Psi(\mathbf{S}'_{i,j}) = \frac{1}{N_G}\sum\nolimits_{k=1}^{N_G} \mu_{\mathbf{C}_k}(\mathbf{S}^k_{i,j}), \tag{3}$$

where $\mathbf{S}'_{i,j}$ is the prospective solution of task j obtained in last negotiation round, $\Psi(\mathbf{S}'_{i,j})$ denotes the aggregated satisfaction value for the prospective solution, $\mu_{\mathbf{C}_k}(\mathbf{S}^k_{i,j})$ is the satisfaction degree of the fuzzy constraint $\mu_{\mathbf{C}_k}$ over goal $\mathbf{G}$.

The tightness of solution space $\delta_{i,j}$ indicates the remaining feasible solution space between the aggregated satisfaction value $\rho_{i,j}$ and the acceptable threshold $\epsilon'_{i,j}$ and is obtained by

$$\delta_{i,j} = 1 - (\rho'_{i,j} - \epsilon'_{i,j}), \tag{4}$$

where $\rho'_{i,j}$ is the aggregated satisfaction value, and $\epsilon'_{i,j}$ is the threshold of aggregated satisfaction determined in last negotiation round.

According to the internal state and the opponent responsive state, the level cut τ for the desire of concession $\mathbf{V}$ obtained by

$$\tau = (\mu_\rho(\rho'_{i,j}) \wedge \mu_\delta(\delta'_{i,j}) \wedge \mu_\sigma(\sigma_{i,j}))^{W_c}, \tag{5}$$

where $\mu_\rho(\rho'_{i,j})$, $\mu_\delta(\delta'_{i,j})$, and $\mu_\sigma(\sigma_{i,j})$ denote the desire of concession according to the degree of tightness, aggregated satisfaction and degree of difference, respectively; W_c denotes the weight associated with the desire of concession.

Then, the behavioral state regarded as the threshold of aggregated satisfaction $\epsilon_{i,j}$ is obtained from

$$\epsilon_{i,j} = \epsilon'_{i,j} - \Delta\epsilon = \epsilon'_{i,j} - \mathcal{D}(\mathbf{V}_\tau), \tag{6}$$

where $\epsilon'_{i,j}$ is the threshold of aggregated satisfaction obtained from last negotiation round, and $\Delta\epsilon$ is the concession value transformed from the fuzzy set of the desire of concession $\mathbf{V}$ with τ level cut. $\mathcal{D}$ is the defuzzification method.

To evaluate and generate the offer and counter-offer, an agent has to plan its prospective solution first. Each agent can only plan the prospective solution from its individual area of interests limited by the threshold of aggregated satisfaction $\epsilon_{i,j}$. That is, a set of feasible solution $\mathfrak{P}_{i,j}$ for task j in agent $\mathfrak{E}_i$ can be defined as

$$\mathfrak{P}_{i,j} = \{\mathbf{S}_{i,j} \mid (\mathbf{S}_{i,j} \in_\alpha \mathbf{\Pi}_{i,j}) \wedge (\Psi(\mathbf{S}_{i,j}) \geq \epsilon_{i,j})\}, \tag{7}$$

where ${}_\alpha\mathbf{\Pi}_{i,j}$ is the solution set with α level cut, $\Psi_{i,j}(\cdot)$ is the aggregated satisfaction value, and $\epsilon_{i,j}$ is the threshold of aggregation satisfaction degree of objectives.

Given the counter-offer $\mathbf{B}_{i,j}$ and the feasible solution set $\mathfrak{P}_{i,j}$, the prospective solution $\mathbf{S}^*_{i,j}$ is generated by solution selection method

$$\mathbf{S}^*_{i,j} = \arg_{\mathbf{S}}(\max_{\mathbf{S}\in\mathfrak{P}_{i,j}} \mathcal{H}(\mathbf{S}, \mathbf{B}_{i,j})), \tag{8}$$

where $\mathcal{H}$ is a utility function to evaluate the appropriateness between the feasible schedule $\mathbf{S}$ and the counter-offer set $\mathbf{B}$ in

ISBN: 978-989-95079-6-8

which the utility function $\mathcal{H}$ can be defined by

$$\mathcal{H}(\mathbf{S},\mathbf{B}) = \frac{1}{N_I}\sqrt{\sum_{k=1}^{N_I}(\mathcal{P}(S_k)^{w_p} \wedge \mathcal{S}(S_k,B_k)^{w_s})^2}, \quad (9)$$

where $\mathcal{P}$ is a satisfaction function over the issue I_k, and $\mathcal{S}$ is a similarity function (a distance measure) between the solution $\mathbf{S}$ and the counter-offer $\mathbf{B}$; w_p, and w_q denote the weight associated with the satisfaction and the similarity of the solution of task, respectively.

For the task j, given the feasible solution set $\mathfrak{P}_{i,j}$ and the prospective solution $\mathbf{S}^*_{i,j}$ which involves a set of values $\{\mathbf{S}^1_{i,j},\mathbf{S}^2_{i,j},...,\mathbf{S}^{N_I}_{i,j}\}$ for N_I issues, the offer $\mathbf{A}^*_{i,j}$ for task j is a tuple of fuzzy set $\{A^1_{i,j},A^2_{i,j},...,A^{N_I}_{i,j}\}$ in which each fuzzy set $A^k_{i,j}$ for issue $I^k_{i,j} \in \mathbf{I}_{i,j}$ is the marginal particularized possibility distribution

$$A^k_{i,j} = Proj_{X^k_{i,j}}(\mathfrak{P}_{i,j} \cap \overline{\Pi}_{X^1_{i,j}} \cap ... \cap \overline{\Pi}_{X^{k-1}_{i,j}} \cap \overline{\Pi}_{X^{k+1}_{i,j}} \cap ... \cap \overline{\Pi}_{X^{N_I}_{i,j}}), \quad (10)$$

where $\overline{\Pi}_{X^k_{i,j}}$ is the cylindrical extension of $\Pi_{X^k_{i,j}}$ in the space $(\mathbf{X}^1_{i,j},...\mathbf{X}^{N_I}_{i,j})$, $\Pi_{X^k_{i,j}} = \mathbf{S}^k_{i,j}$, and $\mathbf{X}^k_{i,j}$ is the object of issue $I^k_{i,j}$ for task j.

Meanwhile, when project manager agent $\mathfrak{E}_i$ receives a counter-offer $B_{i,j} = \{B^1_{i,j},B^2_{i,j},...,B^{N_I}_{i,j}\}$ for the task j, the opponent agents' preferred solution $\widehat{\mathfrak{P}}_{i,j}$ can be obtained by

$$\widehat{\mathfrak{P}}_{i,j} = (\overline{B}^1_{i,j} \cap ... \cap \overline{B}^{N_I}_{i,j}), \quad (11)$$

where $\overline{B}^k_{i,j}$ is the cylindrical extension of $B^k_{i,j}$ in the space $\mathbf{X}_l$. Each element $\widehat{\mathfrak{P}}$ represents a solution preferred by opponent agents and the membership degree of each element represents the acceptability of solution for all opponent agents.

The project manager agent $\mathfrak{E}_i$ will accept the counter-offer $B_{i,j}$ proposed by its opponent as an agreement if

$$(\mathfrak{P}_{i,j} \cap \widehat{\mathfrak{P}}_{i,j}) \neq \{\}, \quad (12)$$

The negotiation process for task allocation will be terminated when the project manager agent reach an agreement with one of contractor agents or the project manager agent or all of their contractor agents withdraws the negotiation. A project manager agent will terminate the negotiation when each of tasks in project is assigned to a contractor agent or the project manager agent withdraws the negotiation while failing to reach agreement. And the supply chain planning/scheduling process will be terminated when all of project manager agents are termination.

4 Experiments

To demonstrate the utility of the proposed model for project planning and scheduling problem in supply chain, the experiments are meant to compare AFCN with centralized heuristic (CTR) [14], conventional CNP (CNP) [15], modified CNP (MCNP) [4], market-based auction CNP (MA-CNP) (i.e., extended from CNP) and market-based auction MCNP (MA-MCNP) (extend from MCNP) (i.e., extended from MCNP) in terms of the number of project managers over makespan, total operating cost, and computational time.

The experiment was implemented on a Pentium M PC with 0.8 GHz running windows XP and 256 MB RAM. In the experiment, each project has a linear sequence of tasks, and each task is then specified by its precedence constraints, required resources, processing time and tardiness cost. Each supplier is specified by its unique capacity and processing cost. The total operating cost is the sum of the total processing cost and tardiness cost of each project. For simplicity, the number of tasks per project is identical at five, alternative suppliers for a task are identical at three, and 50 instances are generated for each number of project managers. Figures 2 to 4 show the performance comparisons over makespan, total operating cost, and computational time, respectively.

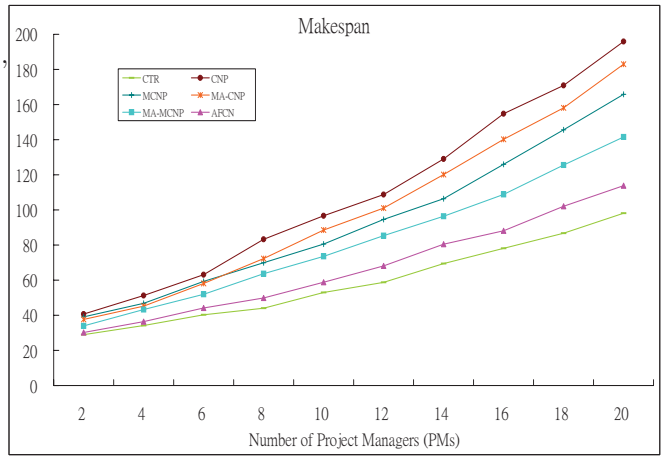


Figure 2: Makespan of various negotiation mechanisms over a number of project managers.

The conventional CNP supports one PA to assign a task at a time and another PA can start only after the PA finishes selecting CAs for all its tasks. In this form of negotiation model, a large number of slacks will be generated since that the conflicts among PAs cannot be coordinated. Each PA's performance highly depends on the order of contracting.

Instead of one PA assigning a task, MCNP supports multiple PAs to contract tasks to multiple CAs simultaneously. In the manner, PAs propose the whole time window of task and price to CAs. CAs resolve the conflicts among tasks according to the urgency of task be completed and the price of the task. It reduces the possibility of violating the due date of the PAs and decreases the tardiness when a project is delayed.

Since the simplicity of negotiation protocol, the single-shot negotiation models (CNP and MCNP) have the best time efficiency in Figure 4. With very limited information sharing and myopic decision-making, however, the global performances can easily get trapped at local optima and are highly unstable. It also can be observed in Figures 2 and 3, these models worse results than other approaches over makespan and total operating cost.

For reflecting the service preference in PA and the task contention in CA, MA-CNP and MA-MCNP incorporate the market-based auction mechanism [16] which is characterized

ISBN: 978-989-95079-6-8

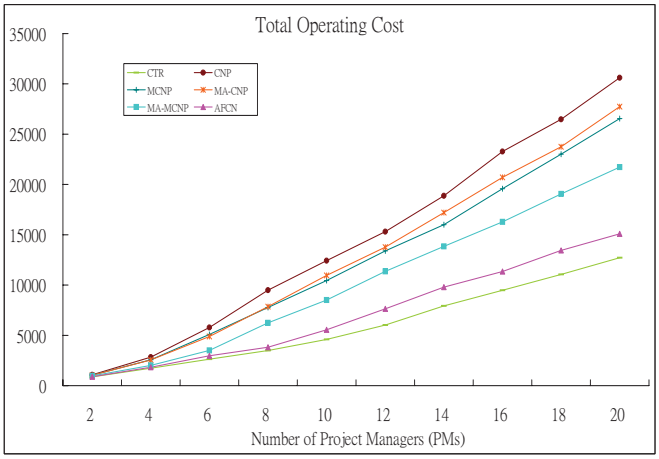


Figure 3: Operating cost of various negotiation mechanisms over a number of project managers.

by iterative bidding among agents. Agents will adjust the prices of bid according to the direction of surplus and deficit of demand. Task allocation processing in CA can be adjusted iteratively according to time boundary and the updating price of tasks. In this way, MA-MCNP with iterative bidding have better makespan and total operating cost than the single-shot negotiation models. However, due to the characteristic of protocol in MA-CNP, MA-CNP even worse than MCNP over makespan and total operational cost when the number of project managers is larger than 8 in Figure 2 and 3, respectively.

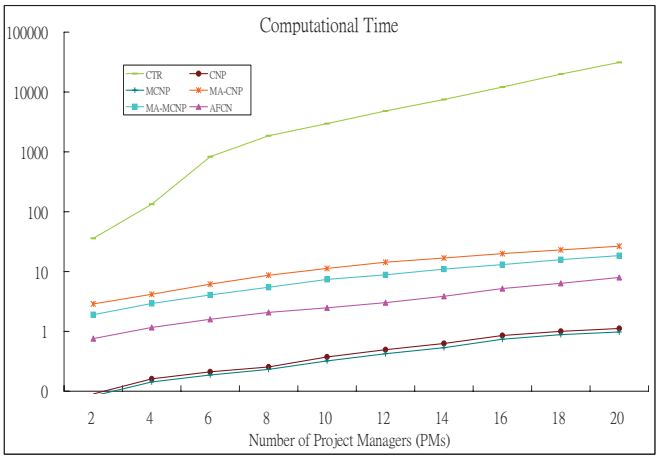


Figure 4: Computational time of various negotiation mechanisms over a number of project managers.

However, the price of task proposed by CA indicates the contention degree of the desired time slots of task, but not indicates where the contention can be resolved. PA might blindly increase price but the schedule of task can not be improved (overpricing). Besides, sharing full of task's time boundary in PA is not practical in reality.

In AFCN, PA generates the offer not only considers how much suppliers can provide this service, but how pressing the task be finished. Instead of the bid information in the market-based mechanism, the negotiation information in the AFCN varies not only depending on the particular combination of task demands/resource contracts, but the objectives and preferences of agent. Both of PA and CA proposes offers/counter-offers which involve the current feasible time slots of task with preference and price. The time slots with preference in offer/counter-offer indicate not the current time boundary of task, but the time window for conflict-free. According to time slots with preference from PA, CA can arrange the task allocation efficiency and can response the conflict in counter-offer directly. Meanwhile, CA generates the counter-offers considering not only the tasks of PAs are satisfied but the profit obtained in the transaction. Based on the counter-offers from CA, PA can avoid requiring the high contention area in CA or can abort lower quality service of suppliers. Thus, PA can bid the desired time slots to preferred suppliers with controllable budget.

In Figure 2 and 3, it can be observed that the AFCN has superior performance than other approaches in makespan and total operating cost. Meanwhile, AFCN performs better in time efficiency than that of market based mechanism. Additionally, as the problem size is increasing, the performance improved by the AFCN grows more significantly.

5 Conclusions

We have presented an agent-based fuzzy constraint-directed negotiation protocol for project planning and scheduling in supply chain. Constraint modeling gives a more direct fusion to the real-world problem descriptions and the impreciseness of knowledge in supply chain can easily be represented by fuzzy constraints with the levels of consistency as well. Experimental results suggest that the AFCN indeed can provide a practical and efficient framework for supply chain planning and scheduling.

References

[1] B. M. Beamon. Supply chain design and analysis: Models and methods. *International Journal of Production Economics*, 55(3):281–294, 1998.

[2] F. R. Lin, H. C. Kuo, and S. M. Lin. The enhancement of solving the distributed constraint satisfaction problem for cooperative supply chains using multi-agent systems. *Decision Support Systems*, 45(4):795–810, 2008.

[3] T. Calosso, M. Cantamessa, and M. Gualano. Negotiation support for make-to-order operations in business-to-business electronic commerce. *Robotics and Computer-Integrated Manufacturing*, 20(5):405–416, 2004. Cited By (since 1996): 7.

[4] J. S. K. Lau, G. Q. Huang, K. L. Mak, and L. Liang. Agent-based modeling of supply chains for distributed scheduling. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 36(5):847–861, 2006.

[5] Y. Aviv. The effect of collaborative forecasting on supply chain performance. *Management Science*, 47(10):1326–1343, 2001.

ISBN: 978-989-95079-6-8

[6] P. J. Byrne and C. Heavey. The impact of information sharing and forecasting in capacitated industrial supply chains: A case study. *International Journal of Production Economics*, 103(1):420–437, 2006.

[7] G. P. Cachon and M. Fisher. Supply chain inventory management and the value of shared information. *Management Science*, 46(8):1032–1048, 2000.

[8] L. Yan and T. G. Woo. Information sharing in a supply chain with dynamic consumer demand pattern. In *Proceedings of the Hawaii International Conference on System Sciences*, volume 37, pages 1143–1152, 2004.

[9] K. R. Lai. *Fuzzy Constraint Processing*. PhD thesis, NCSU, Raleigh, N. C., 1992.

[10] K. Y. Cai. Robustness of fuzzy reasoning and δ-equalities of fuzzy sets. *IEEE Transactions on Fuzzy Systems*, 9(5):738–750, 2001.

[11] D. Dubois, H. Fargier, and H. Prade. Propagation and satisfaction of flexible constraints. In *Fuzzy Sets, Neural Networks, and Soft Computing*, pages 166–187. Van Nostrand Reinhold Company, New York, 1994.

[12] D. Dubois, H. Fargier, and P. Fortemps. Fuzzy scheduling: Modelling flexible constraints vs. coping with incomplete knowledge. *European Journal of Operational Research*, 147(2):231–252, June 2003.

[13] J. Moura-Pires and H. Prade. Specifying fuzzy constraints interactions without using aggregation operators. In *Proceedings of the Ninth IEEE International Conference on Fuzzy Systems*, volume 1, pages 228–233, 2000.

[14] C. G. Li, J. B. Yu, and Liao. X. F. Fuzzy tabu search for solving the assignment problem. In *Communications, Circuits and Systems and West Sino Expositions, IEEE 2002 International Conference on*, volume 2, pages 1151–1155, June-1 July 2002.

[15] R. G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, C-29(12):1104–1113, 1980.

[16] M.P. Wellman, W.E. Walsh, P.R. Wurman, and J.K. MacKie-Mason. Auction protocols for decentralised scheduling. *Games and Economic Behavior*, 14(5):429–439, 2001.

ISBN: 978-989-95079-6-8

Learning and Classification of Events in Monitored Environments

J. Albusac[1] J.J Castro-Schez[2] L.M Lopez-Lopez[2] D. Vallejo[2] L. Jimenez-Linares[2]

1.E.U.P.A, University of Castilla-La Mancha
Almaden, Spain
2.Escuela Superior de Informatica, University of Castilla-La Mancha
Ciudad Real, Spain
Email: {javieralonso.albusac, josejesus.castro, luis.jimenez, lorenzomanuel.lopez, david.vallejo}@uclm.es

Abstract—
This paper presents a prototype system to automatically carry out surveillance tasks in monitored environments. This system consists in a supervised machine learning algorithm that generates a set of highly interpretable rules in order to classify events as normal or anomalous from 2D images without needing to build a 3D model of the environment. Each security camera has an associated knowledge base which is updated when the environmental conditions change. To deal with uncertainty and vagueness inherent in video surveillance, we make use of Fuzzy Logic. The process of building the knowledge base and how to apply the generated sets of fuzzy rules is described in depth for a virtual environment.

Keywords— Automated Video Surveillance, Visual Information Analysis, Machine Learning, and Fuzzy Logic.

1 Introduction

The problem of intelligent surveillance deals with the perception, interpretation, and identification of activities and situations that occur in monitored environments [1]. A typical surveillance scenario consists of a set of CCTV cameras deployed on different places in order to perform explicit surveillance on behalf of people and information storage to carry out forensic analysis if needed [2]. When a security guard watches an environment by means of video cameras, he is able to detect if something is going wrong. In other words, he perceives video events and classifies them as normal or anomalous. If he detects an anomalous behaviour or event, he makes the appropriate decisions to solve the problem as soon as possible. However, the system depends on the human component to classify events and to continuously pay attention to the video stream, which is very tiring and, therefore, error-prone [3]. This is our motivation to face the design of a system able to classify behaviours, that is, to propose a method for automating this task in different domains and scenarios.

Events in real environments can be classified as simple or composite. Really, a composite event is a sequence of simple events which are temporally related [4, 5]. However, the analysis of object characteristics and spatial properties may be enough to understand simple events, i.e, without needing a temporal analysis. In this context, there are many elements which can be learned and identified from data sensors. For instance, the object physics characteristics to determine its class or object type [6], trajectories followed by moving objects [7], entry/exit areas [8], allowed proximity relationships between objects and areas, allowed speed for each type of object, etc.

On the other hand, it is necessary to study how to deal with the uncertainty in real-time to carry out surveillance. Normally, an artificial surveillance system cannot totally ensure what is happening from data sensors in most cases. Uncertainty, imprecision, and vagueness are frequently present when these systems try to solve real world problems. For this reason, one of the main problems of this type of systems is the high number of false alarms due to incorrect interpretations.

Several authors have addressed these issues in the literature. Foresti et al. [9] proposed a visual-based surveillance system for real-time event detection and classification, which is based on adaptive high order neural trees. This system carries out object classification, object tracking, and event recognition for understanding normal, suspicious, and dangerous events in parking lots. Buxton and Gong [10] also provided solutions to the problem of event recognition. They proposed techniques based on Bayesian networks for interpreting traffic situations in dynamic scenes. Remagnino et al. [11] described events in a surveillance scenario by using a Bayesian classifier instead of the Hidden Markov Model. The different techniques proposed in this work were employed to model the common event behaviours in car park. On the other hand, fuzzy logic [12, 13] may provide another interesting approach for dealing with the same problem [14]. In [15], a prototype fuzzy system for describing human activity in natural language is described. This description is made by analysing the possible relations among objects in a monitored environment. To do that, the linguistic labels and the fuzzy rules are defined by an expert to classify people behaviour.

The crafting of detailed scene models may provide an effective means of interpreting situations and detecting anomalous behaviours in a static scene, but it is not an appropriate approach in dynamic scenes where the environmental conditions change over time [7]. This is why in this work we use a fuzzy-based machine learning algorithm to generate a set of highly interpretable set of rules to classify real-time events. To do that, the system analyses spatial data from 2D images obtained by cameras in outdoor/indoor environments, and it generates sets of highly interpretable fuzzy rules. The system analysis situations in a similar way as human beings do, that is, it obtains conclusions without needing precise data. For instance, a person does not need to know the exact speed or the absolute position of an object to determine if it is moving fast on the zone in which it is located.

The remainder of this paper is organised as follows. Section 2 describes the problem statement and overviews the processes required to build the knowledge base of a camera and how to apply this knowledge to classify events. Section 3 de-

ISBN: 978-989-95079-6-8

scribes in depth the machine learning algorithm employed to build the different sets of fuzzy rules that compose the knowledge base. Section 4 studies how to apply the system in a well-defined environment. Finally, conclusions are presented in Section 5.

2 Problem Statement

As previously described, one of the main objectives of advanced surveillance systems is to interpret events in an environment from data sensors. Besides, events not only must be identified, but the system must be also able to classify events as normal or abnormal in order to make adequate decisions. This task covers several research areas, we attack two of them: i) how to build the surveillance knowledge base in a concrete scenario and ii) how to apply this knowledge to classify events as normal or abnormal.

On the other hand, a simple event can be defined as a concrete action that occurs at a time instant, and an anomalous simple event may be enough to activate an alarm in a monitored environment. The method proposed in this work determines the normality of a scene according to this type of events. Examples of anomalous simple events are as follows: a vehicle moving too fast in a concrete time instant or an object located in a forbidden zone. We characterise our surveillance domain by means of the following assumptions:

- Video stream is obtained from cameras placed on certain locations of the environment.
- Each camera has an own knowledge base.
- Every camera is fixed, that is, vertical or horizontal movements and zoom imply a new generation of the surveillance knowledge base.
- Video data can be imprecise.

Each video camera has its own knowledge base, which is composed of three set of fuzzy rules, and each set is generated by an independent training set:

1. Set of rules to determine the object's class. This information is critical when classifying events as normal or abnormal because it defines how objects should normally behave.
2. A second set of rules to determine the object's speed taking into account the object's class and its motion vectors obtained in the segmentation process.
3. Finally, a set of rules to infer whether a situation is normal or anomalous according to the object's class, speed, and the areas in which the object could be located.

To generate the first set of fuzzy rules, we employ the size and the position of the objects in the environment from 2D images (frames), and we take into account the camera view point. Thus, the system may learn, for instance, that vehicles are large or very large size objects when they are close to the camera, and small or medium size objects when they are far from the camera. The segmentation algorithm [16] used in this work determines the set of macroblocks in each frame for every object that appears in the scene. A macroblock is the basic unit in a MPEG stream and it is an area of 16 by 16 pixels in which the motion vectors are stored. The displacement between two macroblocks in different frames gives the motion vector and it specifies a distance and a direction. Each object is represented by means of an ellipse that involves the set of macroblocks. As it will be seen further (Section 4), the parameters of the ellipse are used to determine the location and the size of an object.

On the other hand, in order to generate the rules used to determine the object's speed, the position (centre of the ellipse), the displacement of the object between consecutive frames (distance between the centres of the ellipses), and the object class are used as input variables, i.e, each sample of the training set takes a value for every variable. Movements done by people and vehicles may imply different speed. A medium displacement for one person may involve high speed and for a car may involve low or medium speed. The main goal of this stage is to learn to interpret the object speed depending on their sizes and displacements by taking into account the camera point of view. Finally, the last set of fuzzy rules is generated from a training set where each sample represents the situation of an object in a concrete time instant, i.e, its class, speed, and the spatial relations with the zones or areas of the environment.

Once the rules have been generated and the knowledge base has been built, it can be used to classify simple events. The system analyses frames of a MPEG video stream, evaluating where are the objects located in the environment (2D position) and what are their size. From this information, the system obtains the class of each moving object. In this point, the surveillance system possibly knows where are the objects in each moment and what are their classes. Next, the system uses this information and the second set of rules to determine the speed of each object. Finally, the system determines the areas in which each object could be located and the information previously obtained (class and speed) to classify the simple events as normal or anomalous. In other words, the classification process can be defined as the chaining of three sets of fuzzy if-then rules in which the knowledge generated by a set is employed by the next one.

3 Suggested Machine Learning Algorithm

The machine learning algorithm described in this section is based on the work proposed by Castro *et al.* [17]. This method learns a set of maximal structure fuzzy if-then rules from a set of training data instances. This choice has been motivated by the following reasons: (a) The use of fuzzy logic provides a well-defined mathematical framework to deal with the uncertainty and vagueness inherently contained in the visual information used as input data. (b) This algorithm represents the inferred knowledge by means of a set of fuzzy if-then rules expressed in terms of linguistic variables. As a result, we obtain rules that are easily comprehensible by security guards, who are watching the monitors. Thus, an user can easily analyse why an alarm was activated. (c) As a supervised learning approach, once the set of rules has been inferred from the training data set, the algorithm does not require a significant amount of time to classify new data instances.

Basically, Castro's algorithm works in two phases. In the first one, each example of a training set is converted into a

ISBN: 978-989-95079-6-8

specific rule which describes how to act in a concrete situation, that is, how this example must be classified. Therefore, there will be so many particular rules as examples. In a second phase, the algorithm generalises these particular rules in order to act in a wide range of possibilities, that is to say, many examples of a class can be correctly classified by using one generalised rule. To do that, an amplification process extends the particular rules covering situations of the space of possible situations, which have not been covered before.

Formally, the original algorithm starts out the training phase from a set of data instances $\Theta = \{e_1, \ldots, e_m\}$. Each $e_i = ((x_{i1}, \ldots, x_{in}), y_j)$ is a data example which is conveniently described by means of the set of input variables $V = \{v_1, \ldots, v_n\}$ and one output variable y_j, which is called the *class* of e_i. The elements $(x_{i1}, \ldots, x_{in})$ are the concrete values that the example e_i takes for each variable $v_k \in V$. Besides, there is a domain definition DDV_i for each variable in V, being $DDV = \{DDV_1, \ldots, DDV_n\}$ the set of all domains. Each $DDV_i = \{L_1, \ldots, L_p\}$ is composed of a set of linguistic labels which correspond to the fuzzy sets represented by means of trapezoidal functions as specified in expression (1).

$$\prod(u; a, b, c, d) = \begin{cases} 0 & u < a \\ \frac{(u-a)}{(b-a)} & a \leq u < b \\ 1 & b \leq u \leq c \\ \frac{(d-u)}{(d-c)} & c < u \leq d \\ 0 & u > d \end{cases} \quad (1)$$

In this way, each DDV_i verifies the following properties:

1. $\forall L_x \in DDV_i, height(L_x) = 1$
2. $\forall L_x, L_y \in DDV_i, nucleus(L_x) \cap nucleus(L_y) = \emptyset$
3. $\forall x \in X_i, \sum_{j=1}^{|DDV_i|} \mu_{L_j}(x) = 1$, being X_i the domain where v_i is defined.

Besides, the inferred knowledge is represented as a set of fuzzy if-then rules with the following structure:

$$\textbf{if } v_0 \textit{ is } ZD_0 \wedge \ldots \wedge v_n \textit{ is } ZD_n \textbf{ then } y_j \quad (2)$$

where $v_i \in V$ and $ZD_i \subseteq DDV_i$ is a subset of the linguistic labels defined in DDV_i for the variable v_i.

The original algorithm converts each $e_i \in \Theta$ into the fuzzy domain, according to the values $(x_{i1}, \ldots, x_{in})$ that e_i takes for each $v_i \in V$ and their corresponding $DDV_i \in DDV$. For this purpose, each $L_k \in DDV_i$ has associated a function $\mu_{L_k} : X_j \rightarrow [0, 1]$, being X_j the domain where the variable v_j takes its values (i.e $\Re, \aleph$, an interval $[a, b]$, a finite set A, etc). Thus, every $e_i \in \Theta$ is converted into $e'_i = ((L_{1x}, \ldots, L_{nz}), y_j)$ such that L_{jk} is the label which matches best, according to $L_{jk} = max\{\mu_{L_{jk}}(x_j)\}$ of the DDV_j. Each e'_i is an initial rule which belongs to the set of *initial rules*. Second, if some rule of the initial set does not subsume in any rule of the final set, the algorithm proceeds to amplify the rule. A rule R_i can be amplified to $R_{i'}$: **if** v_0 *is* $ZD_{i'0} \wedge \ldots \wedge v_n$ *is* $ZD_{i'n}$ **then** y_p if there is no rule R_j : **if** v_0 *is* $ZD_{j0} \wedge \ldots \wedge v_{jn}$ *is* ZD_n **then** y_q in the set of initial rules that verify $ZD_{jk} \subseteq ZD_{i'k}$ and $y_p \neq y_q$. In other words, an amplification is possible whenever there is no counterexample that conflicts with the amplified rule. This makes possible for this method to generate a set of rules that are as general as possible. A description in detail of the original algorithm by Castro *et al.* can be found in [17].

However, this algorithm may cause troubles when applied in the surveillance context due to the production of rules that are over-generalised. With over-generalised we refer to the tendency of the final rules to cover spaces for which there are no counterexamples. For example, let us suppose three initial rules for classifying a moving object according to its size by taking into account the distance from the camera:

- R_i: **if** $size$ is {SMALL} $\wedge$ $distance$ is {MEDIUM} **then** y_1 is people
- R_j: **if** $size$ is {MEDSMALL} $\wedge$ $distance$ is {MEDIUM} then y_1 is motorbike
- R_k: **if** $size$ is {MEDBIG} $\wedge$ $distance$ is {MEDIUM} then y_1 is car

all of them contained in the set of initial rules. Let DDV_{size} = {VSMALL, SMALL, MEDSMALL, MEDBIG, BIG, VBIG}, if we try to amplify the rule R_i to achieve $R_{i'}$, according to the variable **size**, it could result in:

- $R_{i'}$: **if** $size$ is {VSMALL, SMALL, BIG, VBIG} $\wedge$ $distance$ is {MEDIUM} **then** y_1 is people

Note that $R_{i'}$ means that if an object is located at a *medium* distance from the camera and its size is *very small*, *small*, *big*, or *very big*, it certainly belongs to the class *people*. However, R_j and R_k mean that if an object is located at a *medium* distance from the camera and its size is *medium small* or *medium big*, the object is a *motorbike* or a *car*, respectively. Obviously, $R_{i'}$ lacks of any sense and it can cause problems because of the misclassification of objects. This happens because when amplifying R_i to $R_{i'}$, the labels $VSMALL$, BIG, and $VBIG$ are added to $R_{i'}$ as there are no counterexamples in the set of rules that conflict with it. Moreover, this type of counterexamples might never exist in the training set because it is possible that cars and motorbikes have never a big or very big size, when they are located at a medium distance from the camera.

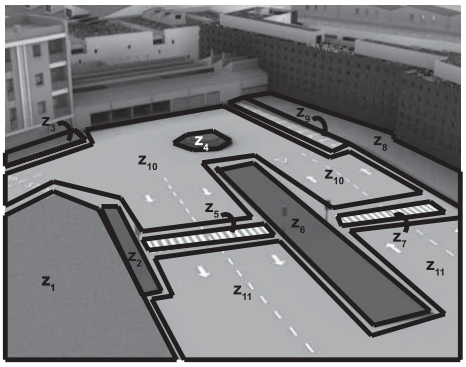
In order to solve this issue, we have modified the original algorithm to restrict the amplification of rules. We are interested in amplifying a rule only if the last label added to the rule in the last amplification step and the label being considered to be added are not too far from each other. Thus, we need to use a measure of separability s to estimate the dissimilarity between linguistic labels L_i and L_j (with $L_i < L_j$), which are expressed in terms of fuzzy sets:

$$s(L_i, L_j) = \frac{(b_{Lj} - c_{Li}) + (a_{Lj} - d_{Li})}{2} \quad (3)$$

being $(a_{Li}, b_{Li}, c_{Li}, d_{Li})$ and $(a_{Lj}, b_{Lj}, c_{Lj}, d_{Lj})$ the trapezoids that define the fuzzy sets L_i and L_j respectively, and verifying that $d_{Li} < a_{Lj}$. If $d_{Li} = a_{Lj}$ or $d_{Li} > a_{Lj}$, then $s(L_i, L_j) = 0$, as the area of the fuzzy set $s(L_i, L_j)$ is not significant enough.

The separability threshold is empirically calculated by observing how the algorithm selects the different values of percentage of the maximum separability between the fuzzy sets

ISBN: 978-989-95079-6-8



Zone Classification	
z_1	Garden
z_2	Sidewalk
z_3	Garden
z_4	Traffic Circle
z_5	Pedestrian Crossing
z_6	Sidewalk
z_7	Pedestrian Crossing
z_8	Garden
z_9	Parking
z_{10}	Road
z_{11}	Road

Figure 1: Definition and classification of areas for the analysed scene.

of a given linguistic variable. If we apply the expression (3) with L_i being the first label and L_j the last label of a linguistic variable, we obtain the maximum possible distance value for that variable. Then, it is possible to choose the 25% or the 10% of such separability as the threshold for amplifying a rule. Note that when the maximum allowed separability is 100%, the new version of the algorithm works as the original one.

Next, the steps of the modified algorithm are described in detail.

1. Convert each training example into an initial fuzzy rule (translation into the fuzzy domain). Each element $e_i \in \Theta$ is translated into a fuzzy rule in which the value of each input variable is represented by means of a linguistic label. This step creates the set of *initial* rules.
2. Take a rule from the set of initial rules.
3. Try to subsume the taken rule in some rule of the set of definitive rules. If that happens so, ignore the taken rule and go back to step 2.
4. If the taken rule does not subsume in any rule of the definitive set, try to amplify it. For each variable:
   (a) For each unconsidered label:
      i. Try to amplify the rule. If it is not possible, go to step 4.a; otherwise, proceed to step 4.a.ii. One rule can be amplified only if:
         A. There is no R_j : **if** v_0 *is* ZD_{j0} $\wedge$ ... $\wedge$ v_n *is* ZD_{jn} **then** y_q in the set of initial rules that $ZD_{jk} \subseteq ZD_{i'k}$ and $y_p \neq y_q$ (we maintain the constraint of the original algorithm).
         B. The separability between the last label added to the rule and the label being considered for amplifying the rule does not exceed a separability threshold.
      ii. Amplify the rule and include it in the set of definitive rules.
5. If there are still unconsidered rules in the initial set of rules, go to step 2. Otherwise, **END**.

4 Applying the Suggested Algorithm to an Example

The proposed machine learning algorithm has been tested on the virtual environment shown in Figure 2 due to the following reasons: (a) the difficulty of obtaining a wide range of real video scenes with anomalous situations, (b) the possibility of generating all abnormal situations as needed without attempting on people security, and (c) the freedom of changing camera position with no configuration cost.

Figure 1 shows the scenario used in this work, which represents a typical urban environment composed of buildings, roads, gardens, traffic signals, and so on. There are also pedestrian areas in which vehicles are not allowed to drive. In the same way, there are only vehicle areas in which pedestrians should not walk.

In our particular application, there are three training sets: Θ_1, Θ_2, and Θ_3, one for each phase of the learning process. Each training set is composed of a set of examples where every one of them e_i is made up from a set of features extracted from the 2D images captured by a video camera, such as the position of a moving object in a frame or the size of the ellipse that contains it. The algorithm described in Section 3 is performed for every training set Θ_i, in order to obtain a set of fuzzy rules. On the other hand, there are three output variables: the class of the moving object (y_1), its speed (y_2), and whether its behaviour is normal or anomalous (y_3).

4.1 Learning Process

As described before, the first step consists in learning the set of rules to determine the object class. Determining the object class is crucial to know if its behaviour in a monitored environment is normal. Each one of these classes has a set of norms or rules. If an object meets the norms associated to its class, then its behaviour will be considered as normal. To do that, we first need to build a training set in which each sample is defined in the following way: $e_i = ((x_{i1}, \ldots, x_{in}), y_j)$, being $(x_{i1}, \ldots, x_{in})$ the variable values used to learn and y_j the output variable. The variables employed in this first phase are $((X_{pos}, Y_{pos}, R_h, R_v)\, y_1)$. X_{pos} represents the moving object horizontal position and Y_{pos} represents the moving object vertical position. The values of X_{pos} and Y_{pos} comes from the coordinate x, y of the ellipse central point that encloses the object. On the other hand, R_h determines the horizontal size measured in pixels of the ellipse that encloses the object and R_v the vertical size measured also in pixels. Finally, y_1 is the output variable, which can take the following values: $\{pedestrian, vehicle\}$.

Each sample refers to a moving object detected in the segmentation process, which has been marked with an ellipse. Precisely, the ellipse parameters (central point coordinates, horizontal radius, and vertical radius) are used to generate the rules in this step. The goal is to determine whether a certain object is a pedestrian or a vehicle depending on its size and position from the camera point of view. Distinguishing be-

ISBN: 978-989-95079-6-8

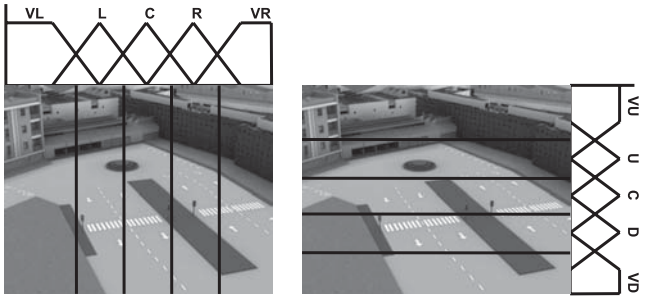


Figure 2: Vertical-horizontal scene division from an image captured by a surveillance camera.

tween people and vehicles depending on size and position is feasible because the difference between them is notable independently of the area in which they are located. On the other hand, to determine the zone in which an object is placed, a vertical-horizontal division as shown in Figure 2 is carried out. In this environment, the far zones from camera position, where objects may move, are located in $X_{pos} = \{\text{centre, right}\}$ and $Y_{pos} = \{\text{up}\}$. However, the close zones from camera position are located in $X_{pos} = \{\text{very left, left, centre}\}$ and $Y_{pos} = \{\text{very down, down}\}$.

An example of a fuzzy rule generated in this stage is as follows:

R : if $Xpos$ is {very right} and $Ypos$ is not {very up, up} and Rh is {very small, small} and Rv is {very small, small} then y_1 is pedestrian

Once applied the first stage, the next step consists in determining the object speed depending on its position and movement (measured in pixels) between the current and previous frame. We consider that two frames per second for the studied environment is enough for detecting anomalous situations. For every object class and zone there is a range of allowed speed values, which should be met by every object. Therefore, the speed study is interesting for the analysis of behaviours in monitored environments. In order to learn the second set of rules, we use input examples with the following form: $((X_{pos}, Y_{pos}, y_1, Mov)\ y_2)$, being Mov the movement of the object measured in pixels between two consecutive frames, and y_2 the output variable which can take the possible following values: $\{slow, normal, fast\}$.

An example of a rule generated in this step is the following:

R : if $Xpos$ is {very left, left} and $Ypos$ is {very down, down} and y_1 is {pedestrian} and Mov is not {medium, low, very low} then y_2 is slow

Finally, the last step in the general learning process consists in generating a set of rules that allows the surveillance system to detect if an object behaves normally or anomalously depending on the object class, its speed and the zones in which it could be located. As in previous stages, a training set is built for the rule generation. Every input sample is represented in the following way: $((y_1, y_2, Garden, Sidewalk, Trafficcircle, Parking, PedestrianCrossing, Road)\ Y_3)$, where every kind of zone represents an input variable whose value sets the intersection grade between the object and the type of zone. Finally, y_3 is the output variable whose values can be $\{normal\ situation, anomalous\ situation\}$.

An example of two rules generated in this final step is:

R : if y_1 is {pedestrian} and y_2 is {slow, normal} and $garden$ is {very high} and $road$ is {out, very low, low} then y_3 is normal situation

R : if y_1 is {pedestrian, vehicle} and y_2 is {fast} and $road$ is {low, medium, high} then y_3 is anomalous situation

4.2 Classification process

Once the learning process is done, the set of rules is available to determine, in each frame, the normality of an object behaviour. To carry out the classification process, first it is necessary to obtain a set of low-level objects data. Next, we expose the description of such data and how to obtain it.

For each frame, the system performs a segmentation process to detect every moving object, which is then bounded within a ellipse. The parameters of the ellipse are employed to determine the position and the size of the bounded object (X_{pos},Y_{pos},R_h,R_v). If an object is moving, the position of its bounding ellipse in the current frame changes respecting to the previous frame. Thus, the movement of an object is calculated through the distance between the origin of the previous and the current ellipses.

To complete the low level data acquisition of the objects, the system performs a proximity analysis of every tracked object to the areas of the environment which are defined according to the camera point of view. The goal of this analysis is to learn in which areas can be located a concrete object class. Thus, an expert is required to define such areas through the definition of their corresponding bounding polygons and the class which every area belongs to. Figure 1 shows the definition and classification of areas performed for the analysed scene.

The next step is to study the intersection between the objects bounding ellipses and the defined areas to learn in which extent an objects is on an area. For this purpose, the system calculates the amount of points in common in an object's bounding ellipse and the polygon defining an area. To do that, we use the following algorithm [18] represented in the programming language C:

```
int pnpoly(int nvert, float *vertx, float *verty,
           float testx, float testy)
{
  int i, j, c = 0;
  for (i = 0, j = nvert-1; i < nvert; j = i++) {
    if ( ((verty[i]>testy) != (verty[j]>testy)) &&
         (testx < (vertx[j]-vertx[i]) * (testy-verty[i])
         / (verty[j]-verty[i]) + vertx[i]) )

         c = !c;
  }
  return c;
}
```

where *nvert* is the number of vertices in the polygon, *vertx* and *verty* the arrays containing the x- and y-coordinates of the polygon vertices and, finally, *testx* and *testy* are the x- and y-coordinate of the test point (the algorithm is performed for each point of the ellipse). The amount of points in common is converted into the fuzzy domain and is classified as *out*, *very low*, *low*, *medium*, *high*, or *very high*.

The classification process proceeds in each frame and it begins once the system knows the low level data of every object in the scene. The system makes use of the three sets of fuzzy rules to determine the object's class and speed, and if its behaviour is normal or, on the contrary, anomalous.

ISBN: 978-989-95079-6-8

In order to evaluate the learning method proposed in this paper, we have run ten tests for every step of the general process. Each training set had a different number of examples, and the 80% of them were used for learning and the rest were used in the classification process. Table 1 resumes the results obtained in the tests, where the average of the generated rules, examples correctly classified, classification errors, and percentage of successful are shown for every step.

Table 1: Results obtained in the tests

Step	Gen. Rules	Correct	Wrong	% Correct
step 1	29	43	2	96.28
step 2	22	79	3	96.34
step 3	50	120	5	96

5 Conclusions

Surveillance systems are being used in a wide range of environments which require more and more sophisticated solutions. In fact, a notable investment has been made during the last years in order to provide surveillance services, both in public and private environments, which increase the efficiency of such systems and manage high-level information to allow users to adequately make decisions and manage crisis situations. In this paper, we have presented a possible solution to classify and describe simple events related to spatial properties in a monitored environment having into account the camera perspective. Moreover, we have also determined whether these events are normal or not.

We have proposed a machine learning algorithm to acquire the necessary knowledge from examples of situations and a method that uses this algorithm. The proposed method includes three calls to the algorithm as has been explained in Section 2.

Our work starts from the fuzzy machine learning algorithm proposed by Castro et al. [17]. A key innovation has been added to this algorithm with the aim of adapting it to the new needs of our problem. This was primarily motivated due to that several generated rules with different consequences could be applied to the same premises, generally scenarios that are not present in the training data. This is due to the generalisation of each particular rule into a definitive rule. The modified algorithm prunes the over-generalisation of a definitive rule during the amplification process. As has been described in Section 3, the over-generalisation is not desirable for us and it will be controlled by means of a separability measurement. This measurement allows one definitive rule to capture the premises that are not present in the training data but close to those evidences in the data that justify that rule. Thanks to this separability measurement the results obtained are improved.

Finally, we want to remark that the separability threshold is empirically calculated by observing how the algorithm selects the different values of percentage of the maximum separability between the fuzzy sets of a given linguistic variable. For this reason, one of our lines of future research consists in designing a new algorithm to decide the best separability value to optimise the results.

Acknowledgment

This work has been founded by the Regional Goverment of Castilla-La Mancha under Research Projects PII2I09-0052-3440 and PII1C09-0137-6488.

References

[1] W. Wang and S. Maybank. A survey on visual surveillance of object motion and behaviors. *Systems, Man and Cybernetics, Part C, IEEE Transactions on*, 34(3):334–352, 2004.

[2] M. Valera and S.A. Velastin. Intelligent distributed surveillance systems: a review. In *Vision, Image and Signal Processing, IEE Proceedings-*, volume 152, pages 192–204, 2005.

[3] G.J.D. Smith. Behind the screens: Examining constructions of deviance and informal practices among cctv control room operators in the uk. *Surveillance and Society*, 2(2/3):376–95, 2004.

[4] J.F. Allen. An interval-based representation of temporal knowledge. In *Proc. 7th International Joint Conference on Artificial Intelligence, Vancouver, Canada*, pages 221–226, 1981.

[5] J. Allen and G. Ferguson. Actions and events in interval temporal logic. In *Journal of Logic and Computation*, volume 4(5), pages 531–579, 1994.

[6] F. Fusier, V. Valentin, F. Bremond, M. Thonnat, M. Borg, D. Thirde, and J. Ferryman. Video understanding for complex activity recognition. *Machine Vision and Applications*, 18(3):167–188, 2007.

[7] H.M. Dee and D.C. Hogg. Navigational strategies and surveillance. In *Proceedings of the IEEE International Workshop on Visual Surveillance*, pages 73–81, 2005.

[8] D. Makris and T. Ellis. Path detection in video surveillance. *Image and Vision Computing*, 20(12):895–903, 2002.

[9] G.L. Foresti, C. Micheloni, and L. Snidaro. Event classification for automatic visual-based surveillance of parking lots. *Proc. of the 17th International Confrence on Pattern Recognition*, 3:314–317, 2004.

[10] H. Buxton and S. Gong. Visual surveillance in a dynamic and uncertain world. *Artificial Intelligence*, 78(1-2):431–459, 1995.

[11] P. Remagnino and GA Jones. Classifying Surveillance Events from Attributes and Behaviour. In *the Proceeding of the British Machine Vision Conference*, pages 10–13.

[12] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[13] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, Jan-Feb 1985.

[14] P. Angelov and X. Zhou. Evolving Fuzzy-Rule-Based Classifiers From Data Streams. *IEEE Transactions on Fuzzy Systems*, 16(6):1462–1475, 2008.

[15] H. Stern, U. Kartoun, and A. Shmilovici. A prototype Fuzzy System for Surveillance Picture Understanding. *IASTED International Conference Visualization, Imaging, and Image Processing (VIIP 2001), Marbella, Spain*, pages 624–629, 2001.

[16] L. Rodriguez-Benitez, J. Moreno-Garcia, J.J. Castro-Schez, J. Albusac, and L. Jimenez-Linares. Automatic objects behaviour recognition from compressed video domain. *Image and Vision Computing*, 27(6):648 – 657, 2009.

[17] J.L. Castro, J.J. Castro-Schez, and J.M. Zurita. Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems. *Fuzzy Sets and Systems*, 101(3):331–342, 1999.

[18] M. Shimrat. Algorithm 112: Position of point relative to polygon. *Communications of the ACM*, 5(8):434, 1962.

ISBN: 978-989-95079-6-8

A Hybrid Method for Soccer Video Events Retrieval Using Fuzzy Systems

Ehsan Lotfi[1], M.-R. Akbarzadeh-T[2], H.R. Pourreza[3], Mona Yaghoubi[1] and F. Dehghan[1]

[1]Department of Computer Engineering, Islamic Azad University,
Mashhad Branch, Iran.
[2]Departments of Electrical Engineering and Computer
Engineering, Ferdowsi University of Mashhad, Iran and IEEE Senior Member.
[3]Departments of Computer Engineering, Ferdowsi University of Mashhad.
esilotf@gmail.com, Akbarzadeh@ieee.org, hpourreza@um.ac.ir, mona.yaghoobi@gmail.com

Abstract— *Most information retrieval systems make indirect use of human knowledge in their retrieval process. The new method we present here aims to efficiently use human knowledge directly in combination with support vector machines for clustering. As illustrated in this paper, this approach is particularly applicable to concept retrieval from soccer-related videos. The first phase consists of extracting suitable features from video shots. Then, using a fuzzy rule base containing the experiences of experts, shots that do not include significant events are removed. Finally, the last phase uses SVM to classify results coming from the fuzzy system. The results of the classification phase are accompanied by a textual description and enables retrieval through text based query. Experimental result show good classification and satisfying retrieval process.*

Keywords— Event Retrieval, Semantic Indexing, Hybrid Method, Fuzzy System, SVM, Soccer Video.

1 Introduction

Videos are used for different purposes such as education, mass media, entertainment, and surveillance systems. In each of these applications a video conveys a significant message to the audience. For example, in a football match video events such as penalties, corners, goals, and the very concept of victory or loss are conveyed when fans watch the video. The meanings contained in a video are quite obvious for men, but the same does not go for computers. A computer cannot understand the scoring of a goal in a football match video. This fact is usually referred to as the semantic gap between man and computers. Imagine if computers were capable of detecting and understanding the meaning of videos then we could use them to summarize sport videos automatically and retrieving meanings from a multimedia database. On the other hand the dramatic increase of multimedia data volumes in recent years will eventually force us to use some sort of retrieval system so that the indexing and retrieval of information is handled automatically. As noted earlier there is a semantic gap between the user and retrieval systems. The user queries the system with his/her ideas and the system presents it findings for such ideas as an answer. Such queries are done in different manners: 1. Query by keywords[11] 2. Sketch-based queries 3. Example-based queries[6] 4. Semantic-based queries[1]. Methods 2 and 3 are usually used in Content-Based Information Retrieval (CBIR). Recent studies have focused on Context instead of Content resulting in the presentation of a range of Semantic-Based Video Retrieval methods which are discussed in the published works section. The system presented here is for assessing significant events included in a database containing shots of football videos. Queries with the user are based on text. For instance, the user can query the system with items such as this: shots of goals scored by the yellow team. The user interface is designed in such a way that the user can ask for the summary of any given video. Published works include several major methods and techniques for retrieval of information-which are common in some features- as summarized here: 1. Context Based Information Retrieval (CBIR); In such methods the media is modelled by features such as colour, texture, etc. and special relationships between objects and movements. In CBIR suitable features are extracted and related to high level concepts and meanings, without the system knowing such meanings. In fact the system displays and models the video contents in such a way that is efficient for content retrieval. For example there is a statistical model of motion in [3]. After that there is a training session on dynamic contents and then the data are recognized and classified. The result is a database of different video shots (football, basketball, meeting, and highways) each divided into minor events. The provided structure classifies shots in various groups based on similarity of their contents. Also there is fuzzy presentation of video contents in [12]. In this paper suitable features of a multidimensional fuzzy histogram are made for each video frame. One feature will be suitable for studying the similarities between frames. Consecutive frames are classified based on the fuzzy histogram. In fact the paper divides the videos into shots by this method, and then it chooses a keyframe for each of these frames. Then the shot related to the keyframe is provided as the answer.

2. Semantic Based Retrieval; One of the most essential phases of designing a Semantic-based retrieval system is defining a meaningful list of concept-oriented meanings based on human's available knowledge. Thus each single

ISBN: 978-989-95079-6-8

concept in the list should be accompanied by correct and true descriptions on the video collection. A new perspective introduced in recent years includes the use of Semantic Description of Object Motion for the retrieval process, which has been studied in [1],[2],[8],[7], and [6]. The method offered in [1] is as follows: First objects included in video frames are extracted and their trajectories are tracked. Once the trajectories are extracted they are classified for the training of object motion models. Finally, a meaningful description is added to these motion models. It should be added that such a meaningful description was added manually.

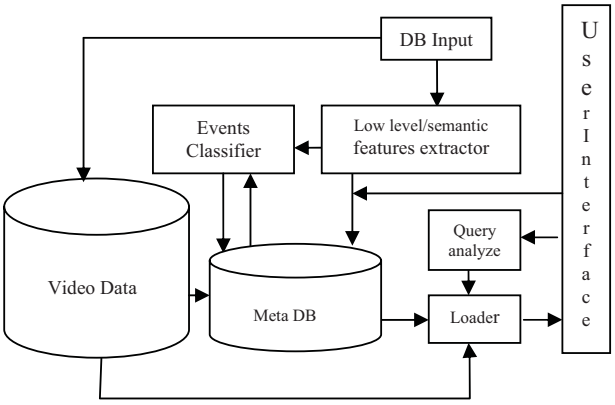


Fig. 1 The overall proposed structure

2 Proposed Structure

Most information retrieval systems make indirect use of human knowledge in their retrieval process(for example in the query by example, all of instance are a priori knowledge and in Semantic Based retrieval a meaningful description as priori knowledge should be added to the visual models manually.) The new method we are presenting in here uses human knowledge directly and in a very efficient way by a fuzzy rule base. The presented structure allows the system to process based on soccer video shots available in the database. The first phase is devoted to extracting shots from each video and making a list of features extracted from each shot. Then a fuzzy system is used to eliminate shots including insignificant events. Finally shots are classified and associated with predefined classed using a SVM. Then shots related to the class associated with the user query are provided as an answer to that query. The user may make queries on different events and concepts such as occurrence of penalties, corners or goals or team attacks throughout the DB. The overall structure is provided in Figure 1. As can been seen each video is processed before entering the DB and its low level features and conceptual properties are extracted, and once the events are classified all collected information is stored along with all data provided by the user (such as names of teams, time and place of the match, etc.) inside the Meta DB, while the video itself is stored in another location. Indeed the assumed database is of Multimedia linked meta database type. Figure 2 demonstrates feature extraction, Shot filtering using fuzzy system and event classification stages.

2.1 Extracting Low-Level and Semantic-Based Features

Used low-level features includes: Gi percent of pixels including grass of the field in keyframe of shot i for which we used the method provided in [5]. Once the grass region was recognized we divided it into 3 sections according to the Fig.3 and calculated Gi(E2) , percent of pixels covering the section E2 of the grass field, for keyframe. These features(Gi,Gi(E2)) may be used for Shot boundary detection and Shot classification and may be used as input to the fuzzy system not only in the Shot Rejection but also in the next phase of Event Classification along with low-level features.

The first step of proposed structure is devoted to extracting shots from each video (Fig. 2) and making a list of features extracted from each shot. Then a fuzzy system is used to eliminate shots including insignificant events. Finally shots are classified and associated with predefined classed using SVM The results of the classification phase are accompanied by a textual description and enables retrieval through text based query. The proposed method is based on the idea of using a fuzzy system for shot rejection and textual description for Soccer Events retrieval. In this section first the general structure is studied and we proceed to describe the extraction of low level features. Afterwards we go to high level features and then development of visual models. The fuzzy system including fuzzy rule base is presented in the section 2.2.

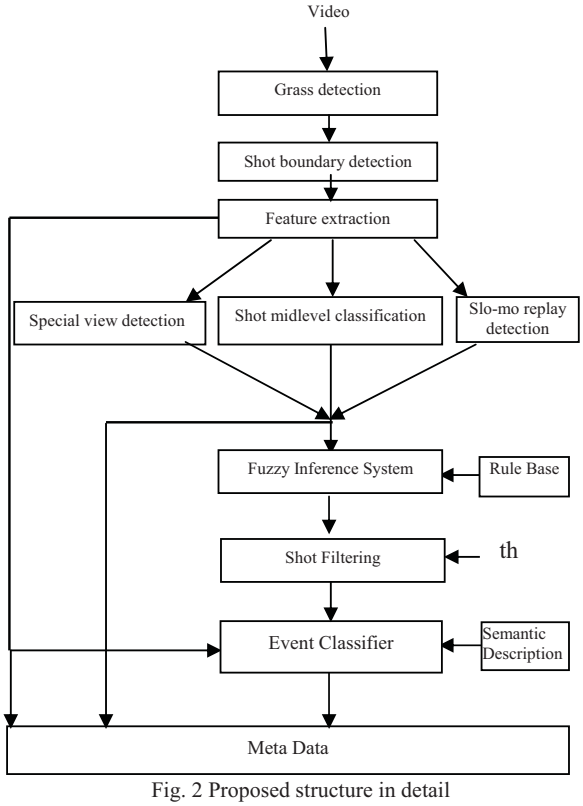


Fig. 2 Proposed structure in detail

ISBN: 978-989-95079-6-8



Fig. 3 Segmenting the area including the grass field

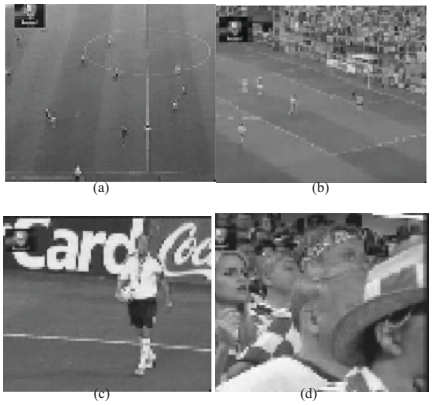


Fig.4 Types of views: a) Far-center, b) Far-side, c) Medium view, d) Out of field

Used semantic based features includes: 1. Far views from midfield, Fig. 4.a. 2. Far views from field sides, Fig. 4.b. 3. Medium views from inside the field 4.c, 4. View of the area outside of the field / closed view, showing the spectators or the upper body of a player, Fig. 4.d 5. Slow-Motion Replay Detection[9]. we have proposed a structural classification method for extracting this views in [5].

2.2 Using a fuzzy system to eliminate shots containing no significant events

Data retrievable from each shot by the above features as input of fuzzy system up to this phase are included in Table 1 along with abbreviations used for each one of them. The goal of designing a fuzzy system is to use the collected data in defining a degree of significance for each shot. This fuzzy system shows how important are the contents of that shot in terms of events occurring in it. For example, the system is expected to give a higher degree of significance to a shot containing a goal scoring than the one containing images of fans and no other important event. The rulebase and 3 triangle membership functions (low, Medium, High) are presented in Figure 5,6. The inference method used here is Mamdani product, for example for one execution of fuzzy inference result see Fig 7. Once each shot is given a degree of significance by the fuzzy system, a thresholding phase is run to eliminate the fuzzy system outputs for some shots. In this phase all shots with a significance degree below th(our system parameter) are eliminated from the classification. by defining a threshold(th) on the output of the fuzzy system we can separate the useful and none-useful shots for the purpose of classification. The threshold must be between 0 and 1. The lowest amount of the threshold is zero. By defining the threshold as 0 the all of shots in the database will be seen in the output of rejection phase as well. the higher the degree of significance the more sensitive the system vice versa.

Table 1 The input/output of fuzzy system with their abbreviations

Shot Degree	Out	Sd
Percent of Far-center view in a shot	In	Fc
Percent of Far-side view in a shot	In	Fs
Percent of Medium view	In	Mv
Percent of Out of field	In	Of
Percent of Frame includes Slow-Motion	In	Sm

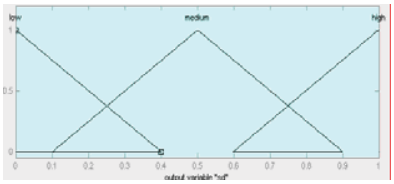


Fig. 5 triangle membership functions (low, Medium, High) for most In/Out variables

1. If (fc is high) then (sd is low) (1)
2. If (fs is high) then (sd is very high) (1)
3. If (fs is medium) then (sd is high) (1)
4. If (mv is high) then (sd is high) (1)
5. If (of is high) then (sd is low) (1)
6. If (cu is high) then (sd is low) (1)
7. If (cu is medium) then (sd is medium) (1)
8. If (fs is low) then (sd is medium) (1)
9. If (sm is high) then (sd is high) (1)
10. If (sm is medium) then (sd is high) (1)

Fig. 6 Fuzzy rule base for none-useful shot rejection

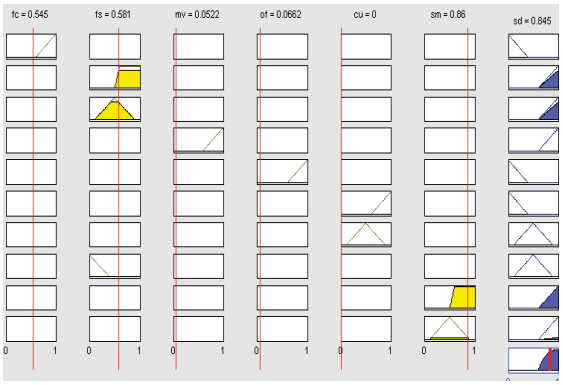


Fig. 7 if(fc=0.545 & fs=0.581 & mv=0.052 & of=0.0682 & cu=0 & sm=0.86) then Shot Degree = 0.845

ISBN: 978-989-95079-6-8

3 Event Retrieval

Up to the previous level we rejected most video shots imported into the DB with high probability of being irrelevant to significant events. So classification of shots in this phase may be done based on the events included in them. We have listed the set of features that may be used to classify the events , features include: G, percent of pixels covering the grass field in shot keyframes, G2, percent of pixels in section E2 on the grass field in shot keyframes and the Sd (the output of fuzzy system) and Fs, Percent of Far-side view in a shot. After normalizing the above 4 features the value of each of them for each shot will range between 0 and 1, so the feature vector for each shot is shown by Fv (I,V) and expressed in the following form:

Fv(i,V)={G,G2,Fs,Sd}
Where: i is the shot number from video V.

We used SVM to classify the events. In this phase events must be associated with 5 predefined classes. All classes have their special keywords which are presented here:

EventClass = { Goal, Penalty, Kroner, Free kick, other}

After the learning phase the videos stored in the DB for making up the Meta Data are processed and then the events are classified. Afterwards the above keywords are associated with the created classes manually and then each new event automatically inherits its keyword and is recorded in the Meta Data for that event. A table is formed in the Meta Data for each single video processed (Figure 8). During implementation the name given to each table of the Meta Data is taken to be the same as the file name recorded in the database, so relations with the raw video are maintained. Also we can save the information entered into the system for each video in Title of main video file to be used in the retrieval process. Such information may include the name of competing teams, time or place of the match or other details.

Shot Index	Start of shot	End of Shot	Type of Shot	Slow-Motion	Class of Event
. . .	. .	...	. .	... ...	

Fig 8 A table is formed in the Meta Data for each single video processed (The field related to Type of Shot shows the view that has been most frequent in the shot.)

The query model we used for this project was the same proposed in [2]. The base expression used in the model included the following:

select video from search_space [where condition];
select shot from search_space [where condition];

Expression SELECT specifies the SEARCH TARGET which may be Videos or Shots. The search space comes after FROM which may be all videos available in the DB or a list of chosen videos from any given video specified with a "where" condition which may only constitute one restriction or a combination of and/or phrases used to create conditions for time dependencies, which we ignore here.The provided system will be able to answer questions such as these:

Q1: "Find all the goal shots from all the soccer videos"
Q2: "Find all the corner kick shots from all the soccer videos where the goal accurse by England team"
Remember that the prerequisite to doing the condition contained in the second question is that the required information are added to the Meta Data by the system manager while the video is being entered into the database. To find the answer it suffices to extract the defined terms from the query expression and search for them in the Meta Data.

4 Evaluation

We have used more than 5.5 hours of football Videos in the database includes: 1 match of the World Cup 2006, 2 matches from the UEFA Champions League 2005, 1 match from the FA Premier League 2004, and more than 5 clips from Euro 2004(Table 2). The file format of the input films was non-compressed avi and the size of output films is 88×79.

Table 2. Names and the length (min:sec) of the clips in the database

Euro001(46:40),Euro002(47:00),Euro003(47:30), Euro004(46:00), Euro005(46:45), FIFA-div01f(12:38), FIFA-div02f(18:23),FIFA-div03f(15:38),FIFA-div04f(17:40),FIFA-div05f(18:22),Eng1(42:7) ,Eng2(20:51)

In the training phase we used 119 shots to learning for SVM and the rate of correct class in training data is 87%. As noted earlier the output of fuzzy system is a degree for each shot and the system rejects no useful shots using a threshold(th) as system input parameter. The input margin of the system for evaluation is 0.2 and then inputs include 0.2, 0.4, 0.6, 0.8 and 1. The results of changing the input amount of the system and the accuracy of the events classification are shown in tables 3,4,5,6. In this evaluation scenario we used 187 shots includes 5 Goals, 5 Penalties, 6 Free kicks, 9 Corners and 162 shots containing no significant events. Table 3 where the input th is 0.2 then system rejected 23

ISBN: 978-989-95079-6-8

shots before classification, table 3 shows the result of classification after rejection phase with input 0.2, Table 4 where the input th is 0.4 then system rejected 86 shots. Table 5 where th is 0.6 and the system rejected 114 shots and Table 6 where th is 0.8 and the system rejected 172 shots.
When th is 0.6, 4 shots containing corner and free kicks are rejected incorrectly(see the '# of rejected' row of table 5) and When th is 0.8, 10 shots containing very important events are rejected incorrectly (see the '# of rejected row' of table 6).
In the tables accuracy formula is:
Accuracy =100* Correct/(Total + False)
The Tables indicate that the higher the threshold, the higher the number of rejected shots will be. Also, the higher the threshold, the lower the number of incorrect shot detection (see 'False' row of tables) will be. See Table 6 where th = 0.8 and False detection for all events is zero but 1 Goal, 5 Free kicks and 4 corner are rejected incorrectly.

Table 3. The result of proposed method with th = 0.2

	Goal	Penalties	Free kick	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	0	0	23	23
Correct	5	5	6	9	86	111
False	13	9	17	14	0	53
Accuracy	27%	36%	26%	39%	-	-

Table 4. The result of proposed method with th = 0.4

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	0	0	86	86
Correct	5	5	5	8	51	74
False	6	0	13	6	1	27
Accuracy	45%	100%	32%	53%	-	-

Table 5. The result of proposed method with th = 0.6

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	2	2	110	114
Correct	5	5	4	7	42	63
False	2	0	5	3	0	10
Accuracy	71%	100%	36%	58%	-	-

Table 6. The result of proposed method with th = 0.8

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	1	0	5	4	162	172
Correct	4	5	1	5	0	15
False	0	0	0	0	0	0
Accuracy	80%	100%	16%	55%	-	-

5 Conclusions

The proposed method uses human expert knowledge for soccer video events retrieval by fuzzy systems. After feature extraction, each shot is given a degree of significance by the fuzzy system, thereby significantly reducing time complexity of video processing. By defining a threshold on the output of the fuzzy system, the useful and none-useful shots are separated for the purpose of event classification. Support vector machines are then applied for final classification of video information.

References

[1] Weiming Hu, Dan Xie, Zhouyu Fu, Wenrong Zeng, and Steve Maybank, "Semantic-Based Surveillance Video Retrieval", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 4, APRIL 2007.

[2] Cees G. M. Snoek, Bouke Huurnink, Laura Hollink,"Adding Semantics to Detectors for Video Retrieval",IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 9, NO. 5, AUGUST 2007.

[3] Gwenaëlle Piriou, Patrick Bouthemy, and Jian-Feng Yao,"Recognition of Dynamic Video Contents With Global Probabilistic Models of Visual Motion' , IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 11, NOVEMBER 2006.

[4] Cees G.M. Snoek , "Multimedia Event based Video Indexing using Time Intervals," IEEE TRANSACTIONS ON MULTIMEDIA, 2004.

[5] Ehsan Lotfi, H.R Pourreza, "Event Detection and Automatic Summarization in Soccer Video ," 4th Iranian Conference on Image Processing and Machine Vision, Mashhad, Iran, Feb 2007.

[6] Z. Q. Liu, L. T. Bruton, J. C. Bezdek, J. M. Keller, S. Dance, N. R. Bartley, and C. Zhang, "Dynamic image sequence analysis using fuzzy measures," IEEE Trans. Syst., Man, Cybern., vol. 31, no. 4, pp. 557–571, Aug. 2001.

[7] Haiping Sun, Joo-Hwee Lim, Qi Tian, Mohan S. Kankanhalli, "SEMANTIC LABELING OF SOCCER VIDEO", Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on MultimediaVolume 3, Issue , 15-18 Dec. 2003.

[8] B. Bell and L. F. Pau, "Context knowledge and search control issues in object-oriented prolog-based image understanding," Pattern Recognit. Lett., vol. 13, no. 4, pp. 279–290, 1992

[9] Ahmet Eklin,A.murat Tekalp, and Rajiv Mehrotra,"Automatic Soccer Video Analysis and Summarization", IEEE Transactions on Image processing Volume 12, Issue 7, July 2003.

[10] N.-S. Chang and K.-S. Fu, "Query-by-Pictorial-Example," IEEE Transactions on Software Engineering, vol. 6, no. 6, pp. 5 19-524, November 1980.

[11] Yu-Lin Kang ,"SOCCER VIDEO EVENT DETECTION WITH VISUAL KEYWORDS", ICICS-CM 200315-18 December 2003 Singapore.

[12] Anastasios D. Doulamis*, Nikolaos D. Doulamis, Stefanos D. Kollias "A fuzzy video content representation for video summarization and content-based retrieval" , ELAIVIER, Signal Processing 80 (2000) 1049.

[13] " Inference and retrieval of soccer event", SUN Xing-hua, YANG Jing-yu, Journal of Communication and Computer, Volume 4, No.3 (Serial No.28) , Mar. 2007.

ISBN: 978-989-95079-6-8

Pointwise defined CRI-based aggregation distributive operators are trivial

Siegfried Gottwald[1] Philipp Hüsam[2]

1. Universität Leipzig, Abt. Logik und Wissenschaftstheorie, Institut für Philosophie
Beethovenstr. 15, 04107 Leipzig, Germany
2. Universität Leipzig, Institut für Mathematik
Augustusplatz 9, 04109 Leipzig, Germany
Email: gottwald@uni-leipzig.de, philipp@huesam.de

***Abstract*— *In a rather general, essentially aggregation operator based discussion of the traditional fuzzy control strategies known as FATI and FITA strategies, a way to reduce these strategies to one another has been to define pairs of aggregation distributive aggregation operators. In this paper it is shown that for some often used special cases this reduction condition allows only the set theoretic union as aggregation operator.*

***Keywords*—** fuzzy control, FATI strategy, FITA strategy, aggregation operators, aggregation distributivity

1 Introduction

The standard paradigm behind standard approaches toward fuzzy control is that one supposes to have given, as an incomplete and fuzzy description of a control function Φ from an input space $\mathbb{X}$ to an output space $\mathbb{Y}$, a family

$$\mathcal{D} = (\langle A_i, B_i\rangle)_{1\leq i\leq n} \tag{1}$$

of (fuzzy) input-output data pairs to characterize this function Φ.

In the usual approaches such a family of input-output data pairs is provided by a finite list

$$\text{IF } x \text{ is } A_i \text{ THEN } y \text{ is } B_i, \tag{2}$$

(with $i = 1, \ldots, n$) of linguistic control rules, also called fuzzy IF-THEN rules.

The basic examples of fuzzy control approaches are Zadeh's original approach via fuzzy relations and the compositional rule of inference (CRI), as prototypically realized by the Mamdani–Assilian approach in [7], and the Holmblad–Ostergaard approach toward fuzzy control of a cement kiln as explained in [5].

Derived from these two approaches there is the well known distinction between FATI and FITA strategies to evaluate systems of linguistic control rules w.r.t. arbitrary fuzzy inputs from $\mathcal{F}(\mathbb{X})$.

2 Preliminaries

The core idea of a FITA strategy is that it is a strategy which **F**irst **I**nfers (by reference to the single rules) and **T**hen **A**ggregates starting from the actual input information A. Contrary to that, a FATI strategy is a strategy which **F**irst **A**ggregates (the information in all the rules into one fuzzy relation) and **T**hen **I**nfers starting from the actual input information A.

Both these strategies use the set theoretic union as their aggregation operator. Furthermore, both of them refer to the CRI as their core tool of inference.

In general, however, the interpolation operators may depend more generally upon some inference operator(s) as well as upon some aggregation operator.

2.1 Aggregation operations and fuzzy control strategies

By an *inference operator* we mean here simply a mapping from the class of fuzzy subsets of the input space to the class of fuzzy subsets of the output space.[1]

And an *aggregation operator* $\mathbf{A}$, as explained e.g. in [1, 2], is a family $(f^n)_{n\in\mathbb{N}}$ of ("aggregation") operations, each f^n an n-ary one, over some partially ordered set $\mathbf{M}$, with ordering $\leqslant$, with a bottom element $\mathbf{0}$ and a top element $\mathbf{1}$, such that each operation f^n is non-decreasing, maps the bottom to the bottom: $f^n(\mathbf{0}, \ldots, \mathbf{0}) = \mathbf{0}$, and the top to the top: $f^n(\mathbf{1}, \ldots, \mathbf{1}) = \mathbf{1}$.

Such an aggregation operator $\mathbf{A} = (f^n)_{n\in\mathbb{N}}$ is a *commutative* one iff each operation f^n is commutative. And $\mathbf{A}$ is an *associative* aggregation operator iff

$$\begin{aligned} &f^n(a_1, \ldots, a_n) = \\ &\quad f^r(f^{k_1}(a_1, \ldots, a_{k_1}), \ldots, f^{k_r}(a_{m+1}, \ldots, a_n)) \end{aligned} \tag{3}$$

for $n = \sum_{i=1}^{r} k_i$ and $m = \sum_{i=1}^{r-1} k_i$.

Our aggregation operators further on are supposed to be commutative as well as associative ones.[2]

As in [3, 4], we now consider operators Ψ of FATI-type operators Ξ of FITA-type and which have the abstract forms

$$\Psi_{\mathcal{D}}(A) = \mathbf{A}(\theta_{\langle A_1,B_1\rangle}(A), \ldots, \theta_{\langle A_n,B_n\rangle}(A)), \tag{4}$$

$$\Xi_{\mathcal{D}}(A) = \widehat{\mathbf{A}}(\theta_{\langle A_1,B_1\rangle}, \ldots, \theta_{\langle A_n,B_n\rangle})(A). \tag{5}$$

Here we assume that each one of the "local" inference operators θ_i is determined by the single input-output pair $\langle A_i, B_i\rangle$. This restriction is in general sufficient. For the present purpose we assume that our inference operators are CRI-based, i.e. we assume that $\theta_{\langle A_1,B_1\rangle}(A)$ has

[1] This terminology has its historical roots in the fuzzy control community. There is no relationship at all with the logical notion of inference intended and supposed here; but–of course–also not ruled out.

[2] It seems that this is a rather restrictive choice from a theoretical point of view. However, in all the usual cases these restrictions are satisfied.

ISBN: 978-989-95079-6-8

the form

$$\theta_{\langle A_i,B_i\rangle}(A) = R_i'' A \tag{6}$$

for some fuzzy relation R. In this case we call the inference operator $\theta_{\langle A_1,B_1\rangle}$ CRI-*based*.

Furthermore $\mathbf{A}$ has to be an aggregation operator for fuzzy subsets of the universe of discourse $\mathbb{X}$, and $\widehat{\mathbf{A}}$ has to be an aggregation operator for inference operators.

2.2 Stability conditions

If $\Theta_{\mathcal{D}}$ is a fuzzy inference operator of one of the types (4), (5), then the interpolation property one likes to have realized is that one has

$$\Theta_{\mathcal{D}}(A_i) = B_i \tag{7}$$

for all the data pairs $\langle A_i, B_i\rangle$. In the particular case that the operator $\Theta_{\mathcal{D}}$ is detrmined by the CRI-methodology, this is just the usual problem to solve a system (7) of fuzzy relation equations.

In the present generalized context the property (7) has been called the $\mathcal{D}$-*stability* of the fuzzy inference operator $\Theta_{\mathcal{D}}$.

To find $\mathcal{D}$-stability conditions on this abstract level seems to be rather difficult in general. However, the restriction to fuzzy inference operators of FITA-type makes things easier.

To explain some of the known results it is necessary to have a closer look at the aggregation operator $\mathbf{A} = (f^n)_{n\in\mathbb{N}}$ involved in (4) which operates on $\mathbb{F}(\mathbb{Y})$, of course with the inclusion relation for fuzzy sets as partial ordering.

Definition 1 *Having $B, C \in \mathbb{F}(\mathbb{Y})$ we say that C is* $\mathbf{A}$-negligible w.r.t. *B iff $f^2(B, C) = f^1(B)$ holds true.*

The core idea here is that in any aggregation by $\mathbf{A}$ the presence of the fuzzy set B among the aggregated fuzzy sets makes any presence of C superfluous.

For examples and further interesting properties of aggregation operators the interested reader may consult [3, 4].

Now we are in a position to state one of the results from [3, 4] to give an impression of what becomes of interest in the present context.

Proposition 1 *Consider a fuzzy inference operator $\Psi_{\mathcal{D}}$ of FITA-type (4). It is sufficient for the $\mathcal{D}$-stability of $\Psi_{\mathcal{D}}$, i.e. to have*

$$\Psi_{\mathcal{D}}(A_k) = B_k \qquad \text{for all } k = 1, \ldots, n \tag{8}$$

that one always has

$$\theta_{\langle A_k,B_k\rangle}(A_k) = B_k \tag{9}$$

and additionally that for each $i \neq k$ the fuzzy set

$$\theta_{\langle A_k,B_k\rangle}(A_i) \quad \text{is } \mathbf{A}\text{-negligible w.r.t.} \quad \theta_{\langle A_k,B_k\rangle}(A_k)\,. \tag{10}$$

This result has two quite interesting specializations which themselves generalize well known results about fuzzy relation equations. The interested reader may consult [3, 4].

To extend such considerations from inference operators (4) of the FITA type to those ones of the FATI type (5) let us consider the following notion.

Definition 2 *Suppose that $\widehat{\mathbf{A}}$ is an aggregation operator for inference operators, and that $\mathbf{A}$ is an aggregation operator for fuzzy sets. Then $(\widehat{\mathbf{A}}, \mathbf{A})$ is an* application distributive *pair of aggregation operators iff*

$$\widehat{\mathbf{A}}(\theta_1, \ldots, \theta_n)(X) = \mathbf{A}(\theta_1(X), \ldots, \theta_n(X)) \tag{11}$$

holds true for arbitrary inference operators $\theta_1, \ldots, \theta_n$ and fuzzy sets X.

Using this notion it is easy to see that one has on the left hand side of (11) a FATI type inference operator, and on the right hand side an associated FITA type inference operator. So one is able to give a reduction of the FATI case to the FITA case, assuming that such application distributive pairs of aggregation operators exist.

Proposition 2 *Suppose that $(\widehat{\mathbf{A}}, \mathbf{A})$ is an application distributive pair of aggregation operators. Then a fuzzy inference operator $\Xi_{\mathcal{D}}$ of FATI-type is $\mathcal{D}$-stable iff its associated fuzzy inference operator $\Psi_{\mathcal{D}}$ of FITA-type is $\mathcal{D}$-stable.*

3 Application distributivity

Based upon the notion of application distributive pair of aggregation operators the property of $\mathcal{D}$-stability can be transferred back and forth between two inference operators of FATI-type and of FITA-type if they are based upon a pair of application distributive aggregation operators.

What has not been discussed previously was the existence and the uniqueness of such pairs. Here are some results concerning these problems.

The uniqueness problem has a simple solution.

Proposition 3 *If $(\widehat{\mathbf{A}}, \mathbf{A})$ is an application distributive pair of aggregation operators then $\widehat{\mathbf{A}}$ is uniquely determined by $\mathbf{A}$, and conversely also $\mathbf{A}$ is uniquely determined by $\widehat{\mathbf{A}}$.*

And for the existence problem we have a nice reduction to the two-argument case.

Theorem 1 *Suppose that $\mathbf{A}$ is a commutative and associative aggregation operator. For the case that there exists an aggregation operator $\widehat{\mathbf{A}}$ such that $(\widehat{\mathbf{A}}, \mathbf{A})$ form an application distributive pair of aggregation operators it is necessary and sufficient that there exists some operation G for fuzzy inference operators satisfying*

$$\mathbf{A}(\theta_1(X), \theta_2(X)) = G(\theta_1, \theta_2)(X) \tag{12}$$

for all fuzzy inference operators θ_1, θ_2 and all fuzzy sets X.

However, there is an important restriction concerning the existence of such pairs of application distributive aggregation operators, at least for the interesting particular case that the application operation is determined by the compositional rule of inference (CRI). And this means simply that the inference operations θ_i are determined via suitable fuzzy relations R_i.

ISBN: 978-989-95079-6-8

Definition 3 *An aggregation operator* $\mathbf{A} = (f^n)_{n\in\mathbb{N}}$ *for fuzzy subsets of a universe of discourse* $\mathbb{X}$ *is* pointwise defined *iff for each* $n \in \mathbb{N}$ *there exists a function* $g_n : [0,1]^n \to [0,1]$ *such that for all* $A_1, \ldots, A_n \in \mathcal{F}(\mathbb{X})$ *and all* $x \in \mathbb{X}$ *there hold*

$$f^n(A_1, \ldots, A_n)(x) = g_n(A_1(x), \ldots, A_n(x)) \,. \tag{13}$$

And an aggregation operator $\widehat{\mathbf{A}}$ *for inference operators is* pointwise defined *iff it can be reduced to a pointwise defined aggregation operator for fuzzy relations.*

From the isotonicity behavior of the aggregation operator $\mathbf{A}$ it follows that also these characterizing functions g_n are isotonic, and similarly in the case of $\widehat{\mathbf{A}}$.

The restrictive result, first proved in [6], now reads as follows.

Theorem 2 *Suppose that all inference operators are CRI-based. Then the pair* $(\bigcup, \bigcup)$ *is the only application distributive pair among the commutative, associative, and pointwise defined aggregation operators.*

Proof: Obviously the considerations can restricted to the binary case $n = 2$. So let us start in this CRI-based case with an aggregation operator $\widehat{\mathbf{A}}$, which has to give a fuzzy relation $\widehat{\mathbf{A}}(R_1, R_2)$ for any two inference operators θ_1, θ_2 determined by the fuzzy relations R_1, R_2, respectively. Because $\widehat{\mathbf{A}}$ has to be pointwise defined, according to Definition 3 there has to be a function $\widehat{g} : [0,1] \times [0,1] \to [0,1]$ such that one has for the membership degrees of the corresponding fuzzy relations

$$\widehat{\mathbf{A}}(R_1, R_2)(x, y) = \widehat{g}(R_1(x, y), R_2(x, y)) \,. \tag{14}$$

In a similar way, again by Definition 3, a corresponding aggregation operator $\mathbf{A}$ has to be determined by a function $g : [0,1] \times [0,1] \to [0,1]$. Assuming that these aggregation operators $(\mathbf{A}, \widehat{\mathbf{A}})$ form an aggregation distributive pair, gives for arbitrary fuzzy inputs A the condition

$$\begin{aligned}
&\widehat{\mathbf{A}}(R_1, R_2)''(A)(y) \\
&= \bigvee_x T\Big(A(x), \widehat{\mathbf{A}}(R_1, R_2)(x, y)\Big) \\
&= \bigvee_x T\Big(A(x), \widehat{g}(R_1(x, y), R_2(x, y))\Big) \qquad (15) \\
&= g\Big(\bigvee_x T(A(x), R_1(x, y)), \bigvee_x T(A(x), R_2(x, y))\Big) \\
&= \mathbf{A}(R_1'' A, R_2'' A)(y) \,,
\end{aligned}$$

which has to be satisfied for arbitrary fuzzy sets A and fuzzy relations R_1, R_2. Of course, T here is the t-norm involved in the CRI application process.

So let be always $A(x) = 1$ and furthermore $R_1(x, y) = a, R_2(x, y) = b$ for some $a, b \in [0,1]$. Now routine calculations yield $\widehat{g}(a, b) = g(a, b)$, which means equality of the functions $\widehat{g} = g$ which determine the aggregation operators $\widehat{\mathbf{A}}, \mathbf{A}$, respectively.

So application distributivity of the pair $(\mathbf{A}, \widehat{\mathbf{A}})$ becomes a condition which has to be satisfied by the characterizing function g, and this condition reads

$$\begin{aligned}
&\bigvee_x T\Big(A(x), g(R_1(x, y), R_2(x, y))\Big) \; = \qquad (16) \\
&\qquad g\Big(\bigvee_x T(A(x), R_1(x, y)), \bigvee_x T(A(x), R_2(x, y))\Big) \,.
\end{aligned}$$

To continue our discussion and to finish the proof of Theorem 2, we insert two lemmata.

Lemma 1 *Suppose that* $g : [0,1] \times [0,1] \to [0,1]$ *determines a commutative and associative pointwise defined (binary) aggregation operator. Then the condition*

$$\sup_{i\in I} g(a_i, b_i) = g(\sup_{i\in I} a_i, \sup_{i\in I} b_i) \tag{17}$$

is equivalent to the fact that g *is left continuous and satisfies*

$$g(a, b) = g(\max\{a, b\}, \max\{a, b\}) \,. \tag{18}$$

It is easy to see that (17) implies the left continuity of g. So assume (17) and that (18) is not generally satisfied. Then there are $a_0, b_0 \in [0,1]$ such that

$$g(a_0, b_0) \neq g(\max\{a_0, b_0\}, \max\{a_0, b_0\}) \tag{19}$$

and additionally, w.l.o.g., also $b_0 \leq a_0$. This last condition forces even its strengthening $b_0 < a_0$, and together with the isotonicity of g yields $g(a_0, b_0) < g(a_0, a_0)$. But this now means

$$\begin{aligned}
&\max\{g(a_0, b_0), g(b_0, a_0)\} = g(a_0, b_0) \\
&\quad < \quad g(a_0, a_0) = g(\max\{a_0, b_0\}, \max\{a_0, b_0\}) \,, \qquad (20)
\end{aligned}$$

contradicting (17). So the $(\Rightarrow)$-part of the lemma is proved.

If otherwise g is left continuous and satisfies (18), one has

$$g(\sup_{i\in I} a_i, \sup_{i\in I} b_i) \geq \sup_{i\in I} g(a_i, b_i) \tag{21}$$

by the isotonicity of g. But $g(a_j, b_j) \leq \sup_{i\in I} g(a_i, b_i)$ and (18) yield

$$g(c_j, c_j) \leq \sup_{i\in I} g(a_i, b_i) \tag{22}$$

for $c_j = \max\{a_j, b_j\}$, and by the left continuity and isotonicity of g this gives

$$g(\sup_{i\in I} c_i, c_j) \leq \sup_{i\in I} g(c_i, c_i) \leq \sup_{i\in I} g(a_i, b_i), , \tag{23}$$

and thus also

$$g(\sup_{i\in I} a_i, \sup_{i\in I} b_i) \leq g(\sup_{i\in I} c_i, \sup_{i\in I} c_i) \leq \sup_{i\in I} g(a_i, b_i) \,. \tag{24}$$

All together gives (17) and hence the $(\Leftarrow)$-part of the lemma.

The condition (17) obviously means that the corresponding aggregation operator commutes with the supremum.

ISBN: 978-989-95079-6-8

Lemma 2 *Suppose that* $g : [0,1] \times [0,1] \to [0,1]$ *determines a commutative and associative pointwise defined (binary) aggregation operator. Then condition (17) is satisfied iff there exists a left continuous isotonic function* $h : [0,1] \to [0,1]$ *satisfying* $h(0) = 0$, $h(1) = 1$ *and*

$$g(a,b) = \max\{h(a), h(b)\}\,. \tag{25}$$

For the $(\Rightarrow)$-part consider the function $h(x) = g(0,x)$. It is left continuous as well as isotonic, satisfies $h(0) = 0$ and $h(1) = 1$, and one has by Lemma 1

$$\begin{aligned} g(a,b) &= g(\max\{a,b\}, \max\{a,b\}) \\ &= g(\max\{0, \max\{a,b\}\}, \max\{0, \max\{a,b\}\}) \\ &= g(0, \max\{a,b\}) \;=\; \max\{h(a), h(b)\}\,. \end{aligned} \tag{26}$$

The $(\Leftarrow)$-part follows from routine calculations.

So we come back to the proof of Theorem 2. What we did not discuss up to now is that for having an aggregation distributive pair of pointwise defined operators the characterizing function g has to be distributive relative to the t-norm T which determines the CRI-application. So we need to have always satisfied

$$T(a, g(b,c)) = g(T(a,b), T(a,c))\,, \tag{27}$$

which means, via Lemmata 1 and 2, to have always satisfied

$$\begin{aligned} &\max\{T(a,h(b)), T(a,h(c))\} \\ &\quad = \quad \max\{h(T(a,b)), h(T(a,c))\}\,, \end{aligned} \tag{28}$$

which forces that one always has to have

$$T(a,h(b)) = h(T(a,b))\,. \tag{29}$$

And this yields $h = \mathsf{id}$: because otherwise there would exist some c with $h(c) \neq c$ and

$$T(c,h(1)) = T(c,1) = c \neq h(c) = h(T(c,1))\,. \tag{30}$$

So Theorem 2 is finally proved.

4 Conclusion

The type of approach explained in Section 2.2 works actually well only in the FITA case. This was the starting point for the considerations on aggregation distributive operator pairs. They give a transfer possibility to the FATI case.

However, as the main result, i.e. Theorem 2 of the present paper shows, these transfer possibilities are quite restricted under some conditions which have, up to now, been considered rather natural ones.

For the authors understanding this result points into two directions. (i) It may be appropriate to try to find other ways and to discuss the FATI case differently. And this way may become essentially different from the reduction strategy toward fuzzy relation equations which stands behind the generalization in [3, 4]. (ii) It may be suitable to move into the realm of aggregation operators which are no longer pointwise defined, and it may also be suitable to leave the world of the CRI-based approaches.

Particularly point (ii) here seems promising, even having in mind that the actual standard cases all fall into the class of pointwise defined aggregation operators. Further investigations into this topic are necessary.

References

[1] Tomasa Calvo, G. Mayor, and Radko Mesiar, editors. *Aggregation Operators: New Trends and Applications*. Physica-Verlag, Heidelberg, 2002.

[2] Didier Dubois and Henri Prade. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142:143–161, 2004.

[3] Siegfried Gottwald. On a generalization of fuzzy relation equations. In *Proc. 11th Internat. Conf. IPMU 2006, Paris*, volume 2, pages 2572–2577, Paris, 2006. Edition EDK.

[4] Siegfried Gottwald. An abstract approach toward the evaluation of fuzzy rule systems. In Paul P. Wang, Da Ruan, and Etienne E. Kerre, editors, *Fuzzy Logic. A Spectrum of Theoretical & Practical Issues*, volume 215 of *Studies in Fuzziness and Soft Computing*, pages 281–297. Springer-Verlag, Berlin, 2007.

[5] L.P. Holmblad and J.J. Ostergaard. Control of a cement kiln by fuzzy logic. In Madan M. Gupta and Elie Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 389–399. North-Holland Publ. Comp., Amsterdam, 1982.

[6] Philipp Hüsam. FATI versus FITA. Ein abstrakter Vergleich gängiger Strategien bei "fuzzy control"-Systemen. Diplomarbeit, Mathematisches Institut, Universität Leipzig, 2008.

[7] Abe Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Internat. J. Man-Machine Studies*, 7:1–13, 1975.

ISBN: 978-989-95079-6-8

A possibilistic approach to bottleneck combinatorial optimization problems with Ill-known weights

Adam Kasperski[1] Paweł Zieliński[2]

1.Institute of Industrial Engineering and Management, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
2.Institute of Mathematics and Computer Science, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
Email: adam.kasperski@pwr.wroc.pl, pawel.zielinski@pwr.wroc.pl

Abstract— *In this paper a general bottleneck combinatorial optimization problem with uncertain element weights modeled by fuzzy intervals is considered. A rigorous possibilistic formalization of the problem and solution concepts in this setting that lead to finding robust solutions under fuzzy weights are given. Some algorithms for finding a solution according to the introduced concepts and evaluating optimality of solutions and elements are provided.*

Keywords— Bottleneck combinatorial optimization, Interval, Fuzzy interval, Fuzzy optimization and design, Possibility theory.

1 Introduction

A *combinatorial optimization problem* consists in finding an object composed of elements of a given ground set E. In a deterministic case, every element of E has a precise weight. In *bottleneck combinatorial optimization problems*, we wish to find an object that minimizes the weight of its heaviest element. Such formulation encompasses a large variety of classical combinatorial optimization problems, for instance: the bottleneck path [1], the bottleneck assignment [2], the bottleneck spanning tree [3] (or a more general the bottleneck matroid base problem [4]) etc. All these problems are efficiently solvable when the weights of all elements are precisely known. Unfortunately, in real world it is not easy to estimate element weights exactly. In many cases, the exact values of weights are not known in advance and the imprecision must be taken into account.

In this paper, we wish to investigate a general bottleneck combinatorial optimization problem with uncertain element weights modeled by fuzzy intervals. The membership function $\mu_{\widetilde{W}}$ of a fuzzy interval $\widetilde{W}$ is a possibility distribution describing, for each value w of the element weight, the extent to which it is a possible value. Equivalently, it means that the value of this weight belongs to a λ-cut interval $\widetilde{W}^{\lambda} = \{t : \mu_{\widetilde{W}}(t) \geq \lambda\}$ with confidence (or degree of necessity) $1 - \lambda$. Now to each feasible solution or element a *degree of possible optimality* and a *degree of necessary optimality* can be assigned. The notion of the necessary optimality of a solution may be weaken by assigning a *degree of necessary soft optimality*. Moreover, all the degrees of optimality of a solution (an element) can be derived from a *fuzzy deviation*, that is a possibility distribution representing the set of plausible values of deviations of the solution (the element) from optimum. In order to choose a "robust solution" under fuzzy weights, we adopt two criteria. The first one consists in choosing a solution of the maximum degree of necessary optimality, called a *best necessarily optimal solution*. Such a solution, if exists, seems to be an ideal choice. The second criterion is weaker than the first one and consists in choosing a solution of the maximum degree of necessary soft optimality, called a *best necessarily soft optimal solution*. This criterion has been originally proposed in [5, 6] for the linear programming problem with a fuzzy objective function. A best necessarily soft optimal solution seems to be a compromise choice and it minimizes a "distance" to the necessary optimality.

In this paper, we provide some methods for the optimality evaluation and for choosing a solution under fuzzy weights. In Section 3, we investigate the interval case, that is the problems in which the element weights are specified as closed intervals. It turns out, that it is possible to construct some polynomial algorithms for such problems if only their deterministic counterparts are polynomially solvable. In consequence, the interval bottleneck problems are easier to solve than the interval problems with a linear sum objective discussed in [7, 8]. We obtain polynomial algorithms for such problems as the bottleneck shortest path, the bottleneck assignment and the bottleneck matroid base. In Section 4, we show that the optimality evaluation and the problem of choosing a solution under fuzzy weights can be reduced to examining a family of interval problems. In particular, we show that a best necessarily soft optimal solution can be computed in polynomial time for a wide class of problems.

2 Preliminaries

In this section, we recall a formulation of a general bottleneck combinatorial problem with precise weights.

Let $E = \{e_1, \ldots, e_n\}$ be a finite ground set and let $\Phi \subseteq 2^E$ be a set of subsets of E called the set of the *feasible solutions*. A nonnegative real weight w_e is given for every element $e \in E$. A *bottleneck combinatorial optimization problem* $\mathcal{BP}$ consists in finding a feasible solution X that minimizes the weight of its heaviest element, namely:

$$\mathcal{BP}: \quad \min_{X \in \Phi} \max_{e \in X} w_e. \tag{1}$$

A solution that minimizes (1) is called an *optimal solution*. We call an element $e \in E$ *optimal* if it is a part of an optimal solution. If a solution (element) is not optimal a natural question arises how far from optimality this solution (element) is. To answer this question one can introduce the concept of a deviation. *A deviation of solution* $X \in \Phi$ and *a deviation of*

ISBN: 978-989-95079-6-8

element $f \in E$ are defined in the following way:

$$\delta_X = \max_{e\in X} w_e - \min_{Y\in\Phi} \max_{e\in Y} w_e,$$
$$\delta_f = \min_{Y\in\Phi_f} \max_{e\in Y} w_e - \min_{Y\in\Phi} \max_{e\in Y} w_e,$$

where Φ_f is the set of all feasible solutions that contain element f. It is clear that solution X (element f) is optimal if and only if $\delta_X = 0$ ($\delta_f = 0$).

In this paper we will assume that problem $\mathcal{BP}$ is polynomially solvable. Some polynomial algorithms for the bottleneck path, the bottleneck assignment, the bottleneck spanning tree, and the bottleneck matroid base problem can be found for instance in [1, 2, 3, 4]. For all these problems the deviations δ_X and δ_f can be computed in polynomial time.

3 Interval-valued bottleneck combinatorial optimization problems

In this section we consider an *interval* version of problem (1), in which the weights of the elements are ill-known and they are modeled by closed intervals $W_e = [\underline{w}_e, \overline{w}_e]$, $e \in E$. Assigning some interval to an element weight means that the actual weight of this element will take some value within the interval but it is not possible at present to predict which one. Every precise instantiation of the element weights is called a *scenario* and we denote it by S. Thus every scenario is a vector $S = (w_e)_{e\in E}$, $w_e \in W_e$ that expresses a realization of the weights. We denote by Γ the set of all the scenarios, i.e. $\Gamma = \times_{e\in E}[\underline{w}_e, \overline{w}_e]$ and we use $w_e(S)$ to denote the weight of element $e \in E$ in a fixed scenario $S \in \Gamma$. Among the scenarios of Γ, we distinguish the *extreme* ones, which belong to $\times_{e\in E}\{\underline{w}_e, \overline{w}_e\}$. Let $A \subseteq E$ be a fixed subset of elements. In scenario S_A^+ all elements $e \in A$ have weights $\overline{w}_e$ and all the remaining elements have weights $\underline{w}_e$. Similarly, in scenario S_A^- all elements $e \in A$ have weights $\underline{w}_e$ and all the remaining elements have weights $\overline{w}_e$. For a given solution $X \in \Phi$, we define its weight under a fixed scenario $S \in \Gamma$ as $F(X, S) = \max_{e\in X} w_e(S)$. We will denote by $F^*(S)$ the value of the weight of an optimal solution under scenario $S \in \Gamma$, that is $F^*(S) = \min_{X\in\Phi} F(X, S)$.

The optimality in the interval-valued problem $\mathcal{BP}$ can be characterized as follows: a given solution $X \in \Phi$ (element $f \in E$) is *possibly optimal* if and only if it is optimal in some scenario $S \in \Gamma$. A given solution $X \in \Phi$ (element $f \in E$) is *necessarily optimal* if and only if it is optimal in all scenarios $S \in \Gamma$. Similarly to the deterministic case, we can express the possible and necessary optimality in terms of the deviation. Let $\delta_X(S) = F(X, S) - F^*(S)$ and $\delta_f(S) = \min_{Y\in\Phi_f} F(Y, S) - F^*(S)$ denote the deviations of solution X and element f under a fixed scenario $S \in \Gamma$, respectively. Consider the following optimization problems:

$$\underline{\delta}_X = \min_{S\in\Gamma} \delta_X(S), \quad \overline{\delta}_X = \max_{S\in\Gamma} \delta_X(S)$$
$$\underline{\delta}_f = \min_{S\in\Gamma} \delta_f(S), \quad \overline{\delta}_f = \max_{S\in\Gamma} \delta_f(S).$$

The solutions to the above problems determine the so called *deviation interval* $\Delta_X = [\underline{\delta}_X, \overline{\delta}_X]$ containing all possible values of deviations for solution X. Similarly $\Delta_f = [\underline{\delta}_f, \overline{\delta}_f]$ is a deviation interval for element f. It is worth pointing that in literature (see e.g. [9]) the quantity $\overline{\delta}_X$ is called the *maximal regret* or *robust deviation* and it expresses the maximal possible deviation from optimum. Obviously, we can easily deduce the optimality of a solution from the deviation interval. Namely, X is possibly (resp. necessarily) optimal if and only if $\underline{\delta}_X = 0$ (resp. $\overline{\delta}_X = 0$). Similarly, an element f is possibly (resp. necessarily) optimal if and only if $\underline{\delta}_f = 0$ (resp. $\overline{\delta}_f = 0$).

3.1 Asserting possible and necessary optimality

In this section we establish some sufficient and necessary conditions for possible and necessary optimality of solutions and elements in the interval-valued problem. We start by proving the following proposition:

Proposition 1. *Let X be a given feasible solution. Then*

$$\underline{\delta}_X = \max\{0, \max_{e\in X} \underline{w}_e - F^*(S_E^+)\}, \tag{2}$$
$$\overline{\delta}_X = \max_{e\in X} \max\{0, \overline{w}_e - F^*(S_{\{e\}}^+)\}. \tag{3}$$

Proof. Equality (3) has been proved in [10]. We now give a proof sketch of (2). It is easy to verify that

$$\underline{\delta}_X \geq \max\{0, \max_{e\in X} \underline{w}_e - F^*(S_E^+)\}. \tag{4}$$

It remains to show that the inequality $\leq$ also holds in (4). Let Y be an optimal solution under S_E^+ and let $g = \arg\max_{e\in Y} \overline{w}_e$. We consider two cases. (i) $\max_{e\in X} \underline{w}_e > \overline{w}_g$. Denote $h = \arg\max_{e\in X} \underline{w}_e$. Consider scenario S such that $w_e^S = \min\{\underline{w}_h, \overline{w}_e\}$ for all $e \in X$ and $w_e^S = \overline{w}_e$ for all $e \in E \setminus X$. Since $\underline{w}_h \geq \underline{w}_e$ for all $e \in X$, $S \in \Gamma$. It is easy to check that $F(X, S) = \underline{w}_h$ and $F^*(S) = F(S_E^+)$. Hence $\underline{\delta}_X \leq \delta_X(S) = \max_{e\in X} \underline{w}_e - F^*(S_E^+) \leq \max\{0, \max_{e\in X} \underline{w}_e - F^*(S_E^+)\}$, which together with (4) yield (2). (ii) $\max_{e\in X} \underline{w}_e \leq \overline{w}_g$. Consider scenario S such that under this scenario all elements $e \in E \setminus X$ have weights $\overline{w}_e$ and all the elements $e \in X$ have weights $\min\{\overline{w}_e, \overline{w}_g\}$. Since $\underline{w}_e \leq \overline{w}_g$ for all $e \in X$, $S \in \Gamma$. One can easily verify that X is optimal under S, which means that $\underline{\delta}_X = 0 \leq \max\{0, \max_{e\in X} \underline{w}_e - F^*(S_E^+)\}$. This, together with (4), give (2). □

Proposition 1 allows us to compute the solution deviation interval Δ_X and leads to the following two corollaries:

Corollary 1. *A solution $X \in \Phi$ is possibly optimal if and only if $F(X, S_E^-) = \max_{e\in X} \underline{w}_e \leq F^*(S_E^+)$.*

Corollary 2. *A solution $X \in \Phi$ is necessarily optimal if and only if $\max_{e\in X} \max\{0, \overline{w}_e - F^*(S_{\{e\}}^+)\} = 0$.*

Making use of Corollaries 1 and 2, we can efficiently evaluate the optimality of a given solution X if only the underlying bottleneck deterministic problem $\mathcal{BP}$ is efficiently solvable. In order to evaluate the possible optimality of X, it suffices to compute the value of $F^*(S_E^+)$ and check if $F(X, S_E^-) \leq F^*(S_E^+)$. This can be done in $O(|X| + f(|E|))$ time, where $f(|E|)$ is the running time of an algorithm for the deterministic $\mathcal{BP}$ problem. Note that evaluating the possible optimality of a next solution, say $X' \in \Phi$, requires only $O(|X'|)$ time. Evaluating the necessary of optimality of X is a little more complex, since it requires computing the difference $\overline{w}_e - F^*(S_{\{e\}}^+)$ for each $e \in X$ and, consequently, the

ISBN: 978-989-95079-6-8

overall running time of evaluating the necessary optimality is $O(|X|f(|E|))$.

Let us consider the problem of computing an element deviation interval. The following proposition is true and its proof is similar to the proof of Proposition 1:

Proposition 2. *Let f be a specified element. Then*

$$\underline{\delta}_f = \max\{0, \min_{X\in\Phi_f} F(X, S_X^-) - F^*(S_E^+)\}.$$

From Proposition 2 we immediately get the following corollary:

Corollary 3. *An element $f \in E$ is possibly optimal if and only if* $\min_{X\in\Phi_f} F(X, S_X^-) \leq F^*(S_E^+)$.

Corollary 3 shows a significant difference between the problems with the bottleneck objective and the problems with the linear sum objective discussed for instance in [7, 8]. For the latter problems, deciding whether $\underline{\delta}_f = 0$ for a given element $f \in E$ may be NP-hard even if a deterministic counterpart is polynomially solvable [8]. For the bottleneck problems the situation is much better. From Proposition 2, it follows that if problem $\mathcal{BP}$ is solvable in $f(|E|)$ time, then the bound $\underline{\delta}_f$ for a given element f can be determined in $O(f(|E|))$ time.

We are unable here to provide a general formula for computing the upper bound of an element deviation $\overline{\delta}_f$. Also, the complexity status of the problem of checking whether f is necessarily optimal is unknown. This is an interesting subject of further research. In the next section we show how to compute efficiently quantities $\underline{\delta}_f$ and $\overline{\delta}_f$ when $\mathcal{BP}$ has a matroidal structure.

3.1.1 Matroidal problems

Let us recall the notion of a matroid (see e.g. [11]). A *matroid* is a system $(E, \mathcal{I})$, where E is a ground set and $\mathcal{I}$ is a set of subsets of E closed under inclusions (if $A \subseteq B$ and $B \in \mathcal{I}$ then $A \in \mathcal{I}$) and fulfilling the so-called *growth* property (if $A, B \in \mathcal{I}$ and $|A| < |B|$ then there is $e \in B \setminus A$ such that $A \cup \{e\} \in \mathcal{I}$). The maximal (under inclusion) elements of $\mathcal{I}$ are called *bases*. We will denote the set of all bases by $\mathcal{B}$. The minimal (under inclusion) sets not in $\mathcal{I}$ are called *circuits*. Matroids have the following property, which will be used in this section. Namely, if $B \in \mathcal{B}$ is a base and f is an element such that $f \notin B$, then $B \cup \{f\}$ contains the unique circuit C. Furthermore, for every $e \in C$, set $(B \cup \{f\}) \setminus \{e\}$ is a base.

In a *matroidal problem* the set of feasible solutions Φ consists of all bases of a given matroid, $\Phi = \mathcal{B}$. An example of a matroidal problem is bottleneck spanning tree, where Φ consists of all bases of a *graphic matroid* [11]. Due to a matroidal structure of the interval problem $\mathcal{BP}$, we can simplify the computation of $\underline{\delta}_f$ and provide a method of computing $\overline{\delta}_f$. We first show how to evaluate the possible and necessary optimality of a given element f.

Proposition 3. *An element $f \in E$ is possibly optimal if and only if $\underline{w}_f \leq F^*(S_E^+)$.*

Proof. Using Corollary 3, it is sufficient to proof that $\underline{w}_f \leq F^*(S_E^+)$ if and only if there is a base $B \in \mathcal{B}_f$ such that $F(B, S_E^-) \leq F^*(S_E^+)$, where $\mathcal{B}_f$ stands for the set of all bases that contain element f.

($\Leftarrow$) Obvious.

($\Rightarrow$) Let B^* be an optimal base in scenario S_E^+. If $f \in B^*$, then $B^* \in \mathcal{B}_f$, $F(B^*, S_E^-) \leq F(B^*, S_E^+) = F^*(S_E^+)$ and we are done. Otherwise, $B^* \cup \{f\}$ contains an unique circuit C. Set $B' = (B^* \setminus \{e\}) \cup \{f\}$, $e \in C \setminus \{f\}$, is a base and $B' \in \mathcal{B}_f$. Since $F^*(S_E^+) \geq \underline{w}_f$, we have $F^*(S_E^+) = \max_{e\in B^*} \overline{w}_e \geq \max_{e\in B'} \underline{w}_e$. Thus $F^*(S_E^+) \geq F(B', S_E^-)$ and the proposition follows. □

Proposition 4. *An element $f \in E$ is necessarily optimal if and only if $\overline{w}_f \leq F^*(S_{\{f\}}^+)$.*

Proof. ($\Rightarrow$) If f is necessarily optimal, then it is optimal under all scenarios, in particular, under $S_{\{f\}}^+$. Thus there is a base B containing f such that $F^*(S_{\{f\}}^+) = F(B, S_{\{f\}}^+)$, which yields $F^*(S_{\{f\}}^+) \geq \overline{w}_f$.

($\Leftarrow$) Consider any scenario $S \in \Gamma$. We first show that inequality $w_f(S) \leq F^*(S)$ holds for S. Define scenario S' where $w_f(S') = \overline{w}_f$ and the remaining elements, $e \neq f$, have weights $w_e(S') = w_e(S)$. Using the assumption we get $\overline{w}_f \leq F^*(S_{\{f\}}^+) \leq F^*(S')$. After decreasing the weight of f to $w_f(S)$, we get inequality $w_f(S) \leq F^*(S)$. Let B^* be an optimal base under S. If $f \in B^*$ then f is optimal under S. Otherwise, $B^* \cup \{f\}$ contains an unique circuit C. Set $B' = (B^* \setminus \{e\}) \cup \{f\}$, $e \in C \setminus \{f\}$, is a base such that $B' \in \mathcal{B}_f$. Since $w_f(S) \leq F^*(S)$, $F(B', S) \leq F(B^*, S) = F^*(S)$. Hence B' is an optimal base and f is optimal under S. □

Using a similar reasoning to that in Propositions 3 and 4 one can prove the following proposition, which allows us to determine the deviation interval $\Delta_f = [\underline{\delta}_f, \overline{\delta}_f]$.

Proposition 5. *Let X be a given feasible solution. Then*

$$\underline{\delta}_f = \max\{0, \underline{w}_f - F^*(S_E^+)\}, \overline{\delta}_f = \max\{0, \overline{w}_f - F^*(S_{\{f\}}^+)\}.$$

Propositions 3 and 4 allow us to evaluate efficiently the possible (resp. necessary) optimality of a specified element $f \in E$. It suffices to replace the interval weights with their exact values determined according to S_E^+ (resp. $S_{\{f\}}^+$), compute the optimal value of $F^*(S_E^+)$ (resp. $F(S_{\{f\}}^+)$) in the resulting deterministic problem and compare it with $\underline{w}_f$ (resp. $\overline{w}_f$). This can be done in $O(|E| \log^*(|E|))$ time [4], where $\log^* |E|$ is the iterated logarithm of $|E|$. Note also that using Proposition 3 we can detect all possibly optimal elements in $O(|E| \log^*(|E|))$ because we need to execute an algorithm for the deterministic problem only once. On the other hand, Proposition 4 does not allow us to detect all necessary optimal elements without extra effort. Computing the deviation interval Δ_f costs the same time as evaluating the possible and necessary optimality of f (see Proposition 5).

3.2 Choosing a robust solution

An important task in the interval-valued problem is to choose a *robust solution*, that is the one which performs reasonably well under any possible scenario. A necessarily optimal solution ($\overline{\delta}_X = 0$) is an ideal choice because it is optimal regardless of weight realizations. On the other hand, the possible optimality of a chosen solution ($\underline{\delta}_X = 0$) is the minimum requirement that should be satisfied. A possibly optimal solution

ISBN: 978-989-95079-6-8

always exists. But a necessarily optimal solution rarely exists. Hence the possible optimality is too weak criterion while the necessary optimality seems to be too strong. A solution that minimizes the maximal regret $\overline{\delta}_X$ seems to be a compromise choice. It minimizes the maximal possible deviation from optimum. In literature [9] such a solution is called an *optimal minmax regret solution*.

Proposition 6. *Every optimal minmax regret solution X is possibly optimal and is composed of possibly optimal elements.*

Proof. We use a proof by contraposition. Assume that X is not possibly optimal. From Corollary 1, we have $\max_{e\in X} \underline{w}_e > F^*(S_E^+)$. Let Y^* be an optimal solution in scenario S_E^+. Define $g = \arg\max_{e\in Y^*} \overline{w}_e$ and $h = \arg\max_{e\in X} \underline{w}_e$. Thus $w_h(S) > \overline{w}_g$ for all scenarios $S \in \Gamma$. But $\overline{w}_g \geq w_e(S)$ for all $e \in Y^*$ in every scenario S, which yields $w_h(S) > w_e(S)$ for $e \in Y^*$. From this we conclude that $F(X,S) > F(Y^*,S)$ for all $S \in \Gamma$, which implies $\overline{\delta}_X > \overline{\delta}_Y$. In consequence, X cannot be an optimal minmax regret solution. It is obvious that every possibly optimal solution is composed of possibly optimal elements. □

Therefore, any optimal minmax regret solution X is possibly optimal and it minimizes a distance to the necessary optimality. In other words its deviation interval is of the form $\Delta_X = [0, \overline{\delta}_X]$ where $\overline{\delta}_X$ is the smallest among all $X \in \Phi$. Fortunately, the problem of determining an optimal minmax regret solution can be efficiently solved if the underlying bottleneck deterministic problem $\mathcal{BP}$ is polynomially solvable [10]. Using Equality (3) we get

$$\min_{X\in\Phi} \overline{\delta}_X = \min_{X\in\Phi} \max_{e\in X} \hat{w}_e, \tag{5}$$

where weights $\hat{w}_e = \max\{0, \overline{w}_e - F^*(S^+_{\{e\}})\}$, $e \in E$, are deterministic. So, the minmax regret problem can be reduced to a deterministic problem $\mathcal{BP}$. It can be shown [10] that the minmax regret problem can be solved in $O(|E|+|X^*|f(|E|))$ time, where X^* is such that $F(X^*, S_E^-) = F^*(S_E^-)$ and $f(|E|)$ is the running time of an algorithm for $\mathcal{BP}$.

4 Fuzzy-valued bottleneck combinatorial optimization problems

In this section, we study problem $\mathcal{BP}$ with uncertain element weights, where the uncertainty is modeled by fuzzy intervals. We give a rigorous possibilistic interpretation, solution concepts and some solution algorithms in this setting.

4.1 A possibilistic formalization of the problem

We now give a formalization of problem $\mathcal{BP}$, in which the weights of elements of E are uncertain and they are modeled by fuzzy intervals $\widetilde{W}_e$, $e \in E$, in the setting of *possibility theory* [12]. Let us recall that a *fuzzy interval* $\widetilde{W}_e$ is a fuzzy set in the space of real numbers $\mathbb{R}$, whose membership function $\mu_{\widetilde{W}_e} : \mathbb{R} \to [0,1]$ is normal, quasiconcave and upper semicontinuous on $\mathbb{R}$. We will additionally assume that the support of $\widetilde{W}_e$ is bounded. A membership function of $\widetilde{W}_e$ is regarded as a *possibility distribution* for the values of the unknown weight $\mathfrak{w}_e$. The possibility degree of the assignment $\mathfrak{w}_e = v$ is $\Pi(\mathfrak{w}_e = v) = \pi_{\mathfrak{w}_e}(v) = \mu_{\widetilde{W}_e}(v)$. Let $S = (v_e)_{e\in E}$ be a scenario that represents a state of the world in which $\mathfrak{w}_e = v_e$, for all $e \in E$. Assuming that the weights are unrelated to one another, the fuzzy intervals associated with the weights induce the following possibility distribution over all scenarios in $\mathbb{R}^n$ (see [13]): $\pi(S) = \min_{e\in E} \Pi(\mathfrak{w}_e = v_e) = \min_{e\in E} \mu_{\widetilde{W}_e}(v_e)$ The degrees of possibility and necessity that a solution $X \in \Phi$ is optimal are defined as follows:

$$\begin{aligned} \Pi(X \text{ is optimal}) &= \sup_{\{S:\, X \text{ is optimal in } S\}} \pi(S), \\ \mathrm{N}(X \text{ is optimal}) &= \inf_{\{S:\, X \text{ is not optimal in } S\}} (1 - \pi(S)). \end{aligned} \tag{6}$$

As in the deterministic and interval cases we can characterize the optimality of a solution using the concept of deviation. In the fuzzy problem a solution deviation belongs to a fuzzy interval $\widetilde{\Delta}_X$, whose membership function $\mu_{\widetilde{\Delta}_X}$ is a possibility distribution for δ_X defined as follows:

$$\mu_{\widetilde{\Delta}_X}(v) = \Pi(\delta_X = v) = \sup_{\{S:\, \delta_X(S)=v\}} \pi(S). \tag{7}$$

Since the statement "X is optimal in S" is equivalent to condition $\delta_X(S) = 0$, it is easily seen that (7) generalizes the degrees of optimality (6), that is

$$\begin{aligned} &\Pi(X \text{ is optimal}) = \Pi(\delta_X = 0) = \mu_{\widetilde{\Delta}_X}(0), \\ &\mathrm{N}(X \text{ is optimal}) = \mathrm{N}(\delta_X = 0) = 1 - \sup_{v>0} \mu_{\widetilde{\Delta}_X}(v). \end{aligned} \tag{8}$$

Replacing X with f in (6)-(8) we get the same formulae for element f.

In practice the requirement $\delta_X = 0$ may be very strong and the degree of necessary optimality of every solution $X \in \Phi$ may be very small or even equal to 0. Suppose that a decision maker knows his/her preferences about δ_X and expresses it by a *fuzzy goal* $\widetilde{G}$. The membership function of the fuzzy goal $\mu_{\widetilde{G}}$ is a nonincreasing mapping from $[0,\infty]$ into $[0,1]$ such that $\mu_{\widetilde{G}}(0) = 1$. The value of $\mu_{\widetilde{G}}(\delta_X)$ expresses the degree to which deviation δ_X satisfies the decision maker. We can now replace the strong requirement "X is optimal" ($\delta_X = 0$) with weaker "X is soft optimal" ($\delta_X \in \tilde{G}$). Notice that $\delta_X \in \tilde{G}$ is a fuzzy event and the necessity that it holds can be computed as follows [5, 6]:

$$\mathrm{N}(X \text{ is soft optimal}) = \inf_S \max\{1 - \pi(S), \mu_{\widetilde{G}}(\delta_X(S))\}. \tag{9}$$

One can check that $\mathrm{N}(X \text{ is soft optimal}) = \alpha$ means that for all scenarios S such that $\pi(S) > 1 - \alpha$ it holds $\mu_{\widetilde{G}}(\delta_X(S)) \geq \alpha$ or equivalently $\delta_X(S) \in \widetilde{G}^\alpha = [0, \mu^{-1}_{\widetilde{G}}(\alpha)]$, which represents the suboptimality of X. Function $\mu^{-1}_{\widetilde{G}} : [0,1] \to \mathbb{R} \cup \{+\infty\}$ is a pseudo-inverse of $\mu_{\widetilde{G}}$ that is $\mu^{-1}_{\widetilde{G}}(\alpha) = \sup\{v : \mu_{\widetilde{G}}(v) \geq \alpha\}$.

4.2 Computing the optimality degrees

Every fuzzy interval $\widetilde{U}$ can be decomposed into its λ-cuts, i.e. sets $\widetilde{U}^\lambda = \{u : \mu_{\widetilde{U}}(u) \geq \lambda\}$, $\lambda \in (0,1]$. We will assume that $\widetilde{U}^0$ is the smallest closed set containing the support of $\widetilde{U}$. It can be easily verified (see e.g. [12]) that if $\widetilde{U}$ is a fuzzy interval with a bounded support, then $\widetilde{U}^\lambda$ is a closed interval for every $\lambda \in [0,1]$. The membership function $\mu_{\widetilde{U}}$ can be retrieved from the family of λ-cuts in the following way:

$$\mu_{\widetilde{U}}(u) = \sup\{\lambda \in [0,1] : u \in \widetilde{U}^\lambda\}, \tag{10}$$

ISBN: 978-989-95079-6-8

and $\mu_{\widetilde{U}}(u) = 0$ if $v \notin \widetilde{U}^0$.

Let us denote by $\mathcal{BP}^\lambda$, $\lambda \in [0,1]$, the interval-valued problem $\mathcal{BP}$ with element weights $\widetilde{W}_e^\lambda = [\underline{w}_e^\lambda, \overline{w}_e^\lambda]$, $e \in E$. Note that the scenario set in $\mathcal{BP}^\lambda$ is composed of all scenarios S such that $\pi(S) \geq \lambda$. Now $\widetilde{\Delta}_X^\lambda = [\underline{\delta}_X^\lambda, \overline{\delta}_X^\lambda]$ is the interval of possible values of the deviation of solution X in problem $\mathcal{BP}^\lambda$. It contains all values of the deviation of X, whose possibility of occurrence is not less than λ. From (8) and (10), it follows easily that

$$\Pi(X \text{ is optimal}) = \sup\{\lambda \in [0,1] : \underline{\delta}_X^\lambda = 0\} \tag{11}$$

and $\Pi(X \text{ is optimal}) = 0$ if $\underline{\delta}_X^0 > 0$. A similar reasoning leads to the following equality:

$$\mathrm{N}(X \text{ is soft optimal}) = 1 - \inf\{\lambda : \overline{\delta}_X^\lambda \leq \mu_{\widetilde{G}}^{-1}(1-\lambda)\}. \tag{12}$$

Replacing expression $\mu_{\widetilde{G}}^{-1}(1-\lambda)$ with 0 in (12) we get a formula for computing the degree of necessary optimality of X. Exactly the same considerations apply to elements. It is enough to replace X with f in formulae (11)–(12).

Equations (11) and (12) form a theoretical basis for calculating the values of the optimality degrees. The function $\underline{\delta}_X^\lambda$ is nondecreasing, so in order to compute the degree of possible optimality of solution X we can apply a binary search technique on $\lambda \in [0,1]$. Condition $\underline{\delta}_X^\lambda = 0$ means that X is possibly optimal in $\mathcal{BP}^\lambda$ and it can be checked efficiently by using Corollary 1. Similarly, $\overline{\delta}_X^\lambda$ is nonincreasing and $\mu_{\widetilde{G}}^{-1}(1-\lambda)$ is nondecreasing function of λ. Hence the degree of necessary soft optimality (and necessary optimality) can also be computed by using a binary search. We obtain $\overline{\delta}_X$ using Corollary 2. The calculations can be done in $\mathcal{O}(I(|E|)\log \epsilon^{-1})$ time, where $\epsilon > 0$ is a given precision and $I(|E|)$ is time required to find the bound (the lower or the upper) of deviation interval $\widetilde{\Delta}_X^\lambda$ in the corresponding interval problem $\mathcal{BP}^\lambda$. Of course the same reasoning can be repeated for elements.

Fuzzy deviations $\widetilde{\Delta}_X$ and $\widetilde{\Delta}_f$ can also be determined via the use of λ-cuts. That is, we compute interval $\widetilde{\Delta}_X^\lambda = [\underline{\delta}_X^\lambda, \overline{\delta}_X^\lambda]$ (resp. $\widetilde{\Delta}_f^\lambda$) in the interval valued $\mathcal{BP}^\lambda$ for suitably chosen λ-cuts. Then the fuzzy quantity $\widetilde{\Delta}_X$ (resp. $\widetilde{\Delta}_f$) is reconstructed from its λ-cuts by using equality (10). This method gives an approximation of $\widetilde{\Delta}_X$ (resp. $\widetilde{\Delta}_f$). Its overall running time is $\mathcal{O}(rI(|E|))$ time, r is the number of chosen λ-cuts and $I(|E|)$ is time required to determine $\widetilde{\Delta}_X^\lambda$ (resp. $\widetilde{\Delta}_f^\lambda$) in the interval problem $\mathcal{BP}^\lambda$.

From the above it follows that the running time of the proposed methods heavily relies on the interval case, because they can be reduced to the optimality analysis in a sequence of the interval-valued problems $\mathcal{BP}^\lambda$. Thus the main difficulty lies in the interval case. Fortunately, the optimality analysis in the interval case can be efficiently done, see Corollaries 1,2 and Proposition 1 for a solution and Corollary 3 and Proposition 2 for an element. Unfortunately, we cannot give an efficient method for asserting necessary optimality and determining the upper bound on possible values of the deviation of an element in a general problem $\mathcal{BP}^\lambda$. However, we have proposed such method when problem $\mathcal{BP}^\lambda$ has a matroidal structure (see Section 3.1.1).

4.3 Choosing a robust solution

We now propose some concepts of choosing a robust solution in the fuzzy-valued problem $\mathcal{BP}$. An ideal choice is a solution with the highest degree of certainty that it is optimal under all possible scenarios, i.e. an optimal solution to the following problem:

$$\max_{X \in \Phi} \mathrm{N}(X \text{ is optimal}) = \max_{X \in \Phi} \mathrm{N}(\delta_X = 0). \tag{13}$$

A solution of (13) is called a *best necessarily optimal solution*. However, this concept has a drawback, since the criterion used in (13) is very strong. Namely, a solution X such that $\mathrm{N}(X \text{ is optimal}) > 0$ may not exist or even if it exists, its necessary optimality degree may be very small. To overcome this drawback, we can replace the degree of necessary optimality with the degree of soft necessary optimality, which leads to the following problem:

$$\max_{X \in \Phi} \mathrm{N}(X \text{ is soft optimal}). \tag{14}$$

A solution to (14) is called a *best necessarily soft optimal solution*. Let us recall that this solution concept was first applied to fuzzy linear programming in [5, 6]. We can see from (12) that problem (14) is equivalent to the following one:

$$\begin{array}{ll} \min & \lambda \\ s.t. & \overline{\delta}_X^\lambda \leq \mu_{\widetilde{G}}^{-1}(1-\lambda), \\ & \lambda \in [0,1], \\ & X \in \Phi. \end{array} \tag{15}$$

If problem (15) is infeasible then $\mathrm{N}(X \text{ is soft optimal}) = 0$ for all solutions $X \in \Phi$. If λ^* is the optimal objective value of (15) and X is a best necessarily soft optimal solution, then $\mathrm{N}(X \text{ is soft optimal}) = 1 - \lambda^*$. Since $\overline{\delta}_X^\lambda$ is nonincreasing and $\mu_{\widetilde{G}}^{-1}(1-\lambda)$ is nondecreasing function of λ, problem (15) can also be solved by the binary search technique on $\lambda \in [0,1]$. Note that $\overline{\delta}_X^\lambda$ is the maximal regret of X in $\mathcal{BP}^\lambda$ (see Section 3.2). Thus to find the optimal λ^*, we seek solution that minimizes $\overline{\delta}_X^\lambda$ for a fixed λ, i.e. an optimal minmax regret solution in $\mathcal{BP}^\lambda$ (see problem (5)). An optimal solution X^* to problem (15), determined by finding an optimal minmax regret solution at λ-cut, has a nice property. Namely, it is an optimal minmax regret solution in $\mathcal{BP}^{\lambda^*}$ and, by Proposition 6, it is also possibly optimal one in $\mathcal{BP}^{\lambda^*}$. Thus $\mathrm{N}(X^* \text{ is soft optimal}) \leq \Pi(X^* \text{ is optimal})$. Consequently,

$$\begin{array}{rcl} \mathrm{N}(X^* \text{ is optimal}) & \leq & \mathrm{N}(X^* \text{ is soft optimal}) \\ & \leq & \Pi(X^* \text{ is optimal}). \end{array}$$

From this, we conclude that the concept of a necessary soft optimality is a natural extension of the minmax regret approach to the fuzzy case. The running time of the method based on the binary search is $\mathcal{O}(I(|E|)\log \epsilon^{-1})$ time, where $\epsilon > 0$ is a given precision and $I(|E|)$ is time required for seeking an optimal minmax regret solution in $\mathcal{BP}^\lambda$. It is polynomial for the bottleneck path, the bottleneck assignment, the bottleneck spanning tree and the bottleneck matroid base problems.

ISBN: 978-989-95079-6-8

4.4 A parametric approach

We now present a parametric approach to finding a best necessary optimal solution. A similar method can be applied to compute fuzzy deviations $\widetilde{\Delta}_X$ and $\widetilde{\Delta}_f$. It is easy to check that the following parametric problem may be associated with problem (15) (see also problems (15) and (5))

$$\overline{\delta}^{\lambda} = \min_{X\in\Phi} \overline{\delta}^{\lambda}_X = \min_{X\in\Phi} \max_{e\in X}\{\hat{w}^{\lambda}_e\}, \quad \lambda \in [0,1], \tag{16}$$

where $\hat{w}^{\lambda}_e = \max\{0, \overline{w}_e - F^*(S^{+\lambda}_{\{e\}})\}$, $e \in E$, are parametric weights (functions of a parameter λ), $S^{+\lambda}_{\{e\}}$ is the scenario in which we fix the weight of element e to $\overline{w}^{\lambda}_e$ and the weights of elements $g \in E \setminus \{e\}$ to $\underline{w}^{\lambda}_g$. In order to solve (16), we need to compute parametric weights $\hat{w}^{\lambda}_e = \max\{0, \overline{w}_e - F^*(S^{+\lambda}_{\{e\}})\}$, $e \in E$. Determining these weights requires computing functions $F^*(S^{+\lambda}_{\{e\}})$ for each $e \in E$. This can be done by applying known parametric techniques. Observe that $F^*(S^{+\lambda}_{\{e\}})$ is a function of parameter $\lambda \in [0,1]$. We wish to compute sequences $0 = \lambda_0 < \lambda_1 < ... < \lambda_k = 1$ and $X_0, \ldots, X_{k-1}$ such that X_i is an optimal solution for $\lambda \in [\lambda_i, \lambda_{i+1}]$, $i = 0, \ldots, k-1$. Having these sequences it is easy to describe function $F^*(S^{+\lambda}_{\{e\}})$ and, in consequence, $\hat{w}^{\lambda}_e$ for $\lambda \in [0,1]$. It turns out that if elements weights are linear functions of λ for each $e \in E$, then for some particular bottleneck problems their parametric counterparts can be efficiently solved (see e.g. [14]). Now, we are ready to solve (16) with parametric weights $\hat{w}^{\lambda}_e$. Again, we compute sequences $0 = \lambda_0 < \lambda_1 < ... < \lambda_k = 1$ and optimal solutions (optimal minmax regret solutions) $X_0, \ldots, X_{k-1}$ and the upper bounds for their deviations $\overline{\delta}^{\lambda}_{X_i}$, $\lambda \in [\lambda_i, \lambda_{i+1}]$, $i = 0, \ldots, k-1$. Values $\overline{\delta}^{\lambda}_{X_i}$, $\lambda \in [\lambda_i, \lambda_{i+1}]$, $i = 0, \ldots, k-1$, provide an analytical description of function $\overline{\delta}^{\lambda}$ (deviations of optimal minmax regret solutions) for $\lambda \in [0,1]$. To find a best necessary optimal solution, we need solve the following problem

$$\lambda^* = \arg \min_{\lambda\in[0,1]} \{\overline{\delta}^{\lambda}, \mu^{-1}_{\widetilde{G}}(1-\lambda)\},$$

which is equivalent to finding an intersection of $\overline{\delta}^{\lambda}$ and $\mu^{-1}_{\widetilde{G}}(1-\lambda)$. An optimal minmax regret solution that corresponds to the optimal value λ^* is a best necessary optimal solution.

5 Conclusions

In this paper, we have studied a general bottleneck combinatorial optimization problem with uncertain element weights modeled by fuzzy intervals, whose membership functions are regarded as possibility distributions for the values of the unknown weights. We have described, in this setting, the notions of possible and necessary optimality of a solution and an element and the necessary soft optimality of a solution. These notions are natural generalizations of the ones introduced in the interval-valued case. In order to choose a robust solution, we compute a best necessary soft optimal solution. This concept of choosing a solution is also a generalization, to the fuzzy case, of the known from literature minmax regret criterion [9]. Thus, we have shown in the paper that there exists a link between interval and possibilistic uncertainty. Hence, we have discussed first the interval-valued case and then we have extended the notions and the methods introduced for the interval-valued problem to the fuzzy-valued one. Indeed, the optimality evaluation and choosing a robust solution in the fuzzy problem boil down to solving a number of interval problems $\mathcal{BP}^{\lambda}$. Both problems can be solved in polynomial time if the corresponding interval counterparts are polynomially solvable, that holds true for a wide class of classical bottleneck combinatorial problems.

Acknowledgment

The first author was partially supported by Polish Committee for Scientific Research, grant N N111 1464 33.

References

[1] Abraham P. Punnen. A linear time algorithm for the maximum capacity path problem. *European Journal of Operational Research*, 53:402–404, 1991.

[2] Harold N. Gabow and Robert E. Tarjan. Algorithms for two bottleneck optimization problems. *Journal of Algorithms*, 9:411–417, 1988.

[3] Paolo M. Camerini. The minimax spanning tree problem and some extensions. *Information Processing Letters*, 7:10–14, 1978.

[4] Abraham P. Punnen. A fast algorithm for a class of bottleneck problems. *Computing*, 56:397–401, 1996.

[5] Masahiro Inuiguchi and Masatoshi Sakawa. Robust optimization under softness in a fuzzy linear programming problem. *International Journal of Approximate Reasonning*, 18:21–34, 1998.

[6] Masahiro Inuiguchi. On possibilistic/fuzzy optimization. In *Foundations of Fuzzy Logic and Soft Computing, IFSA 2007*, Lecture Notes in Artificial Intelligence 4529, pages 351–360. Springer-Verlag, 2007.

[7] Adam Kasperski and Paweł Zieliński. On possibilistic combinatorial optimization problems. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 324–329, Hong Kong, 2008.

[8] Adam Kasperski and Paweł Zieliński. Minmax regret approach and optimality evaluation in combinatorial optimization problems with interval and fuzzy weights. *European Journal of Operational Research*, to appear.

[9] Panos Kouvelis and Gang Yu. *Robust Discrete Optimization and its applications*. Kluwer Academic Publishers, 1997.

[10] Igor Averbakh. Minmax regret solutions for minimax optimization problems with uncertainty. *Operations Research. Letters*, 27:57–65, 2000.

[11] C. Papadimitriou and K. Steiglitz. *Combinatorial Optimization*. Dover Publications, INC, Mineola, New York, 1998.

[12] Didier Dubois and Henri Prade. *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, New York, 1988.

[13] Didier Dubois, Hélène Fargier, and Vincent Galvagnon. On latest starting times and floats in activity networks with ill-known durations. *European Journal of Operational Research*, 147:266–280, 2003.

[14] Timothy M. Chan. Finding the shortest bottleneck edge in a parametric minimum spanning tree. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 232—-240, 2005.

ISBN: 978-989-95079-6-8

Web Usage Mining: users' navigational patterns extraction from web logs using Ant-based Clustering Method

Kobra Etminani [1] Mohammad-R. Akbarzadeh-T. [2] Noorali Raeeji Yanehsari [3]

1 Dept. of Computer Engineering, Ferdowsi University of Mashhad, Iran
2 Depts of Electrical Engineering and Computer Engineering, Ferdowsi University of Mashhad, Iran
3 Iran Khodro, Khorasan, Iran
Email: etminani@wali.um.ac.ir, akbarzadeh@ieee.org, raeeji@ikkco.ir

Abstract— *Web Usage Mining is the process of applying data mining techniques to the discovery of usage patterns from data extracted from Web Log files. It mines the secondary data (web logs) derived from the users' interaction with the web pages during certain period of Web sessions. Web usage mining consists of three phases, namely preprocessing, pattern discovery, and pattern analysis. In this paper, web logs of our university web server logs (http://www.um.ac.ir/) are pre-processed. Then, ant-based clustering is applied to pre-processed logs to extract frequent patterns for pattern discovery and then it is displayed in an interpretable format. Result of this paper would be useful for our university web site owner.*

Keywords— web usage mining, ant-based clustering, frequent pattern extraction, web mining.

1 Introduction

Web usage information mining could help to engage new customers, maintain current customers, track customers who are leaving web site, and so on [1]. Usage information can be extracted to increase web server efficiency by pre-fetching and caching strategies [2].Based on several researches done in the area of web mining, we can broadly classify it into three domains: web content mining, web structure mining, and web usage mining.

Web content mining is the process of extracting knowledge from web documents such as text and multimedia. Knowledge extraction from the structure of web and hyperlink references is called web structure mining. Web usage mining is the process of knowledge exploitation from the secondary data [3]. By secondary data, we mean browser logs, user profiles, web server access logs, registration data, user sessions or transactions, cookies, user queries, mouse clicks and any other data that is the result of interaction with the Web.

In the following section we give an overview over the related work. Section 3 explains the problem and ant-based clustering in more detail. Section 4 goes into detail how we implement the proposed method i.e. the experimental procedure of the proposed method and results are shown. We concluded our work in Section 5.

2 Related Work

There are several methods for pattern extraction from the secondary data (web logs) such as Masseglia et al. [2] [7], Spiliopoulou et al. [8], Bonchi et al. [9], Hay et al. [10], Zhu et al. [11], Nakagawa and Mobasher [12]. In [2], Masseglia et al. invented a new method named PSP (Prefix-tree for Sequential Patterns) that follows the basic principles of GSP (Generalized Sequential Patterns algorithm) in [13]. The tree structure in PSP is similar to the *prefix-tree* used in [14]. At the *k*th step, the tree has a depth of *k*. Each branch from the root to a leaf stands for a candidate sequence.

In [8], Spiliopoulou et al. proposed the WUM (Web utilization miner) tool that determines patterns which are noticeable from the statistical view. So, it emphasizes the frequency (minimum support) of the patterns.

Hay et al. in [10] applied the notion of time embedded in the navigations to cluster user sessions. They used an alignment algorithm to compute the distance between sessions.

Zhu et al. in [11] considered navigating between web sites as a Markov chain and mainly discussed about Markov model problems.

In [12], Nakagawa and Mobasher show that depending on whether the propositions are based on frequent itemsets or frequent sequences, the features of the site have an impact on the quality of the refinement shown to users.

In this paper, we propose a new method for extracting patterns from web logs based on ant clustering algorithm. We apply ant-based clustering for pattern discovery, other similar methods applied ant colony clustering to segregate visitors [15]. Some methods applied Markov models for modeling user web navigation behavior. But the proposed method has the similarity and speed of ant-based clustering algorithm rather than other clustering algorithms.

3 Problem description

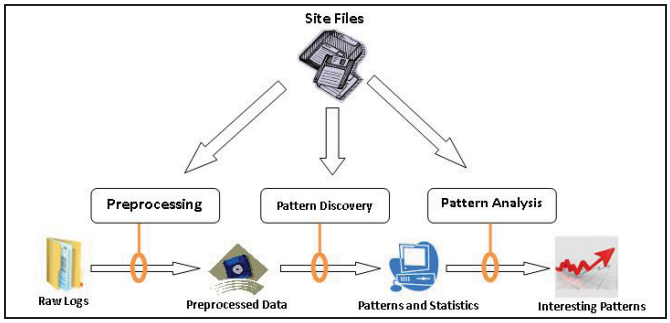


Figure 1: Usage Mining Process.

As shown in Figure 1 from [6], there are three main tasks for performing Web Usage Mining: Pre-processing, Pattern

ISBN: 978-989-95079-6-8

Discovery and Pattern Analysis. This section presents an overview of the tasks for each step.

At least two log file formats exists: Common Log File format (CLF) and Extended Log File format ([16] for more details). Our university log file consists of these fields: Date, Time, client IP address, Method, URI stem, Protocol status, Bytes sent, Protocol version, Host, User Agent and Referrer.

3.1 Pre-processing

As said in [6], pre-processing *"consists of converting the usage, content, and structure information contained in the various available data sources into the data abstractions necessary for pattern discovery"*.

This step can break into at least four substeps: Data Cleaning, User Identification, Session Identification and Formatting.

Unneeded data will be deleted from raw data in web log files in the data cleaning step. When a user requests a page, the request is added to the Log File; but if this page contains images, javascripts, flash animations, video, etc., they are added to the Log file as well. Most of the time, these are not needed for pattern discovery and should be omitted from log files.

There are some methods for User Identification, the second phase: detecting cookies, Identd, through IP address and user name. W3C [17] define cookies as *"data sent by the server to the client, data locally stored in cookies and is sent to the server with each request"*. [18] and [19] use cookies to identify users. But they have two main problems: the users can lock the use of cookies, so server cannot store information locally in the user machine; other problem is the user can delete the cookies.

Identd can be used for user identification. It is a protocol defined in RFC 1413 [20]. It allows us to identify connected users with the unique TCP connection. The problem with Identd is that users should configure with this protocol.

Another way for user detection is through user names added in the log file in field authuser. But this field can be empty (default value dash(-)) according to server/user command.

At last we can identify users through their IP address registered in each record in log file. We used this method although it has several problems: different users can be registered with same IP address; same user can be registered with different IP addresses. But with the help of session identification we can identify a user with IP address and be sure of solving the first problem. The second problem is not important because in this paper, a specific user is not wanted. We want to recognize general user's pattern.

For Session Identification in third phase, first we should define session timeout. Different timeouts considered are between 25-30 minutes .The thirty minute timeout is based on the results of [21]. We assume 30 minutes session timeout for the experimental procedure.

And in the last phase of pre-processing step, we should display pre-processed data in a suitable format, for the second step, pattern discovery.

3.2 Pattern Discovery

As stated in [6], pattern discovery *"draws upon methods and algorithms developed from several fields such as statistics, data mining, machine learning and pattern recognition"*. Several methods and techniques have already been developed for this step as summarized below:

- *Statistical Analysis* such as frequency analysis, mean, median, etc.
- *Clustering* of users help to discover groups of users with similar navigation patterns (provide personalized Web content).
- *Classification* is the technique to map a data item into one of several predefined classes.
- *Association Rules* discover correlations among pages accessed together by a client.
- *Sequential Patterns* extract frequently occurring inter-session patterns such that the presence of a set of items s followed by another item in time order.
- *Dependency Modeling* determines if there are any significant dependencies among the variables in the Web.

We choose clustering to discover users' navigational patterns. Our clustering method is based on Ant-based Clustering algorithm explained in Section 3.4.

3.3 Pattern Analysis

Pattern Analysis is the final stage of WUM (Web Usage Mining), which involves the validation and interpretation of the mined pattern.

- *Validation*: to eliminate the irrelevant rules or patterns and to extract the interesting rules or patterns from the output of the pattern discovery process.
- *Interpretation*: the output of mining algorithms is mainly in mathematic form and not suitable for direct human interpretations.

3.4 Ant-based Clustering

Deneubourg et al. in [22] proposed ant-based clustering and sorting. In the case of ant-based clustering and sorting, two related types of natural ant behaviors are modeled. When clustering, ants gather items to form heaps. And when sorting, ants discriminate between different kinds of items and spatially arrange them according to their properties [23].

Lumer and Faieta in [24] proposed ant-based data clustering algorithm (shown in Figure 2), which resembles the ant behavior described in [22].

As shown in Figure 2, the agents (ants) and data are randomly initialized on a toroidal grid. By moving agents, data is sorted according to its neighbors.

The picking and dropping probabilities, given a grid position and a particular data item *i*, are computed using the density functions:

$$p_{pick}(i) = \left(\frac{k^{+}}{k^{+} + f(i)}\right)^{2} \quad (1)$$

$$p_{drop}(i) = \begin{cases} 2f(i) & if\ f(i) < k^{-} \\ 1 & otherwise, \end{cases} \quad (2)$$

ISBN: 978-989-95079-6-8

where $k+$ and $k-$ are constants, and $f(i)$ is a neighborhood function:

$$f(i) = \max(0, \frac{1}{\sigma^2} \sum_{j \in L} (1 - \frac{d(i,j)}{\alpha})) \quad (3)$$

where, $d(i, j) \in [0, 1]$ is a measure of the dissimilarity between data points i and j , $\alpha \in [0, 1]$ is a data-dependent scaling parameter, and σ^2 is the size of the local neighborhood L.

Handl & Meyer in [25] proposed an extension of this algorithm where the parameter α is adaptively updated during the execution of the algorithm.

We applied Handl & Meyer's Ant-based clustering algorithm for detecting user's patterns.

```
(1)  Procedure Lumer and Faieta
(2)     randomly scatter data items on the toroidal grid
(3)     randomly place agents on the toroidal grid
(4)     for t = 1 to max_iterations
(5)        j = random agent
(6)        move agent j randomly by stepsize grid cells
(7)        l = does agent j carry a data item?
(8)        e = is agent j's grid position occupied by a data
           item?
(9)        if (i = TRUE) and (e = FALSE) then
(10)          i = data item carried by agent j
(11)          drop = (random() ≤ Pdrop(i)) // see
              equations (2) and (3)
(12)          if drop = TRUE then
(13)             Let agent j drop data item i at its current
                 position
(14)          end if
(15)       end if
(16)       if (l = FALSE) and (e = TRUE) then
(17)          i = data item at agent j's grid position
(18)          pick = (random() ≤Ppick(i)) // see equations
              (1) and (3)
(19)          if pick = TRUE then
(20)             let agent j pick up data item i
(21)          end if
(22)       end if
(23)    end for
(24) end procedure
```

Figure 2: Lumer & Faieta 's ant-based clustering algorithm [24].

4 Experimental Procedure and Results

4.1 Experimental Procedure

We used our university (http://www.um.ac.ir/) web server logs of two weeks for the experimental procedure.

For the first step, i.e. pre-processing, we wrote a c++ program compiled using gcc without any optimization flags. As mentioned in section 3.1, this step contains four phases. First, we omit unneeded records from log file. The log file consists of image files (gif, jpg, bmp, jpeg …) and other unneeded resources like javascripts and errors. For user identification, we use IP address and session timeout of 30 minutes. So, a user with an IP address has 30 minutes to navigate in the web site. After a user's navigational sequence is extracted, it is displayed in a suitable format for the second step, Pattern Discovery. We then classify the URLs of the web site into 28 groups and assign a number to each group, as shown in Table 1. Then, each user's requested URL is substituted with its corresponding number. The output of this step is a file that consists of records, each record representing a navigational sequence of users in numbers.

For the second step, Patten Discovery, we used Ant-based Clustering algorithm based on [25]. Julia Handl's written source code is used in Java and changes are made according to our application. Each user's navigational sequence is defined as an array with 28 elements. The element i is 1 if the related user had seen one of the pages in group i; otherwise it is 0.

Dissimilarity of two sequences $s1$ at point i and $s2$ at point j in the grid is computed through the following equation:

$$d(i,j) = \frac{\sqrt{\sum_{k=1}^{28} (s1_k - s2_k)^2}}{28} \quad (4)$$

$d(i,j)$ becomes 1 when two sequences do not have any similar elements, and becomes 0 when they are exactly the same.

Table 1: Classification of URLs.

1	Web site content	2	Newsletter
3	Miscellaneous	4	Black board
5	Web services	6	Search
7	News	8	Help
9	FAQ	10	Societies
11	Publications	12	Abstract
13	Photos	14	Web links
15	RSS	16	People
17	Staff	18	Student
19	Faculty	20	Professor
21	User	22	Guest
23	Research	24	About university
25	Education	26	Download
27	English homepage	28	homepages

The output of this program is a grid that contains numbers: >-1 and =-1 indicates if there is/is not a data item, respectively. So, clusters should be extracted according to these numbers and size of the local neighborhood, σ^2. Figure 3 shows the positions of the data points in different phases of running this algorithm. Figure 3.a shows the distribution of data points at the first step of program execution. As the execution continues, Figures 3.b, 3.c and 3.d shows the results after 200, 400 and 600 iterations, respectively. Clusters are created through moving of the ants.

We examined several numbers of ants for the clustering step. The experiments shows that the smaller the number of ants, the slower will be the rate of convergence, but also better results in clustering the data items. On the other hand, the larger the number of ants, the faster will be the speed of convergence, but also weaker results in clustering. So, empirically, we choose number of ants to be %20 of the number of data items.

At last for the third step, Pattern Analysis, the results are shown in an interpretable way. The output of the second step

ISBN: 978-989-95079-6-8

is the clusters of users' navigational sequence. Each cluster may include lots of members, so we should represent a cluster with one pattern and display it in a suitable format for users.

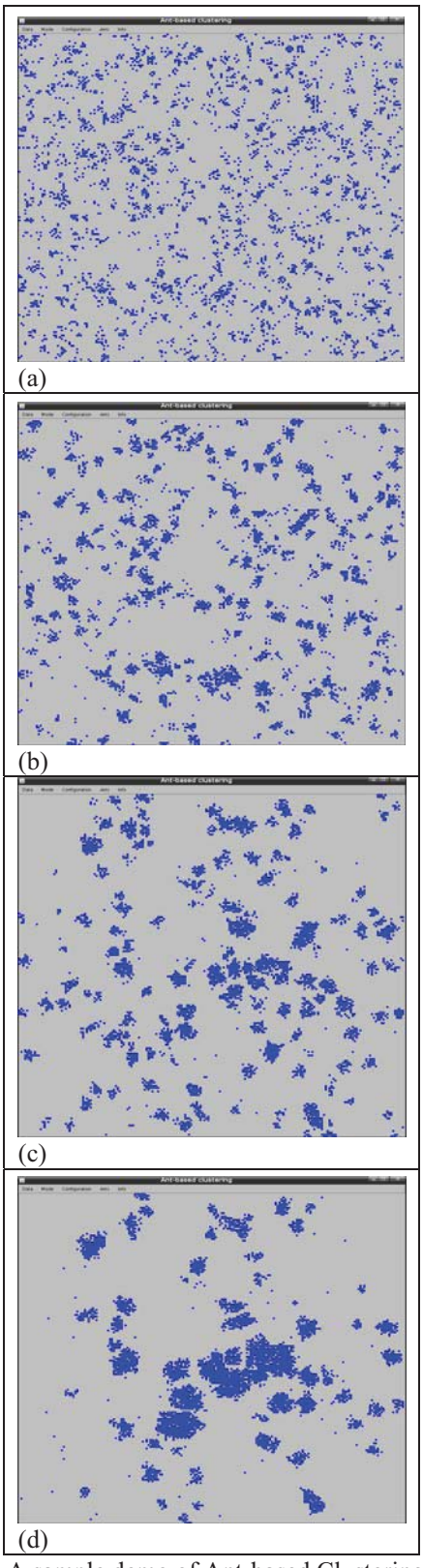


Figure 3: A sample demo of Ant-based Clustering of users' sequences.

The clustering algorithm results in clusters of similar sequences, which is a key element for sequence alignment. The alignment of sequences leads to a weighted sequence (as defined in [26]), represented as follows: $SA = \{I1 : n1, I2 : n2, \ldots, Ir, nr\} : m$. In this representation, m stands for the total number of sequences involved in the alignment. Ip ($1 \leq p \leq r$) is an itemset represented as ($xi1 : mi1 , \ldots xit : mit$), where mit is the number of sequences containing the item xi at the npth position in the aligned sequences. Finally, np is the number of occurrences of itemset Ip in the alignment. Figure 4 shows an example of the alignment process. The approximate sequential pattern can be obtained by specifying k: the number of occurrences of an item in order for it to be displayed. For instance, with the sequence $SA14$ from Figure 4 and $k = 3$, the filtered aligned sequence will be: {2, 4, 8}

To show the results, the aligned patterns that consist of numbers is re-substituted by the groups in Table 1.

S1={	2	4	5	8}
S2={	2		5	8}
SA12={	(2:2):2	(4:1):1	(5:2):2	(8:2):2}
SA12={	(2:2):2	(4:1):1	(5:2):2	(8:2):2}
S3={	2	4		8}
SA13={	(2:3):3	(4:2):2	(5:2):2	(8:3):3}
SA13={	(2:3):3	(4:2):2	(5:2):2	(8:3):3}
S4={	2	4		8}
SA14={	(2:4):4	(4:3):3	(5:2):2	(8:4):4}

Figure 4: Alignment processing example.

4.2 Results

The above mentioned procedure is applied to our university (http://www.um.ac.ir/) web server logs. All the experiments were performed on an Intel Core 2 Duo 2.5GHz PC running Linux (Mint).

The log files are collected in two different weeks in 2008: first week of June and the middle week of August each have 300 MB. Two different weeks are selected and the result of applying the proposed method on each week is compared. Applying ant-based clustering on each week, on average, took 300 seconds for 800 iterations. Average hourly web traffic for each group in these two weeks can be seen in Figures 5 and 6, respectively. The extracted behaviors from these two weeks are listed in Table 2 and Table 3, respectively. As shown in Table 2 and Table 3, behaviors of users in these two weeks were similar in most of the cases. One should notice that these results show just most frequent patterns of users' navigational behaviors, not all of the users' behaviors. Patterns of single navigations are not listed above, i.e. patterns that contain only one navigation.

Comparing Table 2 and Table 3 with Figure 5 and Figure 6, one could understand that pattern <Web site content → News > is the most popular one. <Professor → Homepages> is frequent, too. On the other hand, groups that do not exist in the patterns, like FAQ and Societies, have the least access in hours of a day, too.

Comparing our method to other methods, it has some advantages:

- It is simple. One only needs to define a suitable dissimilarity function for the clustering step and do

ISBN: 978-989-95079-6-8

not need to involve in complex mathematical relations.

- It does not depend on pattern length. Other similar methods are limited in terms of the length of the extracted patterns. The proposed method is able to extract patterns of any length.

Different patterns can be extracted depending on the occurrence of that page group in a cluster. So, even page groups that are less accessed can be extracted.

These kinds of results are useful for web site owners. They can put their advertisements in these sequences, because these are the most frequent ones. They are useful for page pre-fetching, too.

5 Conclusions

In this paper, we have proposed a new method to extract navigational patterns from web logs. Ant-based clustering has been used for this purpose. It needs a neighborhood function to be defined for. After the clustering is completed, alignment processing has been applied to the extracted sequences in each cluster and extract the representative for each cluster.

We apply the following procedure on our university web server logs for two different weeks (http://www.um.ac.ir/) and the results are satisfactory and true according to the hourly web traffic. These kinds of results are suitable for web site owners, for example, to put their advertisements there or to even change the structure of the web site according to users' navigational behavior.

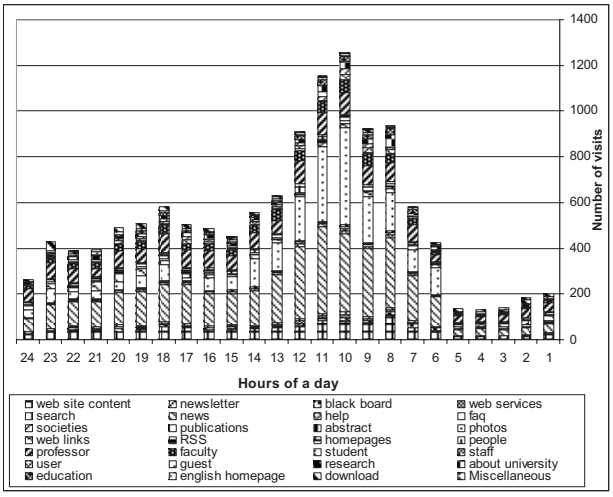


Figure 5: Hourly web traffic of the first week.

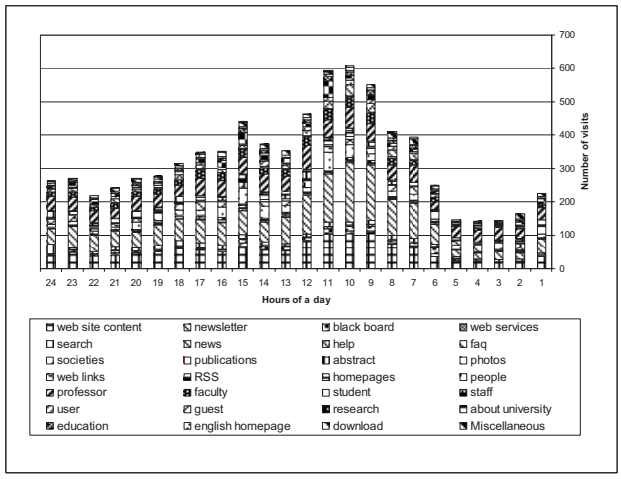


Figure 6: Hourly web traffic of the second week.

Table 2: Extracted behaviors from the first week.

Web site content	→	News						
Web services	→	Search						
Web site content	→	News	→	Faculty	→	Education		
News	→	Faculty	→	Education				
News	→	Professor						
Web site content	→	News	→	Professor	→	Homepages		
Web site content	→	News	→	Help	→	About university	→	Education
Web site content	→	News	→	Search	→	Photos	→	Web links → Professor → About university → Downloads
Web site content	→	Black board						
English homepage	→	Guest						
Professor	→	Homepages						
Professor	→	Faculty						
News	→	Faculty						

Table 3: Extracted behaviors from the second week.

Web site content	→	News						
Faculty	→	English Homepage						
Web site content	→	News	→	Users				
Web site content	→	Users						
Web site content	→	Search	→	News	→	Web links	→	Professor → About university → Download
Web site content	→	Faculty						
News	→	Professor	→	Homepages				
Faculty	→	Education						
Web site content	→	Photos	→	Web links	→	Professor	→	Homepages → English homepage → Download
News	→	Professor						
News	→	Download						

ISBN: 978-989-95079-6-8

References

[1] A. Abraham. Natural Computation for Business Intelligence from Web Usage Mining, *Proceeding of Seventh International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNAC2005),* pp. 3-11, 2005.

[2] F. Masseglia, P. Poncelet, and R. Cicchetti, An Efficient Algorithm for Web Usage Mining, *Networking and Information Systems Journal (NIS)*, 2(5-6), pp. 571-603, 1999.

[3] R. Cooley, Web Usage Mining: Discovery and Application of Interesting patterns from Web Data, Ph. D. Thesis, University of Minnesota, Department of Computer Science, 2000.

[4] P. Pirolli, J. Pitkow, and R. Rao, Silk From a Sow‘s Ear: Extracting Usable Structures from the Web, *Proceeding on Human Factors in Computing Systems (CHI’96)*, ACM Press, pp. 118-125, 1996.

[5] M. Spiliopoulou, and L.C. Faulstich, WUM: A Web Utilization Miner, *Proceeding of EDBT Workshop on the Web and Data Bases (WebDB’98)*, Springer Verlag, pp. 109-115, 1999.

[6] J. Srivastava, R. Cooley, M. Deshpande, and P.N. Tan, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations*, 1(2), pp. 12-23, 2000.

[7] F. Masseglia , P. Poncelet, M. Teisseire, A. Marascu, Web usage mining: extracting unexpected periods from web logs, *Data Min Knowl Disc*, 16, pp.39–65, 2008.

[8] M. Spiliopoulou , L.C. Faulstich , K. Winkler , A data miner analyzing the navigational behavior of web users, *Proceeding of the workshop on machine learning in user modeling of the ACAI’99 international conference Creta*, Greece, 1999.

[9] F. Bonchi , F. Giannotti , C. Gozzi , G. Manco, M. Nanni , D. Pedreschi, C. Renso , S. Ruggieri, Web log data warehousing and mining for intelligent web caching , *Data Knowl Eng*, 39(2), pp. 165–189, 2001.

[10] B. Hay , G. Wets, K. Vanhoof , Mining navigation patterns using a sequence alignment method , *Knowl Inf Syst*, 6(2), pp.150–163, 2004.

[11] Zhu, J., Hong, J., Hughes, J.G. 2002 Using Markov chains for link prediction in adaptive web sites. *Proceeding of soft-ware: first international conference on computing in an imperfect world*, Belfast, UK, pp. 60–73, 2002.

[12] M. Nakagawa, B. Mobasher, Impact of site characteristics on recommendation models based on association rules and sequential patterns. *Proceeding of the IJCAI’03 workshop on intelligent techniques for web personalization*, Mexico, 2003.

[13] R. Srikant, R. Agrawal, Mining sequential patterns: generalizations and performance improvements. *Proceeding of the 5th international conference on extending database technology (EDBT’96)*, pp. 3–17, France, 1996.

[14] A. Mueller, Fast sequential and parallel algorithms for association rules mining: a comparison, Technical report CS-TR-3515, Department of Computer Science, University of Maryland-College Park, 1995.

[15] A. Abraham, V. Ramos, Web Usage Mining Using Artificial Ant Colony Clustering and Genetic Programming, *Congress on Evolutionary Computation (CEC)*, IEEE 2003.

[16] W3C extended log file format. Available at ,http://www.w3.org/TR/WD-logfile

[17] WCA. Web characterization terminology & definitions. Available at ,http://www.w3.org/1999/05/WCA-terms/ .

[18] M. Eirinaki, M.Vazirgiannis, Web Mining for Web Personalization, Athens University of Economics and Business, 2003.

[19] J. Huysmans, B. Baesens , J. Vanthienen , Web Usage Mining: A Practical Study, Katholieke Universiteit Leuven, Dept. of Applied Economic Sciences, 2003.

[20] RFC 1413. Identification Protocol. Available at ,http://www.rfceditor.org/rfc/rfc1413.txt .

[21] L. Catledge, J. Pitkow, Characterizing browsing behaviors on the World Wide Web, *Computer Networks and ISDN Systems*, 27(6), 1995.

[22] J. Deneubourg -L., S. Goss, N. Franks, A. Sendova-Franks, C. Detrain, L. Chrétien, The dynamics of collective sorting: robot-like ants and ant-like robots. *Proceeding of the first international conference on simulation of adaptive behavior*, pp. 356–365, MIT Press, 1991.

[23] J. Handl, B. Meyer, Ant-based and Swarm-based clustering, *Swarm Intelligence*, 1, pp. 95–113, 2007.

[24] E. Lumer, B. Faieta, Diversity and adaptation in populations of clustering ants. *Proceeding of the third international conference on simulation of adaptive behaviour*, pp. 501–508, MIT Press, 1994.

[25] J. Handl, B. Meyer, Improved ant-based clustering and sorting in a document retrieval interface. *Proceeding of the Seventh International Conference on Parallel Problem Solving from Nature*, Vol. 2439 of Lecture Notes in Computer Science, pp. 913-923, Germany: Springer-Verlag, 2002.

[26] H. Kum, J. Pei, W. Wang, D. Duncan, ApproxMAP: approximate mining of consensus sequential patterns, *Proceeding of SIAM International Conference on data mining*, CA, 2003.

ISBN: 978-989-95079-6-8

Multiplicative and implicative importance weighted averaging aggregation operators with accurate andness direction

Henrik Legind Larsen

Department of Electronic Systems
Aalborg University, Denmark
Email: legind@aaue.dk

Abstract— *Importance weighted averaging aggregation operators play a key role in utilizations of electronic data and information resources for retrieving, fusing, and extracting information and knowledge, as needed for decision making. Two central issues in the choice of such operators are the kind of importance weighting and the andness (or conjunction degree) of the operator. We present and discuss two main kinds of importance weighting, namely multiplicative and implicative, each yielding a particular kind of operator for a particular kind of aggregation problems. Importance weighting generalizations of each kind are proposed for each of two classes of averaging operators, namely the Power Means and the Ordered Weighted Averaging operators, each in a De Morgan dual form for increased efficacy. For each class is proposed a function for a rather accurate direct control of the andness. Operators of the same kind appear to behave rather similarly at the same andness, independent of the class of averaging operators generalized.*

Keywords— Aggregation operators, averaging operators, andness, orness, importance weighting, implicative importance.

1 Introduction

With the increasing amount of electronically accessible data in databases, document bases, and data streams, and the need of efficient utilization of such data, aggregation operators have attained new interest. This is due to the central role of such operators in the key reasoning tasks, such as information fusion and pattern recognition, applied in utilizations of many kinds, *e.g.*, information retrieval, information extraction, object recognition, and knowledge discovery. Of particular interest are the aggregations operators between AND (conjunction) and OR (disjunction), *i.e.*, the averaging operators, or generalized conjunction/disjunction functions (GCD) [1], especially the importance weighted generalizations of such operators.

An importance weighted averaging aggregation operator aggregates a number of arguments, each qualified by an importance weight, into a single score. We distinguish between two kinds of importance weighting, namely multiplicative and implicative. Each kind provides an importance weighting generalization of the unweighted averaging operator, in the sense that the latter is retained by the case where all arguments have the same importance weight. We propose such generalizations of two classes of averaging operators, namely the Power Means (PM) and the Ordered Weighted Averaging (OWA) operators [2], each in a De Morgan dual form for increased efficacy, and each with an accurate direct control of the andness, *i.e.*, the degree the operator aggregates like an AND rather than an OR. The generalizations of the PM include the AIWA operators [3].

A few words on the notation applied in the following. $\bigodot_i a_i$, where $\bigodot$ is an operator, is an abbreviation for $\bigodot_{i=1}^{n}(a_i)$; a bold letter like $\mathbf{x}$ denotes the vector $(x_1, \ldots, x_n)$; $\overline{x}$ denotes the standard negation $\overline{x} = 1 - x$; $\overline{\mathbf{x}}$ denotes $(\overline{x_1}, \ldots, \overline{x_n})$. Some specific letters are used all over with the same meaning: I denotes the real unit interval $[0, 1]$; n is the dimension, *i.e.*, number of arguments aggregated by the averaging operator considered; ρ denotes the targeted andness of an averaging operator; $\mathbf{a}$ denotes a vector of arguments $(a_1, \ldots, a_n) \in I^n$; $\mathbf{v}$ denotes an importance weighting vector $(v_1, \ldots, v_n) \in I^n$ that is max-normalized, *i.e.*, $\max_i v_i = 1$; $\mathbf{w}$ denotes an importance weighting vector $(w_1, \ldots, w_n) \in I^n$ that is sum-normalized, *i.e.*, $\sum_i w_i = 1$; $h : I^n \times I^n \to I : (\mathbf{v}, \mathbf{a}) \mapsto h(\mathbf{v}, \mathbf{a})$ is an importance weighted averaging operator, such that v_i is the importance of a_i; for a particular operator h, its name acronym and parameters are attached on the form $h^{\text{acronym}}_{\text{parameters}}$.

Section 2 introduces some basic concepts applied in the following. Sections 3 and 4 introduce, for each class of averaging operators, the PM based and the OWA based, the basis of the class, its multiplicative and implicative importance weighting generalizations, and a function for accurate direct control of the andness of its operators. Sections 5 and 6 discuss, for each kind of importance weighting, the multiplicative and the implicative, the kind of applications they apply to and some key issues of relevance for their application. Section 7 compares and discusses the effect of the two kinds of importance weighting on the common problem of ordering a set of options by their score. Section 8 concludes.

2 Basic concepts

2.1 The andness of an averaging operator

The andness may be viewed as the degree of universal quantification over the arguments, with AND (andness $= 1$) representing *all* and OR (andness $= 0$) representing *at least one*. In this view, the andness is the degree to which all arguments, rather than at least one argument, must support the result of the aggregation. In importance weighted averaging operators, each argument is considered to the degree it is important. The dual interpretation applies to the orness, defined by: orness $= 1 -$ andness.

Since, AND and OR are evaluated by the operators min and max, respectively, the andness of an averaging operator h may be defined as distance between h and the max, relative to the distance between the min and the max, with operators h, max, and min evaluated as the mean of these operators over the ar-

ISBN: 978-989-95079-6-8

gument space, as proposed by Dujmović [4]:

$$\text{andness}(h) = \frac{E(\max_i(\mathbf{x})) - E(h(\mathbf{x}))}{E(\max_i(\mathbf{x})) - E(\min_i(\mathbf{x}))} \qquad (1)$$

where $E(f(\mathbf{x}))$ is the mean of $f(\mathbf{x})$ over the argument space I^n, as defined by $E(f(\mathbf{x})) = \int_{I^n} f(\mathbf{x})d\mathbf{x}$. The terms *andness* and *orness* were coined by Yager for a measure of the OWA operators [2] that is fully consistent with the measure defined by (1).

The andness is, by OWA operators, obtained by positional weights (OWA weights), such that the OWA aggregate is the sum of the products of the i'th OWA weight and the i'th largest argument, whereas it by PM operators is obtained by a power function, yielding emphasis on the smaller or larger valued arguments to the degree the PM represents, respectively, AND and OR (see, *e.g.*, [1]).

2.2 The two kinds of importance weighting

A well-known example of an importance weighted averaging operator is the Weighted Arithmetic Mean (WAM) defined by $h^{\text{WAM}}(\mathbf{v}, \mathbf{a}) = \sum_i (w_i, a_i)$ where, for all i, $w_i = \frac{v_i}{\sum_i v_i}$, for which $\text{andness}(h^{\text{WAM}}) = \text{orness}(h^{\text{WAM}}) = \frac{1}{2}$. The unweighted case, as obtained by $v_i = 1$ for all i, is the Arithmetic Mean (AM): $h^{\text{WAM}}((1, 1, \ldots, 1), \mathbf{a}) = \frac{1}{n}\sum_i a_i = h^{\text{AM}}(\mathbf{a})$. The WAM appears to be the only importance weighted averaging operator that represents both kinds of importance weighting, namely the multiplicative and the implicative. For other degrees and andness than $\frac{1}{2}$, the two kinds of importance weighting provide, as we shall see, different behaviors that apply to different kinds of aggregation problems.

While Multiplicative importance Weighted Averaging (for short, MWA) operators are weighted means, Implicative importance Weighted Averaging (for short, IWA) operators are logic operators for pattern matching inference. In general, the andness of an IWA operator "penalizes" (decreases) the score (the aggregate) to the degree that there are criteria that are important, but not satisfied, while the orness "rewards" (increases) the score to the degree that there are criteria that are important and satisfied.

By the importance weighting, the arguments are, essentially, transformed by the importance weights in a way giving the desired effect, depending on the kind of importance weighting and the andness, in the aggregation. In principle, an argument a with the importance weight w is, by the two kinds of weighting, transformed as follows: by MWA, to the product wa, and by IWA at andness ρ to $\rho(w \Rightarrow a) + \overline{\rho}(\overline{w} \Rightarrow \overline{a})$ where $\Rightarrow$ is a fuzzy implication operator. It is the multiplication in the MWA case and the implication in the IWA case that give name to the two kinds of importance weighting.

As the fuzzy implication $\Rightarrow$ for IWA operators, we choose the Reichenbach implication, $\Rightarrow_{\text{R}}$, as defined by $(v \Rightarrow_{\text{R}} a) = \overline{v\overline{a}} = 1 - v(1-a)$. The reasons for this choice are, first, that it allows us easily to obtain the WAM at $\rho = \frac{1}{2}$, and, second, that the choice of the fuzzy implication appears not to be critical for the behavior of such operators. The latter was tested empirically; the paper doesn't leave space for presenting details from the test.

A key difference between between MWA and IWA operators is that in MWA operators, as opposed to IWA operators, the effect of the importance weights decreases from full effect to no effect as the andness goes from $\frac{1}{2}$ to one of its extremes, 0 or 1. For an appropriate choice of the fuzzy implication operator, the IWA operators represent, like the MWA operators, the WAM at andness $\frac{1}{2}$. These behaviors of the two kind of importance weighted averaging operators are illustrated by Figure 1. We notice, as the figure indicates, that in cases where

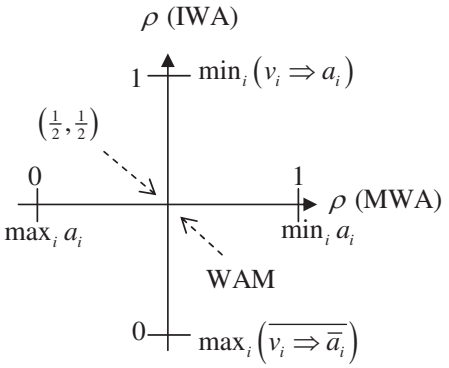


Figure 1: Illustration of MWA and IWA operators

we are completely uncertain about the proper kind of importance weighting and andness for an application, WAM may be a good choice as a starting point.

2.3 The two classes of operators generalized

In the following, we shall in particular consider importance weighted generalizations of two classes of operators, namely the Power Means (PM) and the Ordered Weighted Averaging (OWA) operators, each in a De Morgan dual form for increased efficacy. For each class, we propose two andness-directing functions, $\psi(\rho)$ and $\psi(\rho, n)$, each yielding a value ($\in [1, \infty[$) of the andness controlling parameter of the class. While $\psi(\rho)$ directly yields an andness of the operator that, in general, is somewhat close to the target ρ, $\psi(\rho, n)$ yields the target rather accurately, due to it also being dependent of n, as applied in the definition by (1).

For each class, we provide a common expression of the two kinds of importance weighting generalization, such that the choice of a kind is controlled by a parameter $\gamma \in \{0, 1\}$, yielding the multiplicative generalization by $\gamma = 0$ and the implicative generalization by $\gamma = 1$. The operators are in their common generalized form referred to as, respectively, Andness-directed importance Weighted averaging (AWA) operators and andness-directed importance Weighted OWA operators (W-OWA). Figure 2 gives an overview over the operator name acronyms applied.

Basis	*Without* importance weighting	*With* importance weighting		
		Common class	**M**ultipl. $\gamma = 0$ **MWA**	**I**mpl. $\gamma = 1$ **IWA**
Power Means	AA	AWA	A**M**WA	A**I**WA
OWA	OWA	W-OWA	**M**W-OWA	**I**W-OWA

Figure 2: Operator and operator class name acronyms applied

ISBN: 978-989-95079-6-8

3 AWA operators

3.1 Basis in the Power Means

The AWA (Andness-directed importance Weighted Averaging) operators are based on the Power Means (PM) that (in their unweighted form) are defined by $h^{\text{PM}}_\alpha(\mathbf{a}) = (\frac{1}{n}\sum_i a_i^\alpha)^{\frac{1}{\alpha}}$, $\alpha \in \mathbb{R}$. When α goes from $-\infty$ to $+\infty$, the andness of h^{PM}_α goes from 1 down to 0. It has an asymmetric behavior around andness $= \frac{1}{2}$, as obtained at $\alpha = 1$; in particular, it is "mandatory" for $\alpha \leq 0$, corresponding to andness $\tilde{>}\frac{2}{3}$, in the sense that $h^{\text{PM}}_\alpha(\mathbf{a}) = 0$, if $\min_i(a_i) = 0$, which normally is an undesired property. [1] However, by the De Morgan dual variant PM′, defined by (i) $h^{\text{PM}}_\alpha(\mathbf{a})$, $\alpha \geq 1$, for andness $\leq \frac{1}{2}$, and (ii) $1 - h^{\text{PM}}_\alpha(\overline{\mathbf{a}})$, $\alpha \geq 1$, for andness $\geq \frac{1}{2}$, we obtain a symmetric behavior around andness $= \frac{1}{2}$ and avoid the "mandatory" property. For both (i) and (ii), the arithmetic mean is obtained at $\alpha = 1$; when α goes to $+\infty$, the andness of $h^{\text{PM}'}_\alpha$ goes in case (i) to 0 and in case (ii) to 1.

The AWA operators comprise the multiplicative and implicative importance weighted generalizations of PM′, namely the AMWA (Andness-directed Multiplicative importance Weighted Averaging) operators and the AIWA (Andness-directed Implicative importance Weighted Averaging) operators. The AIWA operators were introduced in [3] where we analyzed their properties and showed that they implement implicative importance weighting by the Riechenbach implication. The AMWA operators is the multiplicative weighting variant of the AIWA operators.

3.2 Definition of AWA operators

The AWA operators are defined by the following semi-recursive expression, where the case of $\rho > \frac{1}{2}$ is defined by its duality to the case of $\rho \leq \frac{1}{2}$:

$$h^{\text{AWA}}_\rho(\mathbf{v},\mathbf{a}) = \begin{cases} \left(\sum_i (y_i a_i^\alpha)\right)^{\frac{1}{\alpha}} & \rho \leq \frac{1}{2} \\ 1 - h^{\text{AWA}}_{\overline{\rho}}(\mathbf{v},\overline{\mathbf{a}}) & \rho > \frac{1}{2} \end{cases} \tag{2}$$

where $\alpha = \psi^{\text{AA}}(\rho) = \overline{\rho}/\rho$, and for all i:

$$y_i = \frac{v_i^{\alpha^\gamma}}{\sum_i v_i^{\alpha^\gamma}} \tag{3}$$

with $\gamma \in \{0,1\}$, $\gamma = 0$ (by which $y_i = \frac{v_i}{\sum_i v_i} = w_i$) yielding the AMWA operators, and $\gamma = 1$ (by which $y_i = \frac{v_i^\alpha}{\sum_i v_i^\alpha}$) yielding the AIWA operators.

By $\mathbf{v} = (1,1,\ldots,1)$ we obtain the common unweighted case of AMWA and AIWA that we shall refer to as Andness-directed Averaging (AA) operators; thus, these operators are, in the case of $\rho \leq \frac{1}{2}$, defined by: $h^{\text{AA}}_\rho(\mathbf{a}) = h^{\text{AWA}}_\rho((1,1,\ldots,1),\mathbf{a}) = (\frac{1}{n}\sum_i a_i^\alpha)^{\frac{1}{\alpha}}$.

3.3 Accurate control of AWA andness

Above, the parameter α is, for a targeted andness ρ, defined by $\alpha = \psi^{\text{AA}}(\rho) = \overline{\rho}/\rho$ in the case $\rho \leq \frac{1}{2}$ (and, by duality, as implemented by the semi-recursive call, $\alpha = \psi^{\text{AA}}(\overline{\rho}) = \rho/\overline{\rho}$ in the case $\rho > \frac{1}{2}$). While by $\psi^{\text{AA}}(\rho)$, andness(h^{AWA}) is equal to ρ, if $\psi^{\text{AA}}(\rho) \in \{0, \frac{1}{2}, 1\}$ and, otherwise, somewhat close to ρ, a much more precise approximation can be obtained by also considering the dimension n, namely through replacing $\psi^{\text{AA}}(\rho)$ by $\psi^{\text{AA}}(\rho, n)$ as defined by:

$$\psi^{\text{AA}}(\rho,n) = (-\log_2 \rho)^{1+(\log_2 n)^{1/3}} \tag{4}$$

An empirical test of the error $\|\rho - \text{andness}(h^{\text{AWA}})\|$, showed that by $\psi^{\text{AA}}(\rho)$ the average error over $\rho \in I$ increases from about 0.03 to about 0.09 as n increases from 2 to 100, while it by $\psi^{\text{AA}}(\rho, n)$ remains around 0.005.

3.4 Decomposability of AWA operators

A property of particular interest for applications of AWA operators is their decomposability, allowing us to update the aggregate if the value of an argument has changed or a new argument has to be considered, without re-computing the whole aggregate. [3] Thus, considering the common case by AWA (2) for $\rho \leq \frac{1}{2}$, let $h^{\text{AWA}}_\rho(\mathbf{v},\mathbf{a}) = c$ to be updated with a new argument a_{n+1} with the importance weight v_{n+1}. Let $\cup$ denote the concatenation of vectors, such that $\mathbf{x} \cup (x_{n+1}) = (x_1,\ldots,x_n) \cup (x_{n+1}) = (x_1,\ldots,x_{n+1})$. Then, the updated aggregate is computed as the AWA aggregate of c and the new argument, as defined by: $h^{\text{AWA}}_\rho(\mathbf{v} \cup (v_{n+1}), (\mathbf{a} \cup (a_{n+1})) = h^{\text{AWA}}_\rho((s, v_{n+1}), (c, a_{n+1})) = \left(\frac{s^{\alpha^\gamma} c^\alpha + v_{n+1}^{\alpha^\gamma} a_{n+1}^\alpha}{s^{\alpha^\gamma} + v_{n+1}^{\alpha^\gamma}}\right)^{1/\alpha}$, with $s = \left(\sum_{i=1}^n v_i^{\alpha^\gamma}\right)^{1/\alpha^\gamma}$.

4 W-OWA operators

4.1 Basis in the OWA operators

The properties of OWA operators [2] are controlled by a vector of sum-normalized position weights $\mathbf{u}$ (called OWA weights) such that the OWA aggregate of an argument vector $\mathbf{a}$ is defined by $h^{\text{OWA}}(\mathbf{a}) = \sum_i (u_i, a_{(i)})$ where $(\cdot)$ is an index permutation such that $a_{(1)} \geq \cdots \geq a_{(n)}$. Two key properties of OWA operators are their andness, defined by andness$(\mathbf{u}) = \frac{1}{n-1}\sum_i ((i-1)u_i)$, and their normalized dispersion (or entropy), defined by $\text{ndisp}(\mathbf{u}) = -\frac{1}{\ln n}\sum_i (u_i \ln u_i)$. This andness measure is consistent with (1).

In [5, 6], Yager proposed an importance weighted generalization of quantifier guided OWA operators that essentially is a multiplicative importance weighting generalization, which we here shall refer to as MW-OWA operators. They are defined by $h^{\text{MW-OWA}}_f(\mathbf{v},\mathbf{a}) = \sum_i (u_i, a_{(i)})$, where $u_i = f(s_i) - f(s_{i-1})$, $i = 1,\ldots,n$, with $s_0 = 0$ and $s_i = \frac{\sum_{k=1}^i v_{(k)}}{\sum_{k=1}^n v_k}$ for $i > 0$, $f : I \to I$ is a regular increasing monotone quantifier (*i.e.*, $f(0) = 0$, $f(1) = 1$, and $f(x) \geq f(y)$ if $x > y$), such as $f(x) = x^\beta$, $\beta \geq 0$, and $(\cdot)$ as before is a permutation of the index such that $a_{(1)} \geq \cdots \geq a_{(n)}$. It is easily seen that when β goes from 0 to $+\infty$, andness$(h^{\text{MW-OWA}})$ goes from 0 to 1, with $h^{\text{MW-OWA}}_f$ representing the max (with andness $= 0$) at $\beta = 0$, the WAM (with andness $= \frac{1}{2}$) at $\beta = 1$, and the min (with andness $= 1$) at $\beta \to +\infty$.

In the case $v_i = 1$, $i = 1,\ldots,n$, we get $s_i = \frac{i}{n}$, and, hence, $\mathbf{u}$ representing the underlying OWA weights (for the OWA operator without importance weighting) that by $f(x) = x^\beta$ are somewhat close to the MEOWA weights [7, 8], *i.e.*, the unique OWA weighting vector with the maximum dispersion at the given andness and dimension n. However, the behavior is not symmetric around andness $= \frac{1}{2}$; in fact, while the dispersion

[1] The mandatory threshold of andness $\tilde{>}\frac{2}{3}$ enforced by the PM is, from a practical point of view, rather arbitrary.

ISBN: 978-989-95079-6-8

of the OWA weights for andness $> \frac{1}{2}$ (*i.e.*, $\beta > 1$) is rather close to the maximum, it is much less for andness $< \frac{1}{2}$ (*i.e.*, $\beta < 1$). For instance, for $n = 5$ the maximum normalized dispersion at andness $= 0.8$ and andness $= 0.2$ is 0.755 in both cases, while the OWA weights with $f(x) = x^\beta$ have the normalized dispersion 0.744 in the first case, but only 0.717 in the second case. Therefore, an improvement with a near optimal dispersion for all degrees of andness, is obtained by defining $h_f^{\text{MW-OWA}}$ as above for andness $\geq \frac{1}{2}$, and by its De Morgan duality for andness $< \frac{1}{2}$, namely (with a change of notation to express the andness ρ) $h_\rho^{\text{MW-OWA}}(\mathbf{v}, \mathbf{a}) = 1 - h_{\overline{\rho}}^{\text{MW-OWA}}(\mathbf{v}, \overline{\mathbf{a}})$. By this improvement, the normalized dispersion in the second case (andness $= 0.2$) above is increased to 0.744, as in the first case, *i.e.*, in both cases only 0.011 lower than the maximum.

4.2 Definition of the W-OWA operators

In the following, we present an extension of the duality based MW-OWA operators to a common class, W-OWA, comprising both the multiplicative importance weighting generalization, MW-OWA, and an implicative importance weighting generalization, IW-OWA, with parametric control of the kind of generalization. The W-OWA operators are defined by:

$$h_\rho^{\text{W-OWA}}(\mathbf{v}, \mathbf{a}) = \begin{cases} \left(\sum_i (u_i b_{(i)})\right) & \rho \geq \frac{1}{2} \\ 1 - h_{\overline{\rho}}^{\text{W-OWA}}(\mathbf{v}, \overline{\mathbf{a}}) & \rho < \frac{1}{2} \end{cases} \tag{5}$$

where, for all i, b_i is defined by:

$$b_i = (v_i^{\gamma(2\rho-1)} \Rightarrow_{\text{R}} a_i) = 1 - v_i^{\gamma(2\rho-1)}(1 - a_i) \tag{6}$$

with $\gamma \in \{0, 1\}$, $\gamma = 0$ (by which $b_i = a_i$) yielding the class of MW-OWA operators, and $\gamma = 1$ (by which $b_i = (v_i^{(2\rho-1)} \Rightarrow_{\text{R}} a_i)$) yielding the class of IW-OWA operators; $u_i = f(s_i) - f(s_{i-1}), i = 1, \ldots, n$, with $s_0 = 0$ and

$$s_i = \frac{\sum_{k=1}^{i} v_{(k)}}{\sum_{k=1}^{n} v_k} \tag{7}$$

for $i > 0$, and $f(x) = x^\beta$, with $\beta = \psi^{\text{W-OWA}}(\rho) = \rho/\overline{\rho}$; finally, $(\cdot)$ is an index permutation such that $b_{(1)} \geq \cdots \geq b_{(n)}$. We notice that the Reichenbach implication applied above may be replaced by any fuzzy implication.

By $\mathbf{v} = (1, 1, \ldots, 1)$ we obtain the common case of MW-OWA and IW-OWA, without importance weighting, namely the duality based, quantifier guided OWA operators that, in the case of $\rho \geq \frac{1}{2}$, are defined by: $h_\rho^{\text{OWA}}(\mathbf{a}) = h_\rho^{\text{W-OWA}}((1, 1, \ldots, 1), \mathbf{a}) = \sum_i (u_i a_{(i)})$ where $u_i = f(\frac{i}{n}) - f(\frac{i-1}{n})$, with $f(x) = x^\beta$.

4.3 Accurate control of W-OWA andness

Above, the parameter β is, for a targeted andness ρ, defined by $\beta = \psi^{\text{W-OWA}}(\rho) = \rho/\overline{\rho}$ in the case $\rho \geq \frac{1}{2}$ (and, by duality, as implemented by the semi-recursive call, $\beta = \psi^{\text{W-OWA}}(\overline{\rho}) = \overline{\rho}/\rho$ in the case $\rho < \frac{1}{2}$). While by $\psi^{\text{W-OWA}}(\rho)$, andness$(h^{\text{W-OWA}})$ is equal to ρ, if $\rho \in \{0, \frac{1}{2}, 1\}$ or $n \to +\infty$, and, otherwise, close to ρ, a more precise approximation can be obtained by also considering the dimension n, namely through replacing $\psi^{\text{W-OWA}}(\rho)$ by $\psi^{\text{W-OWA}}(\rho, n)$ as defined by:

$$\psi^{\text{W-OWA}}(\rho, n) = \frac{0.5 + n\rho}{0.5 + n\overline{\rho}} \tag{8}$$

For instance, by $\rho = 0.8$ (and, through the duality, by $\rho = 0.2$), we get $\beta = \psi^{\text{W-OWA}}(0.8) = \frac{0.8}{1-0.8} = 4$. For $n = 5$, this value of β gives an andness of about of about 0.86, i.e., 0.06 more than the targeted andness, while we by $\beta = \psi^{\text{W-OWA}}(0.8, 5) = \frac{0.5+5\cdot 0.8}{0.5+5(1-0.8)} = 3.0$ obtain the desired andness of 0.8. As n increases from 2 to 100, the average over $\rho \in I$ of the error $\|\rho - \text{andness}(h_\rho^{\text{W-OWA}})\|$ decreases from about 0.07 to about 0.002 by $\psi^{\text{W-OWA}}(\rho)$, and from about 0.0003 to about 0.00001 by $\psi^{\text{W-OWA}}(\rho, n)$.

5 On the multiplicative importance weighting

MWA operators are essentially weighted means that, qua means, are symmetric, monotonic increasing, continuous, and idempotent. They produce means in the interval $[\min_i a_i, \max_i a_i]$, and are monotonically increasing with the orness. If an MWA operator has the andness $= \frac{1}{2}$, it represents the WAM.

5.1 Applications of MWA operators

A common application is for *estimation of a utility variable*. In this case, each argument represents the estimation by an expert (or the measure by some source), while its importance weight represents the decision maker's confidence in the experts's ability to estimate the correct value. The andness and the orness represent decision maker's risk attitude, namely, the degree of, respectively, pessimism and optimism. The aggregate is in this case an estimate of the utility variable, considering the decision maker's risk attitude and, for each expert, the estimate by the expert and the decision masker's confidence in the expert.

Another application is for *selection between (alternative) options*, where the option with the highest score (weighted mean) is winning. In this case, each argument represents the degree to which the option considered has a particular property of interest, while its importance weight represents the importance of having the property to a high degree. The andness represents the degree to which all properties of interest must be present to a high degree. The outcome of the averaging is a ranking of options, possible with a threshold distinguishing acceptable options from unacceptable options.

5.2 Discontinuity property of MWA operators

While MWA and IWA operators both represent the WAM at andness $= \frac{1}{2}$, a key difference is that for MWA, as opposed to IWA, the effect of the importance weights decreases as the andness converges to one of its extremes, 1 or 0. In these cases, MWA evaluates to the smallest argument with a positive weight, yielding a discontinuity, since a small change in an importance weight, from, say, 0.01 to 0, may give a drastic change in the aggregate.

For instance, let h_ρ^{MWA} be an MWA operator (like AMWA or MW-OWA) at andness ρ, and assume $\mathbf{a} = (0.9, 0.1)$. Then $h_1^{\text{MWA}}((1, 0.01), \mathbf{a}) = 0.1$, while $h_1^{\text{MWA}}((1, 0), \mathbf{a}) = 0.9$. Similarly, $h_0^{\text{MWA}}((0.01, 1), \mathbf{a}) = 0.9$, while $h_0^{\text{MWA}}((0, 1), \mathbf{a}) = 0.1$. In both cases, the small decrease from 0.01 to 0 of an importance weight, caused a drastic change (of size 0.8) in the aggregate.

This behavior is an effect of MWA operators as weighted means. It models the attitude of the decision maker, in the cases of extreme pessimism and extreme optimism, to select

ISBN: 978-989-95079-6-8

the most pessimistic, respectively, most optimistic, estimate (argument) by any source with a positive confidence (importance weight). Such behavior is not acceptable in multicriteria aggregation for object recognition; for instance, it is not acceptable that the unimportant second criterion determines the overall satisfaction as by aggregate $h_1^{\text{MWA}}((1, 0.01), \mathbf{a}) = 0.1$. IWA operators avoid this by only considering criteria to the degree they are important; thus $h_1^{\text{IWA}}((1, 0.01), \mathbf{a}) = h_1^{\text{IWA}}((1, 0), \mathbf{a}) = 0.9$.

5.3 Other MWA operators

The PM, on which the AMWA and AIWA operators are based, are in the family of the quasi-arithmetic means (QM) [9, 10]. The weighted QM (WQM) forms the family of MWA operators defined by $h^{\text{WQM}}(\mathbf{v}, \mathbf{a}) = \phi^{-1}(\sum_i (w_i \phi(a_i)))$ where ϕ is a continuous strictly monotonic function, and ϕ^{-1} is the inverse of ϕ. The Weighted PM (WPM), $h^{\text{WPM}}(\mathbf{v}, \mathbf{a}) = (\sum_i (w_i a_i^\alpha))^{1/\alpha}$, $\alpha \in \mathbb{R}$, belong to this family from which they are derived by $\phi_\alpha(x) = x^\alpha$ (with $\phi_\alpha^{-1}(x) = x^{1/\alpha}$).

By $\phi_\alpha(x) = e^{\alpha x}$ (with $\phi_\alpha^{-1}(x) = \frac{\ln x}{\alpha}$), we obtain the Weighted Exponential Means (WEM), $h_\alpha^{\text{WEM}}(\mathbf{v}, \mathbf{a}) = \frac{1}{\alpha} \ln(\sum_i (w_i e^{\alpha a_i}))$, that converges to the WAM for $\alpha \to 0$, and to $\min_i a_i$ and $\max_i a_i$ for α going to, respectively, $-\infty$ and $+\infty$. A nice property of the WEM is its symmetric behavior around andness $= \frac{1}{2}$ as obtained for $\alpha \to 0$. Unlike the WPM, the WEM do not impose the mandatory property.

6 On the implicative importance weighting

6.1 Applications of IWA operators

Implicative importance weighting is applied by IWA operators for pattern matching inference in multicriteria recognition or classification problems, where an observed object (*e.g.*a document or some physical object) is compared to a goal concept (e.g., a query or a class). The goal concept is intensionally characterized by a set of criteria that, in general, are importance weighted. Each criterion expresses a constraint on the values of an object attribute and may, in general, be represented by a fuzzy subset. The importance weight of a criterion expresses the importance of satisfying the criterion in recognizing or classifying an object as an instance of the goal concept.

This scheme applies for querying of the two main kinds, namely object querying and concept querying. [11, 12] In *object querying*, *e.g.*, document retrieval and database querying, the goal concept forms a query. When posed to an object base, the query determines a fuzzy subset of the set of objects, representing the query's fuzzy extension in that set. An object's degree of membership in the extension can be seen as the degree to which the object is an instance of the goal concept and is also referred to as the object's *score* in the goal concept. In *concept querying*, *e.g.*, object recognition and classification, the characterization of the observed object forms a query. When posed to a concept base, *i.e.*, a base of possible goal concepts or classes, the query determines in this case a fuzzy subset of the set of concepts. The answer may in both cases be presented by a ranked list of the objects or concepts (depending on the query type) for which the score is above a given threshold.

6.2 IWA weighting and reasoning scheme

Let C be a goal concept, characterized by the criteria $C_1, \ldots, C_n$ with the importance weights $v_1, \ldots, v_n$, with v_i being the importance of satisfying C_i, and let $a_i = C_i(x) = \mu_{C_i}(x)$ be the degree to which the object x (or, actually, the constrained attribute of x) satisfies C_i. Then, the inference implemented by IWA operators has a Modus Ponens form that in the case of andness = 1 is expressed by:

$$\frac{\begin{array}{l} a_1 = C_1(x), \ldots, a_n = C_n(x) \\ (v_1 \Rightarrow C_1) \wedge \cdots \wedge (v_n \Rightarrow C_n) \to C \end{array}}{C(x) = \wedge_i (v_i \Rightarrow C_i) \ (= h_{\rho=1}^{\text{IWA}}(\mathbf{v}, \mathbf{a}))} \tag{9}$$

Equation 9 expresses that, for an AND aggregation, the goal concept is satisfied by an object to the degree that the requirement *'for all criteria, the criterion is satisfied if it is important'* is met. This may be expressed by $\forall i(\text{important}(\text{C}_\text{i}) \Rightarrow \text{satisfied}(\text{C}_\text{i}))$ that, by $v_i = \text{important}(\text{C}_\text{i})$ and $a_i = \text{satisfied}(\text{C}_\text{i})$, may be written $\forall i(v_i \Rightarrow a_i)$, or, by the standard evaluation of $\forall$ by the min operator, $\min_i(v_i \Rightarrow a_i)$. By other equivalent expressions of $\forall i(v_i \Rightarrow a_i)$, the requirement can be expressed by other words, for instance, by $\neg\exists i(v_i \wedge \neg a_i)$, *'there does not exist a criterion that is important but not satisfied'*.

The transformation of an argument a by an implicative importance weight v for an averaging aggregation at andness ρ is modeled by a function $g_\rho : I^2 \to I : (v, a) \mapsto g_\rho(v, a)$. Thus, for a fuzzy implication $\Rightarrow$, $g_1(v, a) = (v \Rightarrow a)$ is a transformation for an AND aggregation, and $g_0(v, a) = \overline{g_{\overline{0}}(v, \overline{a})} = \overline{v \Rightarrow \overline{a}}$ is its De Morgan dual transformation for an OR aggregation. An important property of g_1 and g_0 is that they for $v = 0$ evaluate to the neutral element for the aggregation, namely, respectively, 1 and 0. In general, for $\rho \in I$, we define g_ρ as the andness-orness weighted sum of g_1 and g_0:

$$g_\rho(v, a) = \rho g_1(v, a) + \overline{\rho} g_0(v, a) \tag{10}$$

which in the case of the Reichenbach implication $\Rightarrow_\text{R}$, with $g_1^\text{R}(v, a) = (v \Rightarrow_\text{R} a) = \overline{v\overline{a}}$ $(= 1 - v(1 - a))$ and $g_0^\text{R}(v, a) = \overline{v \Rightarrow_\text{R} \overline{a}} = va$, evaluates to $g_\rho^\text{R}(v, a) = \rho - v(\rho - a)$. In [13], we presented an (implicative) importance weighting generalization of OWA operators and showed that transformation by g_ρ^R satisfies the requirements for such an importance weighting.

6.3 Other IWA operators

In [3], we introduced another approach to implicative importance weighting generalization of OWA operators, in that case the MEOWA operators [7, 8], namely the IW-MEOWA operators that, by the current notation, are defined by:

$$h_\rho^{\text{IW-MEOWA}}(\mathbf{v}, \mathbf{a}) = \frac{h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, \mathbf{a}) - l}{u - l} \tag{11}$$

where $h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, \mathbf{a}) = \sum_i (u_i b_{(i)})$, $\mathbf{u}$ is the MEOWA weighting vector for the given ρ and n, $b_i = g_\rho^\text{R}(v, a) = \rho - v(\rho - a)$, and $(\cdot)$ is an index permutation, such that $b_{(1)} \geq \cdots \geq b_{(n)}$. Finally, l and u are the lower and upper bounds for $h_\rho^{\text{IW-MEOWA}}(\mathbf{v}, \mathbf{a})$, namely, repectively, $h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, (0, \ldots, 0))$ and $h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, (1, \ldots, 1))$. It was shown that (11) provides a linear transformation of $h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, \mathbf{a})$, such that $h_\rho^{\text{IW-MEOWA}}(\mathbf{v}, \mathbf{a})$ represents the WAM at andness $= \rho = \frac{1}{2}$.

ISBN: 978-989-95079-6-8

7 Comparison of MWA and IWA aggregates

While operators of the same kind, MWA or IWA, tend to behave similarly—in the sense that they provide the same ordering of a given set of options at the same andness and the same importance weights—this is, as we may expect, not the case for operators of difference kinds. This is illustrated by the following examples, where the andness of the MWA and W-OWA classes are targeted by, respectively, $\psi^{\text{AA}}(\rho, n)$ and $\psi^{\text{W-OWA}}(\rho, n)$, as denoted by the aggregation operator symbol; thus, $h^{\text{AMWA}}_{\frac{2}{3},2}$ denotes that h^{AMWA} is applied with $\psi^{\text{AA}}(\frac{2}{3}, 2)$.

Assume that the decision problem is characterized by andness $= \frac{2}{3}$, $n = 2$, $\mathbf{v} = (0.4, 1)$ that is sum-normalized to $\mathbf{w} = (0.286, 0.714)$, and consider the two options represented by the argument vectors $\mathbf{a}_1 = (0.1, 0.7)$ and $\mathbf{a}_2 = (0.9, 0.4)$.

By *MWA aggregation*, the second option, as represented by $\mathbf{a}_2$, get the highest score and is therefore ranked higher than the first option, as the following evaluations, all at andness $= \frac{2}{3}$, show: $h^{\text{AMWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_1) = 0.420 < h^{\text{AMWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_2) = 0.474$ and $h^{\text{MW-OWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_1) = 0.454 < h^{\text{MW-OWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_2) = 0.470$. This is also the case for MWA operators of other classes; thus, for the Weighted Geometric Mean (WGM), $h^{\text{WGM}}(\mathbf{v}, \mathbf{a}) = \prod_i a_i^{w_i}$, that at $n = 2$ has andness $= \frac{2}{3}$, we get $h^{\text{WGM}}(\mathbf{v}, \mathbf{a}_1) = 0.401 < h^{\text{WGM}}(\mathbf{v}, \mathbf{a}_2) = 0.504$; similarly for the Weighted Exponential Means (WEM), $h^{\text{WEM}}_{\alpha}(\mathbf{v}, \mathbf{a}) = \frac{1}{\alpha} \ln \left(\sum_i (w_i e^{\alpha a_i}) \right)$, that at $(\alpha, n) = (-3, 2)$ has andness $= \frac{2}{3}$, we get $h^{\text{WEM}}_{-3}(\mathbf{v}, \mathbf{a}_1) = 0.402 < h^{\text{WEM}}_{-3}(\mathbf{v}, \mathbf{a}_2) = 0.484$.

By *IWA aggregation*, the first option, as represented by $\mathbf{a}_1$, get the highest score and is therefore ranked higher than the second option: $h^{\text{AIWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_1) = 0.579 > h^{\text{AIWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_2) = 0.422$ and $h^{\text{IW-OWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_1) = 0.551 > h^{\text{IW-OWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_2) = 0.473$; similarly, for the IW-MEOWA operator (11), we get $h^{\text{IW-MEOWA}}_{\frac{2}{3}}(\mathbf{v}, \mathbf{a}_1) = 0.530 > h^{\text{IW-MEOWA}}_{\frac{2}{3}}(\mathbf{v}, \mathbf{a}_2) = 0.515$.

This behavior may be explained as follows. By IWA aggregation, the first criterion (argument) is ignored to some degree, due to its rather low importance (0.4); the higher satisfaction of the second, highly important criterion by the first option is the sufficient to give this option the highest rank. By MWA aggregation, the very low value of the first argument in the first option contributes to give this option the lowest rank, despite the rather low importance of the argument.

These observations are supported by the experiment in [3], where the comparison of the IWA operators AIWA and IW-MEOWA for a small data set and a set of implicatively importance weighted queries, indicated a quite similar behavior of the two operators.

8 Conclusion

Importance weighting of two kinds, namely multiplicative and implicative, were proposed as generalizations of two classes of averaging aggregation operators, namely the Power Means (PM) and the Ordered Weighted Averaging (OWA) operators. Each class is applied in a De Morgan dual version, yielding symmetric behavior on both side of andness $= \frac{1}{2}$, and, at the same time, avoiding the mandatory property of the PM and obtaining a near maximum dispersion of OWA at all degrees of andness. The two generalizations of a class are represented by a common expression, where the kind of importance weighting is controlled through a binary parameter.

For each class, were proposed an andness-directing function that allows us to obtain a targeted andness for an operator rather accurately through also considering the number of arguments.

For each kind of importance weighting, we presented and discussed how it works in the operator under different degrees of andness, and the kind of application problems they apply to. Operators of different classes, but with the same kind of importance weighting, appear to have rather similar behavior at the same andness.

For a given choice of the kind of importance weighting, computational issues may affect the choice of operator class. While the PM based operators (AWA) in particular require much computationally rather heavy power functions, OWA based operators (W-OWA) in particular require an ordering (sorting) of the arguments. If reevaluation of a set of options by a modified query occurs frequently, AWA operators have an advantage through their decomposability property.

References

[1] J.J. Dujmović and H.L. Larsen. Generalized conjunction/disjunction. *Int. J. Approx. Reason*, 46(3):423–446, 2007.

[2] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Fuzzy Syst.*, 18(1):183–190, 1988.

[3] H. Legind Larsen. Efficient andness-directed importance weighted averaging operators. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(Suppl.):67–82, 2003.

[4] J.J. Dujmović. Weighted conjunctive and disjunctive means and their application in systems evaluation. *Journal of the University of Belgrade, EE Dept., Series Mathematics and Physics*, (483):147–158, 1974.

[5] R. R. Yager. Multicriteria decision making using fuzzy quantifiers. In *IEEE Conference on Computational Intelligence for Financial Engineering (CIFEr)*, pages 42–47, New York, 1995. Introduces Yager's multiplicative importance weighting of OWA operators.

[6] R.R. Yager. A hierarchical document retrieval language. *Information Retrieval*, 3:357–377, 2000.

[7] D. Filev and R.R. Yager. Analytic properties of Maximum Entropy OWA operators. *Information Sciences*, 85:11–27, 1995.

[8] M. Carbonel, M. Mas, and G. Mayor. On a class of monotonic extended OWA operators. In *Proceeding of FUZZ-IEEE'1997*, pages 1695–1700, Barcelona, 1997.

[9] M. Nagumo. Über eine klasse der mittelwerte. *Japanese Journal of Mathematics*, 7:71–79, 1930.

[10] A.N. Kolmogoroff. Sur la notion de la moyenne. *Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. Sez.*, 12(6):388–391, 1930.

[11] H.L. Larsen and R.R. Yager. The use of fuzzy relational thesauri for classificatory problem solving in information retrieval and expert systems. *IEEE Trans. Syst. , Man, and Cybern.*, SMC-23(1):31–41, 1993.

[12] H.L. Larsen and R.R. Yager. A framework for fuzzy recognition technology. *IEEE Trans. Syst. , Man, and Cybern.*, SMC-30(1):65–76, 2000.

[13] H.L. Larsen. Importance weighted OWA aggregation of multicriteria queries. In *Proceeding of the North American Fuzzy Information Processing Society conference (NAFIPS'99)*, pages 740–744, New York, 10-12 June 1999, 1999.

ISBN: 978-989-95079-6-8

Training a Personal Alert System for Research Information Recommendation

Germán Hurtado Martín [1,2] Chris Cornelis[2] Helga Naessens[1]

1.Dept. of Applied Engineering Sciences, University College Ghent, Ghent, Belgium
2.Dept. of Applied Mathematics and Computer Science, Ghent University, Ghent, Belgium
Email: german.hurtadomartin@{hogent.be,ugent.be}, chris.cornelis@ugent.be, helga.naessens@hogent.be

Abstract— *Information Systems, and in particular Current Research Information Systems (CRISs), are usually quite difficult to query when looking for specific information, due to the huge amounts of data they contain. To solve this problem, we propose to use a personal search agent that uses fuzzy and rough sets to inform the user about newly available information. Additionally, in order to automate the operation of our solution and to provide it with sufficient information, a document classification module is developed and tested. This module also generates fuzzy relations between research domains that are used by the agent during the mapping process.*

Keywords— Automatic classification, Current Research Information Systems, Fuzzy-rough sets, Personal search agents

1 Introduction

Current Research Information Systems (CRISs) are information systems, operating at regional, national, or international level[1], that store and provide access to data on current research carried out by organizations or researchers. These data typically include information on people, projects, organizations, funding possibilities, facilities, etc. CRISs are usually not easy to query, due to several reasons. On the one hand, the data volume is huge. This makes them a very interesting source of information, but can also be overwhelming for the user when he tries to find some specific information in the system. On the other hand, this information is frequently updated, so users are often not aware of new information that could be relevant to them. As a consequence, they are often unable to express their information needs by means of a conventional query. In the last years, personal search agents seem to be the solution to those problems [1]: they gather potentially useful information for users to subsequently present it to them in the form of recommendations. This idea has been received well and can be found not only in research-related applications like CRISs[2] or article recommenders [2, 3], but also in a significant amount of systems from different domains that share the aforementioned problems [4, 5].

However, and despite the fact that these agents can indeed be helpful for the users, they are subject to limitations. Since what they generally do is just to look for exact matches between the user's interests (given as keywords) and the rest of the information in the system, the user will often miss out on useful information as the documents use different terms to refer to the same, or similar, concepts. Furthermore, most CRISs also face other information defects such as missing, ambiguous, or imprecise information [6].

Concepts from fuzzy [7] and rough [8] set theory, upon which the present approach is based, appear as a solution to the problems mentioned above by allowing for a more flexible matching process. Fuzzy sets allow to express partial relationships, which describe reality in a more faithful way than a binary classification, while the rough component provides mechanisms for query expansion: in this way, a user profile and a document may still be matched when they refer to different, but related, keywords, resulting in a higher recall (a higher number of retrieved relevant documents).

To put these ideas into practice, a Personal Alert System (PAS) is currently under development. This system contains profiles of researchers and their interests, activities, papers, etc., as well as documents with information regarding research projects, conference announcements, or funding possibilities. Both these information sources are mapped to a common ontology, which is currently the three-level IWETO[3] taxonomy of the Flemish government. The main goal of the system is to alert users whenever a document can be matched to their research interests by the search agent. A basic prototype has been implemented, using fuzzy-rough algorithms, and in which the user can also influence term relations through a simple feedback process.

While a conceptual version of PAS was described in [9], this work goes further and describes its concrete implementation, paying special attention to the development and evaluation of an automatic classification module which allows the system to classify new documents according to the IWETO taxonomy, as well as to acquire new information with a view to the matching process.

The remainder of this paper is organized as follows: in Section 2, PAS is presented, giving a brief overview of its architecture, and focusing on how the information is represented in the system. In Section 3, the automatic document classification mechanism is described in detail, while in Section 4 we show how the mapper works. Finally, the paper is concluded in Section 5, where we address some issues for future work as well.

2 General design of the system

As said in the introduction, PAS contains information about researchers, projects, funding possibilities, etc., and it will try

[1]Some examples can be found at http://cris.csrees.usda.gov/ (USA), http://www.ris.is/ (Iceland), or http://sicris.izum.si/ (Slovenia). More information can be found at http://www.eurocris.org/, the professional association of CRIS experts and developer of the standard for these systems.

[2]The possibility of using an agent to retrieve funding possibilities is given at EraCareers: http://ec.europa.eu/euraxess/

[3]IWETO, Inventaris Wetenschappelijk en Technologisch Onderzoek Vlaanderen (now FRIS); http://www.iweto.be

ISBN: 978-989-95079-6-8

to match researchers with research information by using the "intelligence" that fuzzy and rough set theories provide. This section first gives a brief, general view on the system, and then focuses on the representation of the information.

2.1 Architecture of the system

Fig. 1 below shows the different modules of the system and how they interact with each other. In brief, the system is organized as follows. Both users (researchers) and documents (information about projects, etc.) are represented by means of profiles. Those profiles, described in more detail in the next subsection, contain several fields, the field storing keywords being the most important one. These keywords represent the researchers' interests (in the cases of the users), or the fields the documents are related to (in the case of the projects), and correspond to the IWETO taxonomy, where they are distributed in three hierarchical levels: a first level with 5 main nodes, a second level with 29 nodes, and a third level containing 359 nodes (in this last level, a node may contain more than one research field, when the fields are closely related). The left-hand side of Fig. 2 shows a sample of the taxonomy. Once a document comes into the system, it is automatically classified (this process will be shown in detail in Section 3), to subsequently be sent to the mapper. The mapper is the central and, thanks to its fuzzy-rough algorithms, intelligent part of the system. It decides whether or not a given document could be interesting for a given user. After this process, and if the mapper has decided so, a notification (about that given document) is sent to the user. Finally, if he desires so, the notified user can give his feedback about his satisfaction degree with the received alert: the system will then use this information to adjust the relations described in Section 2.2.

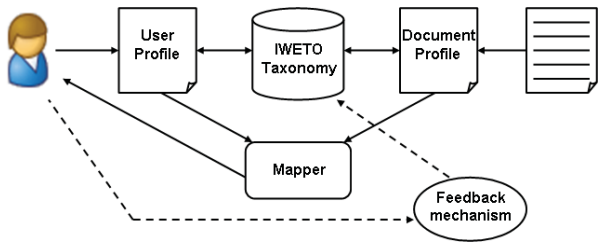


Figure 1: General architecture of PAS

For additional information about the architecture and a more detailed description of the different parts of the system, we refer to [9].

2.2 Representation of the information

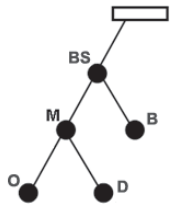
A key issue in PAS is the representation of the information: not only how profiles are represented but also how the IWETO taxonomy is stored.

Profiles are based on IWETO keywords, and their representation is common for both users and documents. Leaving aside the fields inherent to users or documents (i.e. "name" for users, "title" or "description" for documents), both user and document profiles contain a field dedicated to store the keywords that will represent them in the system. The structure used for that is a set where the pairs $<$*interest terms*, *degree of interest*$>$ are stored (*degree of relationship*, in the case of the documents). For example, the list $\{$*(AI,1)*, *(T,0.7)*, *(P,0.2)*$\}$ would correspond to a user very interested in *Artificial Intelligence*, quite interested in *Translation*, and slightly interested in *Physics*. The degree of interest/relation lies always between 0 (no interest/relation) and 1 (strong interest/relation). Every profile can therefore be seen as a fuzzy set in the whole collection of keywords X (since the keywords can come from any of the levels of the taxonomy, its hierarchical structure does not play any role at this point).

In the current implementation, the information contained in the profile keywords has different origins. While documents are automatically classified by using the techniques explained in Section 3, users must select their interests from a list of keywords. Users also select the different interest degrees from a list where these are linguistically represented (to subsequently map them onto a numerical value automatically).

Apart from user-keyword and document-keyword relations, we also consider relationships between keywords. A square matrix is used to represent the different relations between keywords, expressed as a degree ranging between 0 (no relation) and 1 (strong relationship). The right-hand side of Fig. 2 gives an example of this representation. A way of assigning degrees to term pairs is based on their co-occurrence during the training process; the more two terms co-occur in training documents, the higher their degree of relationship (see (10) in Section 4 for more detail).

These degrees make it possible to define a fuzzy relation R reflecting how closely two keywords k_1 and k_2 are related. It is important to remark that R is not a fuzzy tolerance relation, since it is not symmetric. It is reflexive, because a term is perfectly related to itself, so $R(k_1, k_1) = 1$, but $R(k_1, k_2)$ and $R(k_2, k_1)$ are not always equal. This is necessary to depict reality in a more faithful way. For instance, and referring to the example shown in Fig. 2, it is logical to think that there is a stronger relation from *Ophtalmology* towards *Medicine* than vice versa: a document related to *Ophtalmology* will always be related to *Medicine*, but a document related to *Medicine* will not be necessarily relevant to *Ophtalmology*.



	BS	M	B	O	D
BS	1	0.4	0.2	0.1	0.1
M	1	1	0.3	0.3	0.2
B	1	0.3	1	0.1	0
O	1	1	0.1	1	0.1
D	1	1	0	0.1	1

O: Ophtalmology *M: Medicine* *BS: Biomedical Sciences*
D: Dermatology *B: Biophysics*

Figure 2: Example of how the classification is stored

The information about the links between keywords (research fields) contained in the keyword relationship matrix can be used by the system, along with the interest research fields in the profiles, to compute a fuzzy-rough upper approximation for every profile. This process will be explained in more detail in Section 4.

3 Automatic document classification

Clearly, information is the core of any information system, CRISs included. Therefore, it is important for the system to

ISBN: 978-989-95079-6-8

be able to acquire new information easily. Since PAS aims to be as human-independent as possible, it is equipped with an automatic document classification module for this purpose. A representation of the process that documents undergo in this module is shown in Fig. 3 below. To build the module, some of the ideas in [10], such as the usage of keyword vectors for documents and classification, have been used.

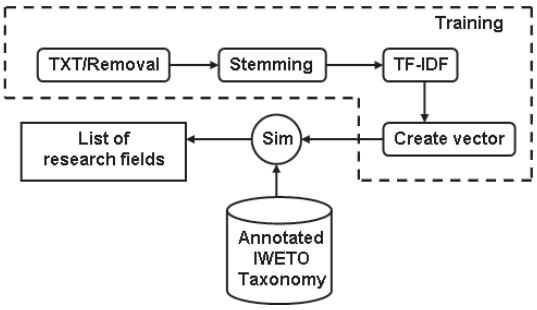


Figure 3: Automatic classification process

3.1 Training

First of all, the system needs to be trained, which is done by means of already classified documents. Currently, these documents come from the FRIS database[4]. Basically, each of them consists of an abstract about a given project, the keywords of that project, and the research fields under which the project was classified in the FRIS database. The documents must first be converted to a format readable by the system, namely txt. Afterwards, every document d is processed by a stemming algorithm. To this purpose, PAS currently uses a purpose-built stemmer, although the list of stopwords that this stemmer uses is an extended version of that which can be found at the site of the implementation of the Porter stemmer [11] in the Snowball language for stemming algorithms[5]. After that, the TF-IDF algorithm [12] is applied to calculate a weight w_i for each term in the document. Those terms with a weight w_i greater than a given threshold γ ($\gamma \geq 0$) are put into a vector $\vec{d}$ that represents the document. This vector $\vec{d}$ is further normalized so that it has unit length.

An important aspect at this point are the terms for which a TF-IDF value is computed. The first implementations of the classifier worked with two-word terms (bigrams), since such terms offer more information and are less ambiguous; for example "information system" is more meaningful than "information" and "system". However, they are also more difficult to match between documents, and some combinations are not so fortunate. So since early tests with one-word terms (unigrams) showed a sensible improvement, the classifier currently works with unigrams. Nevertheless, for some terms composed of more than 1 word, no splitting is applied. This happens for terms extracted via pattern recognition, for example those preceded by the string *"Keywords:"* in a document. In this cases, the terms are added, untouched, to the vector $\vec{d}$, with weight 1.0. Additionally, if the term is not a unigram, another copy of it is added to the rest of the text in the document to be split and processed as described above.

When the training documents have been reduced to their term vector form, the system is ready to start its learning process. It is here where the hierarchy of the IWETO thesaurus is used. First, and based on the information regarding the nodes (research fields) under which the documents have been classified, every vector $\vec{d}$ is assigned to its corresponding node or nodes n. A vector $\vec{n}$ for every node n is then computed. These vectors $\vec{n}$ are normalized vectors containing the terms that are most relevant to node n, or in other words, the terms (and respective weights) contained in the term vectors $\vec{d}$ assigned to n. More formally, let $S(n)$ be the set of subconcepts under concept n (children of node n). Also, let $\{d_1^n, d_2^n, \cdots, d_{k_n}^n\}$ be the individual training documents classified under concept n. $Docs(n)$, the set of all the documents classified under concept n augmented with the documents classified under all its children is defined as:

$$Docs(n) = \left[\bigcup_{n' \in S(n)} Docs(n') \right] \cup \{d_1^n, d_2^n, \cdots, d_{k_n}^n\} \quad (1)$$

The term vector $\vec{n}$ is then computed as:

$$\vec{n} = \left[\sum_{d \in Docs(n)} \vec{d} \right] / |Docs(n)| \quad (2)$$

This term vector $\vec{n}$ is finally normalized into a unit vector.

3.2 Automatic classification

The classification process starts in a similar way to that of the training process, including the application of the stemming and TF-IDF algorithms to create their representative vectors $\vec{d}$. Each test document is then compared to every node in the hierarchical classification. This is done by comparing their representative term vectors $\vec{d}$ and $\vec{n}$ by means of cosine similarity. Those nodes n for which the value of the comparison $sim(\vec{d}, \vec{n})$ is greater than zero are added to a priority queue along with their sim values, ordered with respect to these values. The cosine similarity measure sim for normalized vectors is defined as their inner product, i.e.,

$$sim(\vec{d}, \vec{n}) = \frac{\vec{d} \cdot \vec{n}}{|\vec{d}||\vec{n}|} = \vec{d} \cdot \vec{n} \quad (3)$$

Since the values $sim(\vec{d}, \vec{n})$ reflect how related the document d is to the class represented by node n, they are normalized so that d and n_1, the class with which d has the strongest relation, are related in a degree 1. Finally, the nodes in the N first places of the priority queue (i.e. the nodes with the N greatest values for sim, $N \geq 0$) are retrieved, along with their computed and normalized sim values, as long as they are greater or equal than a given threshold δ, $\delta \geq 0$. Thus, δ controls that only those classes to which d is sufficiently strongly related are taken into account.

The document is then classified under those research fields. This means that a profile is created for the document, with its representative keywords the research fields in which it has been classified, and with the respective sim values giving an idea of how related the document is to them.

[4]FRIS, Flanders Research Information Space; http://www.ewi-vlaanderen.be/fris/

[5]M. Porter, *Snowball: a language for stemming algorithms*; http://snowball.tartarus.org/index.php

ISBN: 978-989-95079-6-8

3.3 Classification results

The last part of this section is dedicated to show some results obtained with the proposed classifier. At the same time, some design decisions are discussed.

First of all, an evaluation measure is needed. Evaluating hierarchical classification is no trivial problem, a problem that gets yet more complicated by the fact of the classification being multi-label, i.e., documents can be labelled with different keywords. In this work we will use an adaptation to fuzzy sets of the measure hF_β proposed by Kiritchenko et al. in [13]. Specifically:

$$hF_\beta = \frac{(\beta^2+1)\cdot hP \cdot hR}{(\beta^2 \cdot hP + hR)}, \beta \in [0, +\infty) \quad (4)$$

where hP and hR are hierarchical precision and hierarchical recall, respectively. These measures are a variation of the traditional precision and recall evaluation measures which also take into account the hierarchical structure of the classes. In particular, rather than comparing the classes a document belongs to and the ones that were predicted for it, we also consider the ancestors of these classes. Formally:

$$\hat{C}_i = \{\cup_{c_k \in C_i} Anc(c_k)\} \qquad \hat{C}'_i = \{\cup_{c_l \in C'_i} Anc(c_l)\} \quad (5)$$

where C_i and C'_i are the actual and the predicted class sets, respectively, and $Anc(n)$ is the set of ancestors of n, n included. Therefore, $\hat{C}_i$ and $\hat{C}'_i$ are the actual and predicted class sets extended by adding them the ancestors of the classes they contain. Note that $\hat{C}'_i$ is treated as a fuzzy set, with its membership values equal to the predicted sim values, and that $\hat{C}_i$ is actually a crisp set, since a document d is related to all the classes under which it is actually classified in a degree of 1. Measures hP and hR are then defined as:

$$hP = \frac{\sum_i |\hat{C}_i \cap \hat{C}'_i|}{\sum_i |\hat{C}'_i|} \qquad hR = \frac{\sum_i |\hat{C}_i \cap \hat{C}'_i|}{\sum_i |\hat{C}_i|} \quad (6)$$

Since we are dealing with fuzzy sets, corresponding operations are used, defining the intersection of $\hat{C}_i$ and $\hat{C}'_i$ by:

$$(\hat{C}_i \cap \hat{C}'_i)(x) = \min(\hat{C}_i(x), \hat{C}'_i(x)) \quad (7)$$

for a keyword x, and defining the cardinality of $\hat{C}'_i$ as:

$$|\hat{C}'_i| = \sum_{x \in X} \hat{C}'_i(x) \quad (8)$$

By this definition, wrongly predicted classes are penalized less strongly when they belong to the same branch as one of the actual classes. The value of β can be chosen; in our case we will use $\beta = 1$, so both hierarchical precision and hierarchical recall have the same weight in hF.

All the results shown throughout the rest of the section have been calculated with this evaluation measure. On the other hand, the tests to obtain those results were carried out with a dataset formed by 9438 IWETO project descriptions, on which 10-fold cross validation was performed.

The results are subject to a number of parameter settings. The parameters with more impact in the results are the ones related with the term vector $\vec{d}$ used to represent a given document d, and the way the system chooses the classes under which the document d will be classified. Those problems are related with the parameter γ mentioned in Section 3.1, and the parameters N and δ mentioned in Section 3.2, respectively. All these parameters are discussed here and some results about their tuning are presented in Table 1.

Parameter γ is necessary to restrict the number of terms in every term vector $\vec{d}$, mainly because of memory usage and system performance reasons, since terms with very low TF-IDF weights would be probably irrelevant anyway if they were included in the term vector. Due to the short length of the documents we used, parameter γ is not very essential for our purposes. Therefore, the value used for our tests was $\gamma = 0$. Nevertheless, as mentioned above, its use is recommended when dealing with longer documents (or documents with an unknown length).

More relevant in our tests was the number of classes in which the system classifies the document: too many or too few classes probably signify a big difference with respect to the actual classification (resulting in too low precision and/or recall), and will also have an important impact on the user's satisfaction (if a document is classified in too many categories, it will be considered for the matching process even when it is not strongly related to a given category; if it is classified in just one category, it probably will not reach its whole target group of users). Parameters N and δ control this.

As said in Section 3.2, N sets the number of research fields in which a document will be classified, i.e., the number of classes that will be retrieved from the first places of the priority queue. On the other hand, a threshold δ is also necessary, in order to avoid cases of weak relationships: N is the number of classes with the highest value for the similarity measure sim, but that does not necessarily mean that all N values are high. Threshold δ tries to solve this. Since it is applied to normalized results, δ must be a value between 0 and 1, and since strong relationships are desired, the value cannot be too low. The results of the different tests are shown in Table 1.

Table 1: Results for tests with different values for N and δ.

δ	N	6	5	4	3	2	1
0.3	hF_1	0.502	0.512	0.523	**0.53**	0.527	0.478
	hP	0.419	0.442	0.472	0.508	0.555	0.623
	hR	0.625	0.609	0.587	0.555	0.501	0.388
0.4	hF_1	0.505	0.514	0.524	**0.53**	0.526	0.478
	hP	0.426	0.447	0.475	0.509	0.556	0.623
	hR	0.62	0.605	0.584	0.553	0.5	0.388
0.5	hF_1	0.509	0.517	0.525	**0.53**	0.526	0.478
	hP	0.439	0.457	0.481	0.513	0.557	0.623
	hR	0.607	0.595	0.577	0.548	0.497	0.388
0.6	hF_1	0.514	0.519	0.525	0.528	0.524	0.478
	hP	0.458	0.473	0.493	0.52	0.56	0.623
	hR	0.584	0.575	0.562	0.538	0.493	0.388
0.7	hF_1	0.516	0.519	0.522	0.524	0.52	0.478
	hP	0.487	0.496	0.511	0.531	0.564	0.623
	hR	0.549	0.546	0.535	0.518	0.483	0.388
0.8	hF_1	0.514	0.514	0.515	0.516	0.514	0.478
	hP	0.528	0.531	0.537	0.55	0.574	0.623
	hR	0.5	0.499	0.495	0.486	0.465	0.388
0.9	hF_1	0.503	0.503	0.503	0.503	0.501	0.478
	hP	0.577	0.577	0.578	0.581	0.591	0.623
	hR	0.445	0.445	0.444	0.443	0.435	0.388

The average number of classes to which a given document from the dataset belongs is 3. Due to this fact, $N = 3$ offers the best values for the evaluation measure hF_1. It can also be

ISBN: 978-989-95079-6-8

seen that $N = 3$ offers the best equilibrium between hP and hR. Moreover, note that the precision hP is inversely proportional to N: the fewer classes d is classified in, the higher the precision. This is logical since a low value for N means that the document is classified in those classes which offered the highest values for the similarity measure sim, classes that are then probably among those under which the document was actually classified. But that small number of classes has a drawback: the recall is of course lower. In the same way, a higher value of N results in higher values of hR: the document is classified under a lot of concepts and therefore it is more probable that they cover all the classes actually linked with the document. Of course, this higher number of classes makes the classifier less precise.

Also δ is important, mainly for higher values of N: these high values imply taking more classes into account when classifying a document, but δ imposes a threshold that leaves out all those "highest values" that are not actually high enough to be considered for the classification. In that way, δ allows higher values of N to be used in order to avoid leaving out potentially adequate classes, but without the risk of considering classes that probably are not so related to the document.

Note that the combination of $N = 1$ and δ is useless, as reflected by the results: since the results of sim are normalized, the highest value in the priority queue (and therefore the only one that will be retrieved when $N = 1$) will be equal to 1.

Although a high precision is desirable, it is necessary to keep in mind that the system aims to alert the researchers about potentially interesting documents. Therefore, the documents are preferably classified under more than one research field (as long as those fields are relatively strongly related to the document). That means that a balance between precision and recall ($\beta = 1$) is recommended in this case. As the combination $\delta = 0.5$, $N = 3$ offers the best results ($\delta = 0.4$ and $\delta = 0.3$ obtain similar values for hF but a lower precision), they are the values currently used by the system.

There are nonetheless some remarks to be made. In this case, the hierarchical classification problem gets complicated by the fact that the documents in the dataset have been manually classified by different people. This means that in some cases the document has been indexed under all the possible research fields while in other cases it has been vaguely classified under a parent concept. As previously explained, this can cause big differences between the set containing the research fields under which the documents were actually classified and that predicted by the system.

4 Mapper

Since the mapper is the most important part of the system, this section is used to explain in detail how it works. The mapper determines whether a document is interesting enough to notify a user about it. In this process, as discussed in the introduction, the system should be able to identify interesting documents even when their keywords do not exactly match those in the user's profile, but are semantically related to them. To achieve this added intelligence, some ideas from fuzzy-rough set theory are used.

Specifically, fuzzy-rough query expansion is applied. In particular, the system currently uses an adaptation of the approach described in [14]: to assess how well a document profile D matches a user profile U (both represented as fuzzy sets in the set X of keywords) the algorithm first uses a fuzzy relation R to generate their respective upper approximations $R{\uparrow}D$ and $R{\uparrow}U$, where the upper approximation of a fuzzy set A in X under a fuzzy relation R is defined as:

$$(R{\uparrow}A)(y) = \sup_{x \in X} \min(R(x, y), A(x)), \forall y \in X \qquad (9)$$

A keyword belongs to this upper approximation to the extent that it is related, by means of the fuzzy relation R, to at least one of the keywords in A. In other words, (9) defines the set of objects possibly belonging to A to a certain degree. In the current system, the relatedness of two terms is based on their co-occurrence during the training process, as explained in Section 2.2. To be precise, the relation R between two terms x and y can be defined as:

$$R(x, y) = \frac{|Docs(x) \cap Docs(y)|}{|Docs(x)|} \qquad (10)$$

The similarity between D and U is then computed, using α-cuts[6], as

$$Sim_\alpha(U, D) = 1 - \frac{|B_{u\alpha}|}{|(R{\uparrow}D)_\alpha|} \qquad (11)$$

where

$$B_u = R{\uparrow}D \ominus [R{\uparrow}U \cap R{\uparrow}D] \qquad (12)$$

with $\ominus$ the difference of A and B, defined by $(A \ominus B)(x) = \max(0, A(x) - B(x))$.

If $|(R{\uparrow}U)_\alpha| = 0$, the similarity is defined to be 0. Currently, the system uses a fixed value $\alpha = 0.5$, which turned out to yield the best performance experimentally in preliminary tests. It is important to point out that Sim is asymmetric and that therefore the order of U and D is relevant. This keeps the focus of the comparison on the document (B_u is the set of terms in the upper approximation of D that are not shared with the upper approximation of U) and allows the document to reach more users, as long as it is of relevance for them.

The obtained similarity between user and document is then compared to a user-set notification threshold (entered linguistically and then mapped onto a numerical value): if the similarity is greater than or equal to the threshold, user U will be notified by the system about document D. In this way, the user can tune the sensitivity of the mapper: if he only wants to receive alerts about documents that are definitely relevant, he can impose a higher threshold.

5 Conclusion and future work

We have shown how fuzzy relations are used to represent gradual relationships between research fields, and how these relations can be generated from a given dataset. It has also been shown how to construct user and document profiles as fuzzy sets, and how these sets can be matched by applying ideas from fuzzy-rough set theory. In addition, we have also proposed and explained a method to automatically classify documents, in order to keep the system easily updated. However, and despite the promising performance offered by the current

[6] The α-cut of A is defined as $A_\alpha = \{x \in X | A(x) \geq \alpha\}$, where $\alpha \in [0, 1]$

ISBN: 978-989-95079-6-8

implementation of the system, there is still a lot of work to be done.

Since the main goal of the project is to match researchers with documents that can be potentially useful for them, the main effort will be put into the mapping algorithms. New algorithms will be developed and tested to replace the one currently used. This current algorithm, though very interesting, is quite basic, so we think that there is a lot of room for improvement. For example, different similarity measures to compare the upper approximation can be investigated, and we plan to set up an experiment with a group of real test users.

The performance of the fuzzy-rough algorithms is of course also linked to the availability of sufficient document profiles and the quality of the classifier. However, the current version of this module of the system is quite simple, and further improvements could be very interesting and fruitful. For instance, more complex techniques from language technology could be used to refine the keyword extraction. Also, the fuzzy weights of the relationships between concepts can play a more important role, in the classification process as well as in the evaluation of results. The possibility of semi-automatically enriching the IWETO thesaurus with a fourth level, for the sake of precision, will be studied too. Besides, an alternative representation of the information, with the profiles as copies of the classification (which would allow the user to assign his own weights to the relations), will be considered.

Other modules will be added as well. In order to achieve a greater autonomy of the system, the feedback mechanism will be automated, and different techniques will be used to monitor the user's behavior while using the system, so that it can automatically update his list of interests and preferences, adding information that can be useful for the mapper. Also, the classifier could be used to extract additional information from the researcher's publications, when available. All this information can also be used, along with that provided by the feedback mechanism, to refine and adjust the relationships between concepts from the thesaurus. The user will also be free to explore the system on his own: a browsing utility and a search engine will be added with that finality. This engine can also use the same query expansion techniques that are used for personalized notification.

Acknowledgment

The authors thank the Flanders Research Information Space program for its cooperation. Also, Chris Cornelis thanks the Research Foundation-Flanders for supporting his research.

References

[1] P. Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7):31–40, 1994.

[2] T. Bogers and A. van den Bosch. Recommending scientific articles using citeulike. In *Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys'08)*, pages 287–290, 2008.

[3] S.M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S.K. Lam, A.M. Rashid, J.A. Konstan, and J. Riedl. On the recommending of citations for research papers. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work (CSCW'02)*, pages 116–125, 2002.

[4] M. Montaner, B. López, and J.L. de la Rosa. A taxonomy of recommender agents on the internet. *Artificial Intelligence Review*, 19(4):285–330, 2003.

[5] G. Adomavicius and A. Tuzhilin. Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[6] A. Motro and P. Smets. *Uncertainty management in information systems: from needs to solutions*. Springer, New York, NY, 1997.

[7] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[8] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11(5):341–356, 1982.

[9] G. Hurtado Martín, C. Cornelis, and H. Naessens. Personalizing information retrieval in CRISs with fuzzy sets and rough sets. In *Proceedings of the 9th International Conference on Current Research Information Systems (CRIS2008)*, pages 51–59, 2008.

[10] A. Sieg, B. Mobasher, and R. Burke. Learning ontology-based user profiles: a semantic approach to personalized web search. *IEEE Intelligent Informatics Bulletin*, 8(1), 2007.

[11] M. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[12] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, NY, 1983.

[13] S. Kiritchenko, S. Matwin, R. Nock, and A. Fazel Famili. Learning and evaluation in the presence of class hierarchies: application to text categorization. In *Canadian Conference on AI*, pages 395–406, 2006.

[14] P. Srinivasan, M.E. Ruiz, D.H. Kraft, and J. Chen. Vocabulary mining for information retrieval: rough sets and fuzzy sets. *Information Processing and Management*, 37(1):15–38, 2001.

ISBN: 978-989-95079-6-8

On Reduced Semantics for Fuzzy Predicate Logics

Pilar Dellunde[1,2]

1.Universitat Autònoma de Barcelona
2.Artificial Intelligence Research Institute IIIA - CSIC
08193 Bellaterra, Catalonia
Email: pilar.dellunde@uab.cat

Abstract— Our work is a contribution to the model-theoretic study of equality-free fuzzy predicate logics. We present a *reduced semantics* and we prove a completeness theorem of the logics with respect to this semantics. The main concepts being studied are the *Leibniz congruence* and the *relative relation*. On the one hand, the Leibniz congruence of a model identifies the elements that are indistinguishable using equality-free atomic formulas and parameters from the model, a reduced structure is the quotient of a model modulo this congruence. On the other hand, the relative relation between two structures plays the same role that the isomorphism relation plays in classical predicate languages with equality.

Keywords— Equality-free Language, Fuzzy Predicate Logic, Model Theory, Reduced Structure, Relative Relation.

1 Introduction

This work is a contribution to the model-theoretic study of equality-free fuzzy predicate logics. Model theory is the branch of mathematical logic that studies the construction and classification of structures. Construction means building structures or families of structures, which have some feature that interest us. In our case we devote our investigation to the class of reduced structures, that help us to shed light to the characteristic role played by equality in predicate fuzzy logics.

Classifying a class of structures means grouping the structures into subclasses in a useful way, and then proving that every structure in the collection does belong in just one of the subclasses. The most basic classification in classical model theory is given by the relations of elementary equivalence and isomorphism. Our purpose in the present article is to investigate and characterize the structure-preserving maps between structures in a fuzzy setting. In classical predicate logics with equality, homomorphisms are structure-preserving, but if they are not isomorphisms, they don't necessarily preserve all the formulas of the language. On the contrary, in equality-free fuzzy predicate logics, σ-homomorphisms preserve all the formulas, but unlike isomorphisms, the relation between structures of one being the σ-homomorphic image of another is not an equivalence relation in the class of structures.

The main concepts studied in this work are the *Leibniz congruence* and the *relative relation*. The notion of Leibniz congruence arises in a very natural way. It is said that two elements of a model are related by this congruence when they satisfy exactly the same equality-free atomic formulas with parameters in the model. This congruence always exists, and it turns out to be the greatest congruence of the model. This idea has its origin in the Principle of the identity of the indiscernibles of G. W. Leibniz.

Given a model $(\mathbf{M}, \mathbf{B})$ the quotient structure modulo the Leibniz congruence $\Omega(\mathbf{M}, \mathbf{B})$ is denoted by $(\mathbf{M}, \mathbf{B})^r$ and is called its *reduction*. When we make the quotient modulo the Leibniz congruence, we identify the elements that are indistinguishable using equality-free atomic formulas and parameters in the model, thus the Leibniz congruence of the reduction of a model is always the identity. An structure with the property that its Leibniz congruence is the identity is said to be a *reduced* structure. The importance of reduced structures in equality-free logic comes from the fact that the reduction of a model is a σ- homomorphic image of the model and therefore, the model and its reduction satisfy exactly the same equality-free sentences.

The other main concept analysed is the relative relation. It is said that two structures are *relatives* when they have isomorphic reductions. Along this work we will give different characterizations of this relation. Our aim is to point out that it plays the same role in equality-free logic that the isomorphism relation plays in logic with equality. The actual interest of the Leibniz congruence and the relative relation comes from the work of W. Blok and D. Pigozzi. They introduced the concept of relative relation for the special case of logical matrices in [1], and in [2] they made an extensive use of what they named the Leibniz congruence.

Different definitions have been introduced so far for basic model-theoretic operations on structures. For instance, the notion of *elementary submodel, morphism* and *congruence of a fuzzy model* of [3], *elementary embeddings and submodels* of [4], *fuzzy submodel, elementary fuzzy submodel* and *isomorphism* of structures of first-order fuzzy logic with graded syntax of [5], *complete morphism and congruence* in languages with a similarity predicate of [6] and the notion of σ-embedding of [7]. Being our starting point all these works, in the Preliminaries section we introduce the notions of homomorphism and congruence of a model, trying both to encompass the most commonly used definitions in the literature and to extend the corresponding notions of classical predicate logics.

In section 3 we introduce the notion of reduced structure and some basic model-theoretic properties of this kind of structures. In section 4 we characterize when two structures are relative and we prove that the relative relation is the transitivization of the relation of being a σ-homomorphic image. Finally, section 5 is devoted to future work.

ISBN: 978-989-95079-6-8

2 Preliminaries

Our study of the model theory of fuzzy predicate logics is focused on the basic fuzzy predicate logic MTL∀ and stronger t-norm based predicate calculi, the so-called *core fuzzy logics*. We start by introducing the notion of core fuzzy logic in the propositional case.

Definition 1 *A propositional logic L is a* core fuzzy logic *iff L satisfies:*

1. *For all formulas ϕ, φ, α, $\varphi \equiv \phi \vdash \alpha(\varphi) \equiv \alpha(\phi)$.*
2. (LDT) Local Deduction Theorem: *for each theory and formulas ϕ, φ:*

   $$T, \varphi \vdash \phi \text{ iff for some natural number } n,\ T \vdash \varphi^n \to \phi.$$

3. *L expands* MTL.

For a thorough treatment of core fuzzy logics we refer to [4], [8] and [7]. A *predicate language* Γ is a triple (P,F,A) where P is a non-empty set of predicate symbols, F is a set of function symbols and A is a function assigning to each predicate and function symbol a natural number called the *arity of the symbol*. Functions f for which $\mathrm{A}(f) = 0$ are called *object constants*. Formulas of the predicate language Γ are built up from the symbols in (P,F,A) together with logical symbols $(\forall, \exists, \&, \to, \overline{0}, \overline{1})$, variables and punctuation. Throughout the paper we consider the equality symbol as a binary predicate symbol not as a logical symbol, we work in equality-free fuzzy predicate logics. That is, the equality symbol is not necessarily present in all the languages and its interpretation is not fixed.

Let L be a fixed propositional core fuzzy logic and $\mathbf{B}$ an L-algebra, we introduce now the semantics for the fuzzy predicate logic $L\forall$. A $\mathbf{B}$-structure for predicate language Γ is a tuple $\mathbf{M} = (M, (P_{\mathbf{M}})_{P\in\Gamma}, (F_{\mathbf{M}})_{F\in\Gamma}, (c_{\mathbf{M}})_{c\in\Gamma})$ where M is a non-empty set and

1. For each n-ary predicate $P \in \Gamma$, $P_{\mathbf{M}}$ is a $\mathbf{B}$-fuzzy relation $P_{\mathbf{M}} : M^n \to \mathbf{B}$.
2. For each n-ary function symbol $F \in \Gamma$, $F_{\mathbf{M}} : M^n \to M$.
3. For each constant symbol $c \in \Gamma$, $c_{\mathbf{M}} \in M$.

Let $\mathbf{M}$ be a $\mathbf{B}$-structure, an $\mathbf{M}$-*evaluation* of the variables is a mapping v which assigns to each variable an element from M. By $\phi(x_1, \ldots, x_k)$ we mean that all the free variables of ϕ are among $x_1, \ldots, x_k$. If v is an evaluation such that for each $0 < i \leq n$, $v(x_i) = d_i$, and λ is either a Γ-term or a Γ-formula, we abbreviate by $\|\lambda(d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}$ the expression $\|\lambda(x_1, \ldots, x_n)\|^{\mathbf{B}}_{\mathbf{M},v}$. Let ϕ be a Γ-sentence, given a $\mathbf{B}$-*structure* $\mathbf{M}$, it is said that $\mathbf{M}$ is a *model* of ϕ iff $\|\phi\|^{\mathbf{B}}_{\mathbf{M}} = 1$.

From now on, we say that $(\mathbf{M}, \mathbf{B})$ is a Γ-structure instead of saying that $\mathbf{M}$ is a $\mathbf{B}$-structure in the language Γ. We say that a structure is *safe*, if a truth value is defined for each formula and evaluation. We assume that all our structures are safe. It is denoted by $(\mathbf{M}, \mathbf{B}) \equiv (\mathbf{N}, \mathbf{A})$ when these two structures are elementarily equivalent. In this section we have presented only a few definitions and notation, a detailed introduction to the syntax and semantics of fuzzy predicate logics can be found in [9].

Definition 2 *Let $(\mathbf{M}_1, \mathbf{B}_1)$ be a Γ_1-structure and $(\mathbf{M}_2, \mathbf{B}_2)$ be a Γ_2-structure with $\Gamma_1 \subseteq \Gamma_2$. We say that the pair (f, g) is a* homomorphism *of $(\mathbf{M}_1, \mathbf{B}_1)$ into $(\mathbf{M}_2, \mathbf{B}_2)$ iff*

1. *$g : \mathbf{B}_1 \to \mathbf{B}_2$ is a L-algebra homomorphism of $\mathbf{B}_1$ into $\mathbf{B}_2$.*
2. *$f : M_1 \to M_2$ is a mapping of M_1 into M_2.*
3. *For each constant symbol $c \in \Gamma_1$, $f(c_{\mathbf{M}_1}) = c_{\mathbf{M}_2}$.*
4. *For each n-ary function symbol $F \in \Gamma_1$ and elements $d_1, \ldots, d_n \in M_1$,*

   $$f(F_{\mathbf{M}_1}(d_1, \ldots, d_n)) = F_{\mathbf{M}_2}(f(d_1), \ldots, f(d_n))$$

5. *For each n-ary predicate $P \in \Gamma_1$ and elements $d_1, \ldots, d_n \in M_1$,*

   $$g(P_{\mathbf{M}_1}(d_1, \ldots, d_n)) = P_{\mathbf{M}_2}(f(d_1), \ldots, f(d_n))$$

We say that (f, g) is a σ-homomorphism *if g preserves the existing infima and suprema (that is, if I is a non-empty set and $\sup_{i\in I} a_i$ and $\sup_{i\in I} g(a_i)$ exist, then $g(\sup_{i\in I} a_i) = \sup_{i\in I} g(a_i)$ and analogously for the infima).*

It is denoted by $(\mathbf{M}, \mathbf{B}) \cong (\mathbf{N}, \mathbf{A})$ when these two structures are isomorphic (that is, there is a homomorphism (f, g) from $(\mathbf{M}, \mathbf{B})$ into $(\mathbf{N}, \mathbf{A})$ with f and g onto and one-to-one). It is easy to check, by induction on the complexity of the formulas, that homomorphisms preserve quantifier-free formulas. Note that, by definition, homomorphisms are not always σ-complete, as are in [3] or [6], and unlike [6] homomorphisms are crisp on the algebraic reduct of the first-order structure. If (f, g) is a σ-homomorphism of $(\mathbf{M}_1, \mathbf{B}_1)$ into $(\mathbf{M}_2, \mathbf{B}_2)$ such that f is onto, then for each formula $\phi(x_1, \ldots, x_n) \in \Gamma_1$ and elements $d_1, \ldots, d_n \in M_1$,

$$g(\|\phi(d_1, \ldots, d_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}) = \|\phi(f(d_1), \ldots, f(d_n))\|^{\mathbf{B}_2}_{\mathbf{M}_2} \quad (1)$$

The proof can be found in [3] (Propositions 6.1 and 6.2). We will refer to homomorphisms satisfying condition (1) as *elementary homomorphisms*.

Definition 3 *A* congruence *on a Γ-structure $(\mathbf{M}, \mathbf{B})$ is a pair (E, θ) where:*

1. *θ is an L-congruence on the algebra $\mathbf{B}$.*
2. *E is an equivalence relation $E \subseteq M \times M$ such that:*

   - *For each n-ary function symbol $F \in \Gamma$ and elements $d_1, \ldots, d_n, e_1, \ldots, e_n \in M$, if for each $0 < i \leq n$, $(d_i, e_i) \in E$, then*

     $$(F_{\mathbf{M}}(d_1, \ldots, d_n), F_{\mathbf{M}}(e_1, \ldots, e_n)) \in E$$

   - *For each n-ary predicate $P \in \Gamma$ and elements $d_1, \ldots, d_n, e_1, \ldots, e_n \in M$, if for each $0 < i \leq n$, $(d_i, e_i) \in E$, then*

     $$(P_{\mathbf{M}}(d_1, \ldots, d_n), P_{\mathbf{M}}(e_1, \ldots, e_n)) \in \theta$$

Now, given a congruence (E, θ) on $(\mathbf{M}, \mathbf{B})$ we define the *quotient structure* $(\mathbf{M}/E, \mathbf{B}/\theta)$ by:

ISBN: 978-989-95079-6-8

- For each constant symbol $c \in \Gamma$, $c_{(\mathbf{M}/E,\mathbf{B}/\theta)} = [c_{(\mathbf{M},\mathbf{B})}]_E$.
- For each n-ary function symbol $F \in \Gamma$ and elements $d_1, \ldots, d_n \in M$,

$$F_{\mathbf{M}/E}([d_1]_E, \ldots, [d_n]_E) = [F_{\mathbf{M}}(d_1, \ldots, d_n)]_E$$

- For each n-ary predicate $P \in \Gamma$ and elements $d_1, \ldots, d_n \in M$,

$$P_{\mathbf{M}/E}([d_1]_E, \ldots, [d_n]_E) = [P_{\mathbf{M}}(d_1, \ldots, d_n)]_\theta$$

where, given an element $d \in M$ and $b \in B$, $[d]_E$ and $[b]_\theta$ denote respectively the equivalence classes of d modulo E and of b modulo θ. We will say that a (E, θ) is an *elementary* congruence (σ-congruence, respectively) if its canonical mapping (f_E, g_θ) is an elementary homomorphism (σ-homomorphism, respectively).

3 Reduced Structures

In this section we introduce the notions of Leibniz congruence and of reduced structure and we establish some basic model-theoretic properties of this kind of models, giving some examples of first-order theories with reduced models. The study of reduced structures and Leibniz congruences for classical predicate logics was done in [10]. In the context of fuzzy predicate logics, X. Caicedo introduced this notion in [11] for the particular case of models of first-order Rational Pavelka's logic in a language with the $\approx$ symbol. At the end of this section we study similarities on reduced structures.

Definition 4 *Let $(\mathbf{M}, \mathbf{B})$ be a Γ-structure and θ an L-congruence on $\mathbf{B}$. We define the relation $\Omega(\mathbf{M}, \mathbf{B}, \theta) \subseteq M \times M$ as follows: for every $d, e \in M$, $(d, e) \in \Omega(\mathbf{M}, \mathbf{B}, \theta)$ iff for every atomic formula, $\phi(y, x_1, \ldots, x_n) \in \Gamma$ and elements $d_1, \ldots, d_n \in M$,*

$$(\|\phi(d, d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}, \|\phi(e, d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

Fixed an L-congruence θ on the L-algebra $\mathbf{B}$, next lemma shows that $\Omega(\mathbf{M}, \mathbf{B}, \theta)$ is the greatest E such that (E, θ) is a congruence on the model $(\mathbf{M}, \mathbf{B})$.

Lemma 5 *Let $(\mathbf{M}, \mathbf{B})$ be a Γ-structure and θ an L-congruence on $\mathbf{B}$, then*

1. *$(\Omega(\mathbf{M}, \mathbf{B}, \theta), \theta)$ is a congruence on $(\mathbf{M}, \mathbf{B})$.*
2. *For every (E, θ) congruence on $(\mathbf{M}, \mathbf{B})$, $E \subseteq \Omega(\mathbf{M}, \mathbf{B}, \theta)$.*

Proof: 1. Since θ is an equivalence relation, by definition, $\Omega(\mathbf{M}, \mathbf{B}, \theta)$ is also an equivalence relation. For each n-ary predicate $P \in \Gamma$ and elements $d_1, \ldots, d_n, e_1, \ldots, e_n \in M$, if for each $0 < i \leq n$, $(d_i, e_i) \in \Omega(\mathbf{M}, \mathbf{B}, \theta)$, then by using the definition of $\Omega(\mathbf{M}, \mathbf{B}, \theta)$, for every $0 < i \leq n$, we have the following chain:

$$(\|P(d_1, d_2, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}, \|P(e_1, d_2, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

$$(\|P(e_1, d_2, d_3, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}, \|P(e_1, e_2, d_3, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

$$\vdots$$

$$(\|P(e_1, e_2, \ldots, e_{n-1}, d_n)\|^{\mathbf{B}}_{\mathbf{M}}, \|P(e_1, \ldots, e_n)\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

Assume now that $F \in \Gamma$ is an n-ary function symbol and $d_1, \ldots, d_n, e_1, \ldots, e_n \in M$ such that for each $0 < i \leq n$, $(d_i, e_i) \in \Omega(\mathbf{M}, \mathbf{B}, \theta)$. Let $\overline{k} = k_1, \ldots, k_s \in M$, $\phi(y, x_1, \ldots, x_s) \in \Gamma$ an atomic formula and ϕ' the formula obtained from ϕ by substitution of the variable y for the term $F(z_1, \ldots, z_n)$ (where $z_1, \ldots, z_n$ are new variables not occurring in ϕ). By definition of $\Omega(\mathbf{M}, \mathbf{B}, \theta)$, since ϕ' is also atomic, we can build a chain similar to the one defined in the predicate case and then, we obtain $(\|\phi'(d_1, \ldots, d_n, \overline{k})\|^{\mathbf{B}}_{\mathbf{M}}, \|\phi'(e_1, \ldots, e_n, \overline{k})\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$, consequently,

$$(\|\phi(F_{\mathbf{M}}(d_1, \ldots, d_n), \overline{k})\|^{\mathbf{B}}_{\mathbf{M}}, \|\phi(F_{\mathbf{M}}(e_1, \ldots, e_n), \overline{k})\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

and then $(F_{\mathbf{M}}(d_1, \ldots, d_n), F_{\mathbf{M}}(e_1, \ldots, e_n)) \in \Omega(\mathbf{M}, \mathbf{B}, \theta)$.

2. By definition of $\Omega(\mathbf{M}, \mathbf{B}, \theta)$, because (E, θ) is a congruence. □

Definition 6 *A Γ-structure $(\mathbf{M}, \mathbf{B})$ is* reduced *iff $\Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}})$ is the identity relation.*

From now on we denote $(\Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}}), Id_{\mathbf{B}})$ simply by $\Omega(\mathbf{M}, \mathbf{B})$ and we call it the *Leibniz congruence* of $(\mathbf{M}, \mathbf{B})$. Since the identity map clearly preserves infima and suprema, $\Omega(\mathbf{M}, \mathbf{B})$ is always a σ-congruence, therefore for every $(d, e) \in \Omega(\mathbf{M}, \mathbf{B})$, every formula, $\phi(y, x_1, \ldots, x_n) \in \Gamma$ and elements $d_1, \ldots, d_n \in M$,

$$\|\phi(d, d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}} = \|\phi(e, d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}$$

We will denote by $(\mathbf{M}, \mathbf{B})^r$ the quotient structure modulo the Leibniz congruence $\Omega(\mathbf{M}, \mathbf{B})$ and call it the *reduction* of $(\mathbf{M}, \mathbf{B})$.

Lemma 7 *For every Γ-structure $(\mathbf{M}, \mathbf{B})$ we have:*

1. *$(\mathbf{M}, \mathbf{B}) \equiv (\mathbf{M}, \mathbf{B})^r$.*
2. *$(\mathbf{M}, \mathbf{B})^r$ is a reduced structure.*
3. *$((\mathbf{M}, \mathbf{B})^r)^r \cong (\mathbf{M}, \mathbf{B})^r$.*
4. *If there exists a σ-congruence (E, θ) on $(\mathbf{M}, \mathbf{B})$, then*

$$(\mathbf{M}, \mathbf{B}) \equiv (\mathbf{M}/\Omega(\mathbf{M}, \mathbf{B}, \theta), \mathbf{B}/\theta)$$

Proof: 1. holds because $(\mathbf{M}, \mathbf{B})^r$ is a σ-homomorphic image of $(\mathbf{M}, \mathbf{B})$, 2. and 3. by definition of the Leibniz congruence and of quotient structure. To prove 4. use the fact that for every σ-congruence (E, θ), $(\Omega(\mathbf{M}, \mathbf{B}, \theta), \theta)$ is also a σ-congruence. □

Corollary 8 [Completeness Theorem] *Let $L\forall$ be a fuzzy predicate logic and $\mathbb{K}$ a class of structures such that $L\forall$ is $\mathbb{K}$-complete (strong or finite strong $\mathbb{K}$-complete, respectively), then $L\forall$ is $\mathbb{K}^r$-complete (strong or finite strong $\mathbb{K}^r$-complete, respectively), where $\mathbb{K}^r$ is the class of reductions of the structures in $\mathbb{K}$.*

ISBN: 978-989-95079-6-8

Reduced structures are of common use in computer science and in mathematics. The Rado Graph (infinite random graph) and fuzzy linear orders are examples of reduced structures. However, we can see a more developed example of this process of reduction in the well-known case of similarities. In Section 5 of [9], P. Hájek studies similarities and applies the obtained results to the analysis of fuzzy control in Chapter 7 of of [9]. For a reference about model-theoretic properties of algebras with fuzzy equalities see [8] and [6]. Now (and only for the rest of this section) we assume that our predicate language Γ contains a binary predicate symbol $\approx$. Similarity is understood as fuzzified equality (for a reference see [12] or [13]). Given a core fuzzy logic L, let our axiomatic system for $L\forall$ contain also the following axioms:

1. (Reflexivity) $\forall x\ x \approx x$
2. (Symmetry) $\forall x \forall y (x \approx y \rightarrow y \approx x)$
3. (Transitivity) $\forall x \forall y \forall z ((x \approx y \& y \approx z) \rightarrow x \approx z)$
4. For each n-ary function symbol $F \in \Gamma$, $\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall y_n ((x_1 \approx y_1 \& \ldots \& x_n \approx y_n) \rightarrow (F(x_1, \ldots, x_n) \approx F(y_1, \ldots, y_n))$
5. For each n-ary predicate $P \in \Gamma$, $\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall y_n ((x_1 \approx y_1 \& \ldots \& x_n \approx y_n) \rightarrow (P(x_1, \ldots, x_n) \leftrightarrow P(y_1, \ldots, y_n))$

Axioms 1-3 are called *Similarity Axioms* (Sim) and axioms 4-5 *Congruence Axioms* (Cong).

Definition 9 *A Γ-structure $(\mathbf{M}, \mathbf{B})$ has the* equality property (EQP) *if the following condition holds: for every $d, e \in M$, $\|d \approx e\|^{\mathbf{B}}_{\mathbf{M}} = 1$ iff $d = e$.*

Lemma 10 *Given a set Σ of Γ-sentences, $\Sigma \cup Sim \cup Cong$ is satisfiable iff Σ has a model $(\mathbf{M}, \mathbf{B})$ that has* EQP.

Proof: Let $(\mathbf{M}, \mathbf{B})$ be a model of $\Sigma \cup Sim \cup Cong$. If we define $E = \{(a, b) \in M \times M : \|a \approx b\|^{\mathbf{B}}_{\mathbf{M}} = 1\}$, then $(E, Id_{\mathbf{B}})$ is a congruence. Thus, in the quotient structure $(\mathbf{M}/E, \mathbf{B})$, $\|x \approx y\|^{\mathbf{B}}_{\mathbf{M}/E} = 1$ iff $x = y$. □

In [11] X. Caicedo called *reduced structure* to a model of first-order Rational Pavelka's logic with the property EQP. Next lemma shows that Caicedo's notion coincides with ours when we consider axiomatic systems including axioms $Sim \cup Cong$.

Lemma 11 *$(\mathbf{M}, \mathbf{B})$ is a reduced structure iff $(\mathbf{M}, \mathbf{B})$ has property* EQP.

Proof: Assume that $(\mathbf{M}, \mathbf{B})$ is a reduced structure, by definition, the Leibniz congruence $\Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}})$ is the identity on M. Then, if we define $(E, Id_{\mathbf{B}})$ as in the previous proof, by Lemma 5, $E \subseteq \Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}})$, consequently E is also the identity and thus $(\mathbf{M}, \mathbf{B})$ has the EQP. Conversely, assume that $(\mathbf{M}, \mathbf{B})$ has the EQP. If $d, e \in M$ and $(d, e) \in \Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}})$, since $x \approx y$ is an atomic formula and $\|d \approx d\|^{\mathbf{B}}_{\mathbf{M}} = 1$, by definition of the Leibniz congruence, $\|d \approx e\|^{\mathbf{B}}_{\mathbf{M}} = 1$. Then, by EQP, we have that $d = e$, therefore $(\mathbf{M}, \mathbf{B})$ is a reduced structure. □

Note that the interpretation of the $\approx$ symbol in a reduced structure is not necessarily crisp. Adding a new axiom it is possible to obtain crisp interpretations: Crispness Axiom (Crisp) $\forall x \forall y (x \approx y \vee \neg(x \approx y))$ (for the details about this axiom see Chapter 5 of [8]).

Corollary 12 *Let T be a set of Γ-sentences containing axioms $Sim \cup Cong \cup Crisp$. Then, for every formula $\phi(x_1, \ldots, x_n) \in \Gamma$, the following holds:* $T \vdash \forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall y_n ((x_1 \approx y_1 \& \ldots \& x_n \approx y_n) \rightarrow (\phi(x_1, \ldots, x_n) \leftrightarrow \phi(y_1, \ldots, y_n))$

Proof: Since axiom Crisp holds, the interpretation of the $\approx$ symbol in a reduced structure is the identity. Therefore, since the logic is complete with respect to its reduced models, we obtain the desired result. □

4 The relative relation

We now present the notion of relative relation, a relation between structures that will play in fuzzy predicate languages the same role that the isomorphism relation plays in classical predicate languages with equality. This notion was introduced by G. Zubieta in [14], but only for relational structures, and independently by W. Blok and D. Pigozzi in [1], for the special case of logical matrices. A characterization of the relative relation for classical first-order logics can be found in [10].

Definition 13 *Let $(\mathbf{M}_1, \mathbf{B}_1)$ and $(\mathbf{M}_2, \mathbf{B}_2)$ be two Γ-structures, we say that the pair (R, T) is a* relative relation *between $(\mathbf{M}_1, \mathbf{B}_1)$ and $(\mathbf{M}_2, \mathbf{B}_2)$ iff*

1. *$T \subseteq B_1 \times B_2$ is a relation such that $dom(T) = B_1$, $rg(T) = B_2$ and*
   (a) *for every connective $\delta \in L$, if $(a_i, b_i) \in T$, then*
   $$(\delta_{\mathbf{B}_1}(a_1, \ldots, a_n), \delta_{\mathbf{B}_2}(b_1, \ldots, b_n)) \in T$$
   (b) *for every $a \in B_1$, if $b, b' \in rg(a)$, then $dom(b) = dom(b')$.*
   (c) *for every $b \in B_2$, if $a, a' \in dom(b)$, then $rg(a) = rg(a')$.*
2. *$R \subseteq M_1 \times M_2$ is a relation such that $dom(R) = M_1$, $rg(R) = M_2$ and*
   (a) *For each constant symbol $c \in \Gamma$, $(c_{\mathbf{M}_1}, c_{\mathbf{M}_2}) \in R$.*
   (b) *For each n-ary function symbol $F \in \Gamma$, if for every $0 < i \leq n$, $(a_i, b_i) \in R$,*
   $$(F_{M_1}(a_1, \ldots, a_n), F_{M_2}(b_1, \ldots, b_n)) \in R$$
   (c) *For each n-ary predicate $P \in \Gamma$, if for every $0 < i \leq n$, $(d_i, e_i) \in R$,*
   $$(\|P(d_1, \ldots, d_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}, \|P(e_1, \ldots, e_n))\|^{\mathbf{B}_2}_{\mathbf{M}_2}) \in T$$

where for every $a \in B_1$, $rg(a) = \{b \in B_2 : (a, b) \in T\}$ and for every $b \in B_2$, $dom(b) = \{a \in B_1 : (a, b) \in T\}$. We denote by $(R, T) : (\mathbf{M}_1, \mathbf{B}_1) \sim (\mathbf{M}_2, \mathbf{B}_2)$ when (R, T) is a relative relation (or simply by $(\mathbf{M}_1, \mathbf{B}_1) \sim (\mathbf{M}_2, \mathbf{B}_2)$ when there is a relative relation between them).

ISBN: 978-989-95079-6-8

Theorem 14 *Let* $(\mathbf{M}_1, \mathbf{B}_1)$ *and* $(\mathbf{M}_2, \mathbf{B}_2)$ *be two* Γ*-structures. The following are equivalent:*

1. *There is a relative relation* $(R, T) : (\mathbf{M}_1, \mathbf{B}_1) \sim (\mathbf{M}_2, \mathbf{B}_2)$
2. *There are congruences* (E_1, θ_1) *and* (E_2, θ_2) *such that*

$$(\mathbf{M}_1/E_1, \mathbf{B}_1/\theta_1) \cong (\mathbf{M}_2/E_2, \mathbf{B}_2/\theta_2)$$

Proof: 1. $\Rightarrow$ 2. Assume that (R, T) is a relative relation. We define $\theta_1 = \{(a, a') \in B_1 \times B_1 : rg(a) = rg(a')\}$, $\theta_2 = \{(b, b') \in B_2 \times B_2 : dom(b) = dom(b')\}$, $E_1 = \Omega(\mathbf{M_1}, \mathbf{B_1}, \theta_1)$ and $E_2 = \Omega(\mathbf{M_2}, \mathbf{B_2}, \theta_2)$.

It is clear by definition that θ_1 and θ_2 are equivalence relations. Now we show that they are L-congruences, we prove that, for every connective $\delta \in L$, for every $0 < i \leq n$ and $a_i, a_i' \in B_1$, if $rg(a_i) = rg(a_i')$, then $rg(\delta_{\mathbf{B}_1}(a_1, \ldots, a_n)) = rg(\delta_{\mathbf{B}_1}(a_1', \ldots, a_n'))$. Let us assume that for every $0 < i \leq n$, $rg(a_i) = rg(a_i')$. Since $dom(T) = B_1$, for every $0 < i \leq n$, we choose $b_i \in B_2$ such that $(a_i, b_i) \in T$. Thus, by assumption, since $rg(a_i) = rg(a_i')$, we have also that $(a_i', b_i) \in T$. By condition 1.(a) of the definition of relative relation,

$$(\delta_{\mathbf{B}_1}(a_1, \ldots, a_n), \delta_{\mathbf{B}_2}(b_1, \ldots, b_n)) \in T$$

and $(\delta_{\mathbf{B}_1}(a_1', \ldots, a_n'), \delta_{\mathbf{B}_2}(b_1, \ldots, b_n)) \in T$, finally, by condition 1.(c) of the definition of relative relation, we have that

$$rg(\delta_{\mathbf{B}_1}(a_1, \ldots, a_n)) = rg(\delta_{\mathbf{B}_1}(a_1', \ldots, a_n')).$$

In order to show that θ_2 is a congruence, we can follow an analogous proof, using condition 1.(b) of the definition of relative relation, instead of 1.(c). Now we define a mapping $g : \mathbf{B}_1/\theta_1 \to \mathbf{B}_2/\theta_2$. First we fix enumerations (possibly with repetitions) $(a_i : i \in I)$ and $(b_i : i \in I)$ of $\mathbf{B}_1$ and $\mathbf{B}_2$ respectively, with the property that, for every $i \in I$, $(a_i, b_i) \in T$. And then let, for every $i \in I$, $g([a_i]_{\theta_1}) = [b_i]_{\theta_2}$. By using the definition of relative relation it is easy to check that g is well-defined, and it is an L-isomorphism.

Now, since $dom(R) = M_1$ and $rg(R) = M_2$, we can fix enumerations (possibly with repetitions) $(d_j : j \in J)$ and $(e_j : j \in J)$ of M_1 and M_2 respectively, with the property that, for every $j \in J$, $(d_j, e_j) \in R$. And then let, for every $j \in J$,

$$f([d_j]_{\Omega(\mathbf{M_1}, \mathbf{B_1}, \theta_1)}) = [e_j]_{\Omega(\mathbf{M_2}, \mathbf{B_2}, \theta_2)}$$

Let us see that f is well defined. Let $(d_j, d_j') \in \Omega(\mathbf{M_1}, \mathbf{B_1}, \theta_1)$, we show that $(e_j, e_j') \in \Omega(\mathbf{M_2}, \mathbf{B_2}, \theta_2)$. Let $\phi(y, x_1, \ldots, x_n)$ be an atomic formula, and a sequence of elements $k_1, \ldots, k_n \in M_2$. Since $rg(R) = M_2$, we can choose $l_1, \ldots, l_n \in M_1$ such that for every $0 < i \leq n$, $(l_i, k_i) \in R$. Remark that, since $(d_j, e_j) \in R$ and for every $0 < i \leq n$, $(l_i, k_i) \in R$, by conditions 2.(a) and 2.(b) of the definition of relative relation, we have for every Γ-term t, $(t_{M_1}(d_j, l_1, \ldots, l_n), t_{M_2}(e_j, k_1, \ldots, k_n)) \in R$. Assume that the formula $\phi(y, x_1, \ldots, x_n)$ is of the form $P(t_1, \ldots, t_s)(y, x_1, \ldots, x_n)$, where P is a s-ary predicate symbol and $t_1, \ldots, t_s$ are Γ-terms. Since $(d_j, d_j') \in \Omega(\mathbf{M_1}, \mathbf{B_1}, \theta_1)$, $(\|P(t_1, \ldots, t_s)(d_j, l_1, \ldots, l_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}$, $\|P(t_1, \ldots, t_s)(d_j', l_1, \ldots, l_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}) \in \theta_1$, thus, by 2.(c) of the definition of relative relation,
$\|P(t_1, \ldots, t_s)(e_j, k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}$ and
$\|P(t_1, \ldots, t_s)(e_j', k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}$
$\in rg(\|P(t_1, \ldots, t_s)(d_j, l_1, \ldots, l_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1})$

and by 1.(b), $dom(\|P(t_1, \ldots, t_s)(e_j, k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}) = dom(\|P(t_1, \ldots, t_s)(e_j', k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2})$ and we obtain the desired result:

$$(\|\phi(e_j, k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}, \|\phi(e_j', k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}) \in \theta_2.$$

In an analogous way it is easy to check that f is one-to-one and that (f, g) is an isomorphism.

2. $\Rightarrow$ 1. Let (f, g) be an isomorphism. Define (R, T) in the following way: for every $b_1 \in B_1, b_2 \in B_2$,$(a, b) \in T$ iff $g([a]_{\theta_1}) = [b]_{\theta_2}$ and for every $d \in M_1, e \in M_2$, $(d, e) \in R$ iff $f([d]_{E_1}) = [e]_{E_2}$. Using the fact that (f, g) is an isomorphism, it is easy to check that $(R, T) : (\mathbf{M}_1, \mathbf{B}_1) \sim (\mathbf{M}_2, \mathbf{B}_2)$ is a relative relation. □

Remark that our approach differs from [6] because the relative relation is not a measure of the degree of similarity between structures. We left for future work the study of the relationship between these two notions.

Definition 15 (R, T) *is an* elementary *relative relation if for every formula* $\phi(x_1, \ldots, x_n) \in \Gamma$ *and for every* $0 < i \leq n$*, if* $(a_i, b_i) \in R$*, then*

$$(\|\phi(a_1, \ldots, a_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}, \|\phi(b_1, \ldots, b_n))\|^{\mathbf{B}_2}_{\mathbf{M}_2}) \in T$$

By induction on the complexity of the formulas it is straightforward to show that, in case (E_1, θ_1) and (E_2, θ_2) are elementary congruences in Theorem 14, then (R, T) is also an elementary relative relation and $(\mathbf{M}_1, \mathbf{B}_1) \equiv (\mathbf{M}_2, \mathbf{B}_2)$. The following corollary show us that, when we study structures over the same algebra, we can improve Theorem 14.

Corollary 16 *Let* $(\mathbf{M}_1, \mathbf{B})$ *and* $(\mathbf{M}_2, \mathbf{B})$ *be two* Γ*-structures. The following are equivalent:*

1. *There is a relative relation* $(R, Id_{\mathbf{B}}) : (\mathbf{M}_1, \mathbf{B}) \sim (\mathbf{M}_2, \mathbf{B})$
2. $(\mathbf{M}_1, \mathbf{B})^r \cong (\mathbf{M}_2, \mathbf{B})^r$

Proof: By the proof of Theorem 14. □

Let us see now an example of two fuzzy equivalence relations that are relatives but there is no homomorphism from one onto the other. Let $(\mathbf{M}_1, \mathbf{B})$ and $(\mathbf{M}_2, \mathbf{B})$ be defined as follows: the domains of the structures are $M_1 = \{d_1, d_2, e_1, e_2\}$ and $M_2 = \{d_1', d_2', d_3', e_1'\}$ respectively. The fuzzy equivalence relation E_1 is defined by: for every $i, j \in \{1, 2\}$, $E_1(d_i, d_j) = 1 = E_1(e_i, e_j)$ and $E_1(d_i, e_j) = r$, where $r \neq 1$ is a fixed element of $\mathbf{B}$. And the fuzzy equivalence relation E_2 is defined by: for every $i, j \in \{1, 2, 3\}$, $E_2(d_i', d_j') = 1 = E_2(e_1', e_1')$ and $E_2(d_i', e_1') = r$. It is easy

ISBN: 978-989-95079-6-8

to check that $(R, Id_{\mathbf{B}}) : (\mathbf{M}_1, \mathbf{B}) \sim (\mathbf{M}_2, \mathbf{B})$, where R is the relation $R = \{(d_i, d'_j), (e_i, e'_j) : i, j \in \{1, 2, 3\}\}$.

The relation of being either a homomorphic image or a homomorphic counter-image is not in general transitive. Its transitivization is precisely the relative relation, as the following propositions show.

Lemma 17 *Given an L-algebra* $\mathbf{B}$*,* $\sim$ *is an equivalence relation in the class of* $\mathbf{B}$*-structures.*

Proof: By Corollary 16, because $\cong$ is an equivalence relation on the class of reduced structures. □

Notation: Given Γ-structures $(\mathbf{N}, \mathbf{B})$ and $(\mathbf{O}, \mathbf{B})$, we denote by $(\mathbf{N}, \mathbf{B}) \in \mathbf{H}(\mathbf{O}, \mathbf{B})$ the fact that there exists a mapping f from O onto N such that $(f, Id_{\mathbf{B}})$ is a homomorphism.

Proposition 18 *Let* $(\mathbf{M}, \mathbf{B})$ *and* $(\mathbf{N}, \mathbf{B})$ *be two* Γ*-structures. The following are equivalent:*

1. $(\mathbf{M}, \mathbf{B}) \sim (\mathbf{N}, \mathbf{B})$.
2. *There is a natural number* n *and* Γ*-structures* $(\mathbf{O}_1, \mathbf{B}), \ldots, (\mathbf{O}_n, \mathbf{B})$ *such that* $(\mathbf{M}, \mathbf{B}) = (\mathbf{O}_1, \mathbf{B})$*,* $(\mathbf{N}, \mathbf{B}) = (\mathbf{O}_n, \mathbf{B})$ *and for every* $0 < i < n$*, either* $(\mathbf{O}_{n+1}, \mathbf{B}) \in \mathrm{H}(\mathbf{O}_n, \mathbf{B})$ *or* $(\mathbf{O}_n, \mathbf{B}) \in \mathrm{H}(\mathbf{O}_{n+1}, \mathbf{B})$.
3. *There is a* Γ*-structure* $(\mathbf{O}, \mathbf{B})$ *such that* $(\mathbf{M}, \mathbf{B}), (\mathbf{N}, \mathbf{B}) \in \mathrm{H}(\mathbf{O}, \mathbf{B})$.
4. *There is a* Γ*-structure* $(\mathbf{O}, \mathbf{B})$ *such that* $(\mathbf{O}, \mathbf{B}) \in \mathrm{H}(\mathbf{M}, \mathbf{B})$ *and* $(\mathbf{O}, \mathbf{B}) \in \mathrm{H}(\mathbf{N}, \mathbf{B})$.

Proof: 4. $\Rightarrow$ 2. and 3. $\Rightarrow$ 2. are clear. 1. $\Rightarrow$ 4. By Corollary 16. 2. $\Rightarrow$ 1. By the definition of relative relation, given two Γ-structures $(\mathbf{O}, \mathbf{B})$ and $(\mathbf{D}, \mathbf{B})$, if $(\mathbf{O}, \mathbf{B}) \in \mathbf{H}(\mathbf{D}, \mathbf{B})$, then $(\mathbf{O}, \mathbf{B}) \sim (\mathbf{D}, \mathbf{B})$ therefore we can apply the transitive property of the relative relation (Lemma 17).

1. $\Rightarrow$ 3. Assume that there is a relative relation $(R, Id_{\mathbf{B}}) : (\mathbf{M}, \mathbf{B}) \sim (\mathbf{N}, \mathbf{B})$. Since $dom(R) = M$ and $rg(R) = N$, we can fix enumerations (possibly with repetitions) $(d_j : j \in J)$ and $(e_j : j \in J)$ of M and N respectively, with the property that, for every $j \in J$, $(d_j, e_j) \in R$. Now we define a structure $(\mathbf{O}, \mathbf{B})$ and homomorphisms $(f^M, Id_{\mathbf{B}})$ and $(f^N, Id_{\mathbf{B}})$ from $(\mathbf{O}, \mathbf{B})$ onto $(\mathbf{M}, \mathbf{B})$ and $(\mathbf{N}, \mathbf{B})$ respectively.

The algebraic reduct of $(\mathbf{O}, \mathbf{B})$ is the algebra Ter_J of Γ-terms generated by the set of variables $V_J = \{v_j : j \in J\}$. We define the function $f_0^M : V_J \to M$ as follows: for every $j \in J$, $f_0^M(v_j) = d_j$. Then we extend f_0^M in the usual way, to a homomorphism f^M from Ter_J onto M. Finally we define the interpretation of the predicate symbols in $(\mathbf{O}, \mathbf{B})$: for each n-ary predicate $P \in \Gamma$ and elements $t_1, \ldots, t_n \in O$, $P_O(t_1, \ldots, t_n) = P_M(f^M(t_1), \ldots, f^M(t_n))$. So defined $(f^M, Id_{\mathbf{B}})$ is clearly a homomorphism onto $(\mathbf{M}, \mathbf{B})$. Now we define f_0^N by: for every $j \in J$, $f_0^N(v_j) = e_j$ and we extend f_0^N as before, to a homomorphism f^N of the terms algebra onto N. Since for every $j \in J$, $(d_j, e_j) \in R$, for every Γ-term t,

$$(t_M(d_{j_1}, \ldots, d_{j_n}), t_N(e_{j_1}, \ldots, e_{j_n})) \in R$$

and using this fact it is easy to check that $(f^N, Id_{\mathbf{B}})$ is also a homomorphism onto $(\mathbf{N}, \mathbf{B})$. □

5 Conclusions

Work in progress includes the development of usual tools of model theory such as the method of diagrams or ultraproducts in order to work in fuzzy predicate logic. The use of a reduced semantics will allow us to show when one structure is either embeddable or elementarily embeddable in another in terms of extensions of the usual diagrams with special sentences. By using relative relations we could define new operations among structures, such as *ultrafilter-products*, more suitable for working with equality-free languages. Future work will be devoted also to provide different characterizations of the relation of elementary equivalence and some strenghtenings of this notion.

Acknowledgment

Research partially supported by the Generalitat de Catalunya, under the grant 2005-SGR-00093 and the Spanish project "Agreement Technologies" (CONSOLIDER CSD2007-0022, INGENIO 2010).

References

[1] W. J. Blok and D. Pigozzi. Protoalgebraic logics. 45:337–369, 1986.

[2] W. J. Blok and D. Pigozzi. *Algebraizable logics*, volume 396 of *Mem. Amer. Math. Soc.* A.M.S., Providence, January 1989.

[3] A. Di Nola and G. Gerla. Fuzzy models of first order languages. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 32:331–340, 1986.

[4] P. Hájek and P. Cintula. On theories and models in fuzzy predicate logics. *The Journal of Symbolic Logic*, 71(3):863–880, 2006.

[5] V. Novák et al. *Mathematical principles of fuzzy logic*. The Kluwer International Series in Engineering and Computer Science, 517. Kluwer Academic Publishers, Boston, MA, 1999.

[6] R. Belohlávek. *Fuzzy Relational Systems: Foundations and Principles.* Springer, 2002.

[7] P. Cintula et al. Distinguished algebraic semantics for t-norm based fuzzy logics: methods and algebraic equivalencies. *Annals of Pure and Applied Logic (article in press)*, 2009. Submitted.

[8] P. Cintula and P. Hájek. Triangular norm based predicate fuzzy logics. In *Fuzzy Sets and Systems ((article in press)*, 2009. Forthcoming.

[9] P. Hájek. *Metamathematics of fuzzy logic*, volume 4 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 1998.

[10] P. Dellunde. *Contributions to the model theory of equality-free logic*. Ph. D. Dissertation, University of Barcelona, 1996.

[11] X. Caicedo. A characterization of first order rational pavelka's logic. Preprint, slides of the talk in the Conference *Logic, Algebra and Truth Degrees*, Siena http://www.mat.unisi.it/ latd2008/slides/Caicedo.pdf (2008).

[12] E. H. Ruspini. On the semantics of fuzzy logic. *International Journal of Approximate Reasoning*, 5(1):45–88, 1991.

[13] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[14] G. Zubieta. Clases aritméticas definidas sin igualdad. *Boletín de la Sociedad Matemática Mejicana*, 2:45–53, 1957.

ISBN: 978-989-95079-6-8

Fuzzy Clustering for Finding Fuzzy Partitions of Many-Valued Attribute Domains in a Concept Analysis Perspective

Yassine Djouadi[1] Basma Alouane[2] Henri Prade[1]

1. IRIT, University Paul Sabatier, Toulouse
118 Route de Narbonne, 31062 Toulouse, Cedex 9, France
2. LIFAB Laboratory, University of Boumerdes
Avenue de l'Independance, Boumerdes, Algeria
Email: djouadi@irit.fr, B_alouane@umbb.dz, prade@irit.fr

Abstract— *Although an overall knowledge discovery process consists of a distinct pre-processing stage followed by the data mining step, it seems that existing formal concept analysis (FCA) and association rules mining (ARM) approaches, dealing with many-valued contexts, mainly focus on the data mining stage. An "intelligent" pre-processing of input contexts is often absent in existing FCA/ARM approaches, leading to an unavoidable information loss. Usually, many-valued attribute domains need to be first fuzzily partitioned. However, it is unrealistic that the most appropriate fuzzy partitions can be provided by domain experts. In this paper, an unsupervised learning stage, based on Fuzzy C-Means algorithm, is proposed in order to get fuzzy partitions that are faithful to data for quantitative attribute domains, and consequently for avoiding the loss of valuable association rules due to the use of empirical fuzzy partitions. More precisely, the paper reports an experiment where it is shown that some rules are no longer found because their support or confidence is too low when using such empirical partitions. Experimental results show that the learned fuzzy partition outperforms human expert fuzzy partitions. More generally, the paper provide discussions about the handling of many-valued attributes in both fuzzy FCA and fuzzy ARM.*

Keywords— Many-valued formal contexts, fuzzy partitions, fuzzy C-means, association rules.

1 Introduction

Association rules mining (ARM for short) [1] is one of the most widely used data mining technique. This model of knowledge represents the patterns of co-occurrence of items in a collection of transactions. Typically, input data set is in the form of sets of items called transactions (e.g. an item may be things we buy in a market). Thus an example of an association rule is an expression of the form *"beer"* $\wedge$ *"sausage"* $\Rightarrow$ *"mustard"* which represents the fact that purchasing beer and sausage implies purchasing mustard with some degree. It is important to point out that, in the classical setting, considered input data sets (called *contexts* in the rest of this paper) are binary (i.e. are expressed in terms of binary attributes).

Formal concept analysis [2] (FCA for short) consists also of learning some knowledge representation model in an unsupervised way. In the classical setting of this theory, the considered input data set is called *formal context* (also called *context* in this paper). It consists of a binary relation between a set of objects and a set of properties. This relation is usually represented as a table with rows corresponding to objects, columns corresponding to properties (or conversely), and table entries containing 1's or 0's depending on whether an object has or not the corresponding property.

It appears that, in their classical setting, both FCA and ARM are concerned with crisp binary contexts. However real-world data usually contain heterogenous kinds of data (binary, categorical, quantitative, etc...). In order to generalize binary settings to many-valued contexts (i.e. with quantitative or categorical attributes) it is usually suggested to partition quantitative attribute domain into many fuzzy intervals.

Few approaches propose a complete pre-processing stage that is intended to get an optimal fuzzy partition. In [3], authors propose a genetic algorithm-based clustering method that adjusts the centroids of the clusters, which are to be handled as midpoints of the triangular fuzzy partitions. Another approach, also based on genetic algorithms, is proposed in [4]. Based on a linguistic representation model, authors propose to perform a genetic lateral tuning of the membership functions. Authors consider also that each attribute has a predefined number of linguistic terms associated to it. For example, *Low*, *Middle* and *High* are the fuzzy linguistic terms covering the domain of the attribute *Age*.

Concerning these approaches, one may notice that they consider fuzzy partitions restricted to triangular membership functions. On the other hand, the number of clusters has to be empirically fixed for each attribute. Then, the fuzzy partitioning may be considered as ad-hoc to some extent and is far to be information lossless. In other words, the distribution of the data into the context is not taken into account for the assessment of discovered patterns. For this purpose, we present an approach that consists of learning the fuzzy partitions by means of the well-known Fuzzy C-Means algorithm. We also consider cluster validity measures in order to get an optimal multi-prototype context representation.

The paper is organized as follows. Section 2 gives some basic notions of FCA and ARM frameworks, while both crisp and fuzzy partitioning methods are reviewed. The next section highlights the information loss problem with empirical or expert-based fuzzy partitions. In the fourth section, data faithful fuzzy partitions are learned using Fuzzy C-Means algorithm. Experimental results on the *adult* database (UCI Machine Learning Repository) show that the proposed approach avoids information loss and outperforms in terms of support and confidence measures, expert-based partitions. Finally, we conclude and give future perspectives.

ISBN: 978-989-95079-6-8

2 Basic Notions

This section introduces main notions and terminology for both FCA and ARM paradigms and then, presents quantitative context representation and its inherent sharp boundary problem.

2.1 Association Rules Mining

Formally, let $\mathcal{X}$ be a set of objects (called also transactions) $\left(\mathcal{X} = \{x_1, x_2, ..., x_{|\mathcal{X}|}\}\right)$ and let $\mathcal{Y}$ be a set of items $\left(\mathcal{Y} = \{y_1, y_2, ..., y_{|\mathcal{Y}|}\}\right)$. A context $\mathbb{K}$ is denoted as $\mathbb{K} = (\mathcal{X}, \mathcal{Y}, \mathcal{R})$ where $\mathcal{R}$ is a binary relation ($x\mathcal{R}y$ indicates that item y satisfies the transaction x). An association rule is of the form $A \Rightarrow B$ where $A, B \subseteq \mathcal{Y}$, $A \cap B = \emptyset$ and, $A, B \neq \emptyset$. The intended meaning of this rule is that the presence of all of the items of A in a transaction implies the presence of all of the items of B in the same transaction. Association rules are assigned support and confidence measures for a given context. The support defines the percentage of transactions that satisfy (contain) both A and B. Whereas the confidence expresses the conditional probability that B is satisfied given that A is satisfied. These measures denoted respectively `Sup` and `Conf` are defined as:

$$\texttt{Sup}\,(A \Rightarrow B) = \frac{|\{x \in \mathcal{X} \mid \forall y \in A \cup B,\, x\mathcal{R}y\}|}{|\mathcal{X}|} \quad (1)$$

$$\texttt{Conf}(A \Rightarrow B) = \frac{\texttt{Sup}(A \Rightarrow B)}{\texttt{Sup}(A)} \quad (2)$$

Hence, an itemset A is said frequent, if its support is greater than a fixed threshold ξ (i.e. $\texttt{Sup}(A) \geqslant \xi$). A large variety of representational and computational approaches for association rules mining have been published in the literature in the case of binary items. Among them, frequent-based [5] and closed-based [6] approaches. While the former generate the lattice of all frequent itemsets, the latter generate a more condensed representation through the so-called closed itemsets lattice. For instance, based on the notion of closed itemsets, and minimal implication rules base (e.g. Duquenne-Guigues base, Luxemburger base) [7], association rules may be easily inferred from a closed itemsets lattice structure without enumerating them [8]. This kind of approach is closely related to formal concept analysis since the closed itemsets lattice is isomorphic to the concepts lattice which is described below.

2.2 Formal Concept Analysis

Formal concept analysis theory, proposed by [2], deals with a particular kind of analysis of data based on a formal context. Let us consider a similar notation as used above. A set $\mathcal{X}$ of objects, a set $\mathcal{Y}$ of attributes and a formal context $\mathbb{K} = (\mathcal{X}, \mathcal{Y}, \mathcal{R})$ with $\mathcal{R}$ a binary relation ($x\mathcal{R}y$ means that object x verifies attribute y). Let A (resp. B) be a subset of $\mathcal{X}$ (resp. $\mathcal{Y}$), two mappings $\uparrow$ and $\downarrow$ are symmetrically defined:

$$A\uparrow = \{y \mid \forall x \in A,\, x\mathcal{R}y\}$$
$$B\downarrow = \{x \mid \forall y \in B,\, x\mathcal{R}y\}$$

From a formal context, one can construct pairs (A, B) such that $A\uparrow = B$ and $B\downarrow = A$ known as formal concepts, where A and B are called respectively the extent and the intent of the corresponding concept. The set $\mathfrak{B}(\mathbb{K})$ of all formal concepts in the formal context $\mathbb{K}$ is equipped with a partial order (denoted $\preceq$) defined as:

$$(A_1, B_1) \preceq (A_2, B_2) \text{ iff } A_1 \subseteq A_2 \text{ (or, equivalently, } B_1 \supseteq B_2)$$

$(\mathfrak{B}(\mathbb{K}), \preceq)$ forms a complete lattice, called the *concept lattice* of $\mathbb{K}$ (for more details on the lattice structure see [9]).

It appears that ARM and more generally FCA both require, in their classical setting, the same context representation which is a binary context.

2.3 Quantitative Context Representation

In order to generalize binary context settings to many-valued contexts, Srikant et al. [10] suggested to partition a quantitative attribute domain into intervals. In their approach, these authors proposed to map a quantitative or categorical attribute value into a Boolean attribute value in the following way. Let y be an attribute whose domain is quantitative (the principle remains the same with a categorical attribute). Let $Dom\,(y)$ denote the domain of y (according to the relational paradigm, $Dom\,(y)$ is necessarily a finite set). For a given object x we denote by $x[y]$ the value v of y for the object x (i.e. $x[y] = v$).

If the set $Dom\,(y)$ consists of few values, the mapping is straightforward. Conceptually, the authors propose to have instead of the attribute y, as many new attributes as the cardinality of the set $Dom\,(y)$. The new attributes are of the form "$y.v_k$" (where $v_k \in Dom\,(y)$). The values of the new binary formal context are given as:

$$x[y.v_k] = \begin{cases} 1 & \text{if } x[y] = v_k \\ 0 & \text{otherwise.} \end{cases}$$

If the set $Dom\,(y)$ is large, it is partitioned into different intervals $I_1, I_2, ..., I_p$. Attributes of the new binary context are of the form "$y.I_k$" and are given as:

$$x[y.I_k] = \begin{cases} 1 & \text{if } x[y] \in I_k \\ 0 & \text{otherwise.} \end{cases}$$

Note that in the two above cases only particular subsets (singletons or definite intervals) are allowed.

An example of a many-valued context is illustrated in Table 1, while Table 2 gives the corresponding binary mapping.

Table 1: An example of a many-valued context.

Obj\Attr	Age	Marital Status	Native Country	Work Class
Alice	23	single	USA	never-worked
Boris	24	married	USA	private
Cyril	31	married	USA	fed-gov
David	60	divorced	Haiti	self-emp

Table 2: An example of a many-valued context mapping.

Obj\Attr	Age.[20,29]	Age.[30,39]	Age.[40,49]	...
Alice	1	0	0	...
Boris	1	0	0	...
Cyril	0	1	0	...
David	0	0	0	...

ISBN: 978-989-95079-6-8

2.4 Crisp vs. Fuzzy Context Representation

In [10], the authors point out a dilemma between support and confidence measures caused by crisp partitions. On the one hand, if the number of intervals for a quantitative attribute is important, the support for any single interval may be low. Hence, without using larger intervals, some rules involving this attribute may not be found because they lack minimal support. On the other hand, if the intervals are too large, some rules may not have minimal confidence. Indeed, when interval becomes larger, the support of the antecedent A of a rule $A \Rightarrow B$ becomes higher. Consequently, the confidence of the rule i.e. the ratio $\texttt{Supp}\left(A \Rightarrow B\right) / \texttt{Supp}\left(A\right)$ decreases and the rule may not have minimal confidence.

Besides, in [11, 12] other authors point out some undesirable threshold effects caused by crisp partitions. Such effects are well-known, for instance, from histograms in statistics: A slight variation of the boundary points of the intervals can have a noticeable effect on the histogram induced by a number of observations. Likewise, the variation of a partition can strongly influence the evaluation of association rules.

Considering the above dilemma and the undesirable threshold effects, fuzzy partitions (in fuzzy intervals) have been widely used instead of crisp partitions (crisp intervals). This has led to fuzzy association rules [13, 14, 15] and to fuzzy concepts analysis [16, 17]. For example, Table 3 illustrates a fuzzy representation for the many-valued context given in Table 1, when choosing the fuzzy partition $\mathcal{F}_1$ (see Fig. 1).

Table 3: Fuzzy representation of a many-valued context.

Obj\Attr	Age.Very Young	Age.Young	Age. Middle Age	...
Alice	0.4	0.6	0	...
Boris	0.2	0.8	0	...
Cyril	0	1	0	...
David	0	0	0	...

3 Information Loss in Quantitative Context

There are two distinct tendencies to transform a many-valued context into a fuzzy partitioned one. The former is based on human expertise, or is even given empirically. The latter uses dedicated heuristics (e.g. [16]). These partitioning methods induce important information loss as it will be showed in this section. Let us before, recall the definition of a fuzzy partition.

Definition1.
A fuzzy partition of a set X on a universe U is a family $\mathcal{F}\left(X\right) = \left(F_1, ..., F_n\right)$ such that:

1. $\forall i = 1..n \quad F_i \neq \emptyset \wedge F_i \neq U$
2. $\forall u \in U, \quad \sum_i F_i\left(u\right) = 1$

3.1 Empirical Fuzzy Partitions

Classical fuzzy set theory has a powerful tool to manage granularity, namely linguistic variables, introduced by Zadeh [18]. Fuzzy partitions are generally based on those linguistic variables. However the complete specification of a linguistic variable may obviously differ from a user to another. For example, in a survey of the literature [11, 12, 14, 15] related to fuzzy association rules discovery, one may find different fuzzy representations of the attribute *Age*. These partitions $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$, $\mathcal{F}_4$ are respectively showed in Fig. 1, Fig. 2, Fig. 3, Fig. 4.

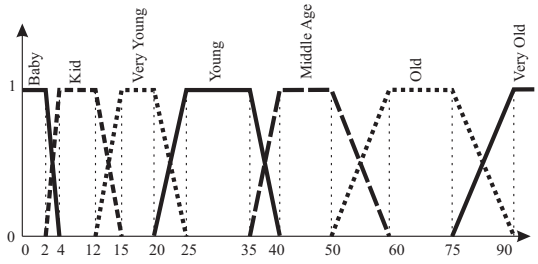


Figure 1: Fuzzy partition $\mathcal{F}_1$ given in [14].

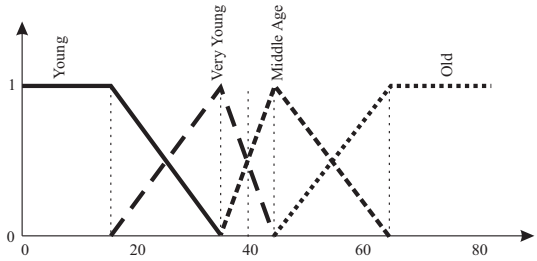


Figure 2: Fuzzy partition $\mathcal{F}_2$ given in [15].

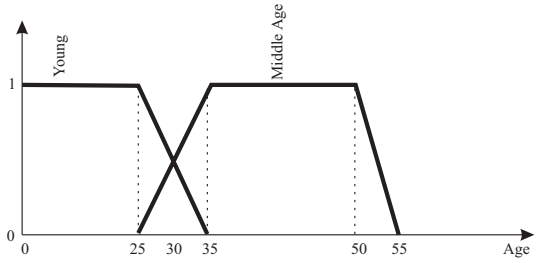


Figure 3: Fuzzy partition $\mathcal{F}_3$ given in [11].

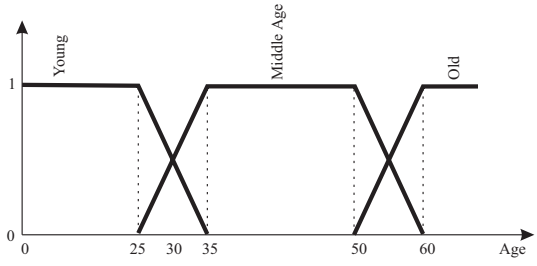


Figure 4: Fuzzy partition $\mathcal{F}_4$ given in [12].

In order to highlight the information loss that may arise with a given fuzzy context compared to another one, we induce all the fuzzy association rules w.r.t a fixed support threshold ξ. The *Adult* benchmark database (UCI Machine Learning Repository) available at http://www.ics.uciedu/mlearn/MLRepositery is taken as a target context for our experiments. Note that, we have restricted the *Adult* context up to six categorical attributes (i.e. *Marital_Status*, *Native_Country*, *Work_Class*, *Salary_Class*, *Education*, *Occupation*) and one quantitative attribute (i.e. *Age*). According to

ISBN: 978-989-95079-6-8

the fuzzy partitions represented respectively in Fig. 1, Fig. 2, Fig. 3, Fig. 4 (namely $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$, $\mathcal{F}_4$) we apply the fuzzy association rules mining algorithm that is fully described in [13] for each of the four fuzzy *Adult* contexts.

Note that, the algorithm in [13] uses a co-occurrence semantics [19] for fuzzy rules. However the information loss phenomenon could be observed as well with gradual or certainty semantics [19] .

Thus among the fuzzy rules induced by each partition $\mathcal{F}_1$-$\mathcal{F}_4$, we have chosen the following association rules to illustrate information loss:
$R1$:(Age.young)⇒(MaritalStatus.single)
$R2$:(Age.middle)⇒(WorkClass.private)
$R3$:(Age.young)⇒(WorkClass.private)
$R4$:(Age.middle)∧(WorkClass.private)⇒(NativeCountry.USA)

By analyzing the results, illustrated in Table 4, it may be remarked that:

1. Different support and confidence measures are found. This confers an uncertainty to the final decision process.

2. More important is a clear information loss materialized by the non discovery of many rules. For example, the rule $R1$ does not exist in the fuzzy context $\mathcal{F}_1$ even though this fuzzy context contains the fuzzy label *young*.

Besides, defining such fuzzy partitions may not be intuitive for humans. This is especially true for attributes where there does not exist a more or less standard linguistic vocabulary which fuzzily partition the attribute domain.

Table 4: Information loss for expert fuzzy partitioning.

	$R1$ Sup/Conf	$R2$ Sup/Conf	$R3$ Sup/Conf	$R4$ Sup/Conf
$\mathcal{F}_1$	$\not\models$	0.26/0.72	$\not\models$	0.13/0.90
$\mathcal{F}_2$	0.08/0.77	0.17/0.68	0.31/0.75	0.08/0.88
$\mathcal{F}_3$	0.05/0.98	$\not\models$	$\not\models$	0.20/0.92
$\mathcal{F}_4$	0.13/0.59	0.47/0.77	$\not\models$	0.30/0.94

where the mark $\not\models$ indicates that the rule could not be discovered since it lacks support.

3.2 Heuristic-Based Contexts Mapping

The second family of approaches deals with a particular generalization of many-valued contexts based on some heuristics depending on the nature of the context. In [16] a "single-pivot" representation is proposed. The so-called pivot is taken as the maximum value in the attribute domain. The equivalent fuzzy context is obtained by dividing the quantitative values by the maximal value of the domain, thus achieving the estimation of a fuzzy proximity to the maximum value. For example, the equivalent fuzzy representation of the attribute *Age* given in Table 1 will be: {(Alice,0.38), (Boris,0.4), (Cyril,0.51), (David,1)}. One may remark that some non uniform data distribution (e.g. with few values near the pivot and almost all values far from it) will lead to an unavoidable information loss for any knowledge discovery process.

As a summary, it appears that empirical, or heuristic-based fuzzy partitioning for data mining applications may be not well-suited. Indeed, in [20] two important stages (i.e. *Processing* and *Transformation*) are intended to get a well-suited representation for the next data mining stage. This is why, we propose a fuzzy learning stage.

4 Learning Fuzzy Partitions

The Fuzzy C-Means algorithm [21, 22] is one of the strongest and widely used fuzzy clustering method. It is based on an iterative optimization of a fuzzy objective function. This function characterizes the partition and is given as:

$$J_m(X,U,V) = \sum_{j=1}^{n}\sum_{k=1}^{c} \mu_{jk}^m \left\|x_j - v_k\right\|^2 \tag{3}$$

where $X = \{x_1, ..., x_n\}$ represents the domain to partition, $V = \{v_1, ..., v_c\}$ represents the set of cluster centers and $U = (\mu_{ki})_{n\times c}$ is the fuzzy partition matrix composed of the membership degree of each item x_j in each cluster k. The exponent m is a parameter called *fuzzifier*.

The distance measure usually used is the Euclidian distance between a datum and a prototype (cluster center). Each cluster V_k is determined as:

$$V_k = \frac{\sum_{j=1}^{n}(\mu_{jk})^m x_j}{\sum_{j=1}^{n}(\mu_{jk})^m} \tag{4}$$

Whereas the elements μ_{jk} of the fuzzy partition matrix are given as:

$$\begin{cases} \mu_{jk} = 0 & \text{if } J_j \neq \emptyset \wedge k \notin J_j \\ \sum_{k\in J_j}\mu_{jk} = 1 & \text{if } J_j \neq \emptyset \wedge k \in J_j \\ \mu_{jk} = \dfrac{1}{\sum_{l=1}^{c}\left(\frac{\|x_j - v_k\|^2}{\|x_j - v_l\|^2}\right)^{\frac{1}{m-1}}} & \text{if } J_j = \emptyset \end{cases} \tag{5}$$

with $J_j = \{k \mid 1 \leqslant k \leqslant c, \|x_j - v_k\| = 0\}$

4.1 Algorithm

The proposed algorithm integrates cluster validity measures [23] in order to obtain an optimal fuzzy partition. It takes as input a minimal (resp. maximal) number of fuzzy partitions allowed, denoted $Cmin$ (resp. $Cmax$). It takes also a fixed threshold tolerance denoted θ. The vector V contains the cluster centers and it is randomly initialized. The algorithm is given as follows:

Begin.
1: $Cmin := 2$; $Cmax := 10$;
2: $\theta := 0.001$;
3: **For** l:=$Cmin$ To $Cmax$ **Do**
4: Initialize $V^{(l)}$ randomly;
5: Compute the partition matrix $U^{(l)}$ with expression (5);
6: Compute the objective function J with (3);
7: Update the cluster centers using (4);
8: Compute the new partition matrix with (5);
9: Compute the new objective function J_{new};
10: If $(J - J_{new}) < \theta$ goto 3 else goto 7;
11: **End For**
12: Determine l s.t. $Vpc(l) = \min\left(Vpc(k)\right)_{k=Cmin,Cmax}$
13: Return $V^{(l)}$, $U^{(l)}$

ISBN: 978-989-95079-6-8

End.

Definition2.
The equivalent fuzzy representation of a quantitative attribute y corresponds to the pair $(\mathcal{F}_y, U_y)$ where:

$$\mathcal{F}_y = \{y.V_1, .., y.V_k, .., y.V_c\}$$
$$U_y\,(x_i, y.V_k) = \mu_{ik}$$

This definition means also that we have chosen to label each fuzzy cluster k by its prototype (cluster center) V_k. Considering that membership degrees to these fuzzy clusters are determined using a distance measure (i.e. Euclidian distance), the semantics attached to the discovered fuzzy association rules $Y.V_k(\overrightarrow{y}) \Rightarrow Z.V_l(\overrightarrow{z})$ is of the type: *the presence of $\overrightarrow{y}$'s near to the prototype $Y.V_k$, imply the presence of $\overrightarrow{z}$'s near to the prototype $Z.V_l$.*

4.2 Cluster validity measures

Our algorithm is implemented considering two options:
i) the end-user may specify the number c of partitions, and even the center of classes (i.e. V_1, ..., V_c),
ii) the learning algorithm determines itself the optimal number of classes using the *Vpc* validity measure as done in the above algorithm.

The partition coefficient (*Vpc*) [23] measure has been chosen since this measure is based on the subsethood degrees μ_{jk}. It is given as:

$$Vpc = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{c}\mu_{jk}^2 \qquad (6)$$

5 Experimental Results

All experiments were performed on a 3.0 Ghz Pentium-IV PC with 512 MB main memory running on Microsoft Windows XP. Algorithms were implemented in C++.

The fuzzy association rules mining algorithm is an A-PRIORI like implementation. Cardinalities of fuzzy sets are computed using a $\Sigma-$count operator. While conjunction in expressions of support and confidence, resp. (1) and (2), is performed using the *min* operator. We have also considered weighted $\alpha-cuts$ for computing support and confidence measures which avoids accumulation of small (irrelevant) cardinalities. The reader is referred to [13] for complete details.

Experiments were performed on the *Adult* database (see section §3.1.). Many objects in the *Adult* context contain missing (null) values. Such objects have been deleted from the initial context which reduces the size of the considered context to 32561 objects. The probability distribution for the *Age* attribute is represented in Fig. 5.

Our fuzzy partition learning process for the *Age* attribute yields two fuzzy clusters and induces the partition:

$$\mathcal{F}_{age} = \{age.25, age.43\}$$

where 25 and 43 correspond to the prototypes (cluster centers). The complete partition is showed in Fig. 6.

According to the partition $\mathcal{F}_{age}$, we show below some new fuzzy association rules that are now discovered.

(Age.25)∧(Workclass.*Private*)⇒(NativeCountry.*USA*)
[support=0.32;confidence=0.89]

(Age.25)∧(Workclass.*Private*)∧(NativeCountry.*USA*)⇒
(Salary_Class.*LeastThan50K$*) [support=0.31;confidence=1]

(Age.43)⇒(Workclass.*Private*)
[support=0.24;confidence=0.69]

(Age.25)⇒(Workclass.*Private*)
[support=0.35;confidence=0.94]

(Age.43)∧(Workclass.*Private*)⇒(NativeCountry.*USA*)
[support=0.32;confidence=0.89]

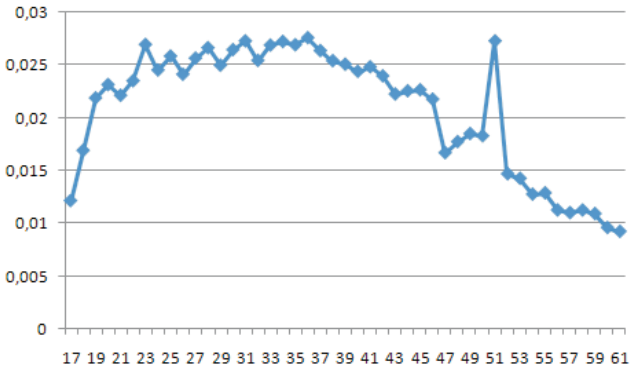


Figure 5: Probability distribution for *Age* attribute in *Adult* context.

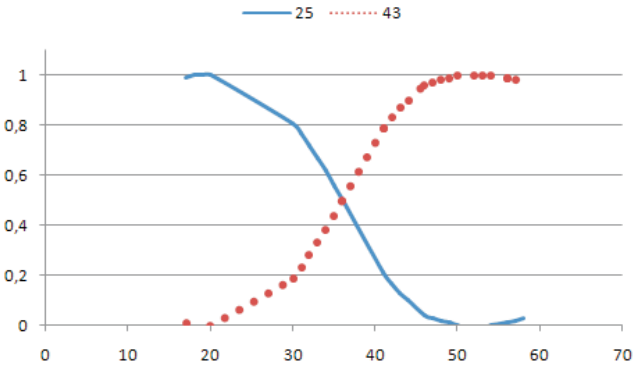


Figure 6: The learned fuzzy partition for *Age* attribute in *Adult* context.

It has been checked that, related to the prototypes, almost all rules are now discovered. For instance, the fourth above rule is a counterpart of the rule $R3$ ((Age.young)⇒(WorkClass.private)), which was not found using the three partitions $\mathcal{F}_1$, $\mathcal{F}_3$, $\mathcal{F}_4$.

6 Conclusion and Perspectives

This paper deals with the transformation stage in a knowledge discovery process. An automatic approach is proposed for the mapping of a many-valued context into a fuzzy context. This mapping is achieved using a fuzzy clustering method. The difficulties of humans for determining fuzzy partitions and fuzzy grades are avoided, which makes things easier for end-users in real life applications. The Fuzzy C-Means algorithm has been

ISBN: 978-989-95079-6-8

chosen as a strong means to get optimized fuzzy partitions. Besides experimental results show that the proposed approach avoids information loss, which departs from the results of almost all existing approaches.

As a first perspective, we intend to enlarge the fuzzy partitions setting by allowing a hierarchical clusters representation. Similar ideas were already explored at least in two different ways. Wolff in [24] proposes a so-called conceptual scale theory which handles many-valued (quantitative) attributes in a fuzzy setting without being restricted to particular subsets (singletons or definite intervals). Bosc et al. in [25] have outlined a knowledge extraction method based on the idea on the fuzzy summarization of data, which may also be used in order to cluster data in a hierarchical-like way.

As pointed out in [26] fuzzy association rules may be also of interest for their increased expressivity. That's why, we intend secondly to investigate other semantics than co-occurrence, namely the one of gradual and certainty rules.

Lastly, it is worth of interest to enlarge fuzzy contexts to incomplete information fuzzy contexts (i.e. with unknown values) [27],[28].

Acknowledgment

The first author is grateful to the A.U.F. (Agence Universitaire de la Francophonie at Brussels-Belgium) who has supported this work under a mobility grant.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining associations rules between sets of items in large databases. *Proc. of the ACM SIGMOD Conf on Management of Data. Washington D.C.*, pages 207–216, 1993.

[2] Rudolf Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. *Rival. I. (Ed): Ordered Sets. Reidel, Dordrecht. Boston*, pages 445–470, 1982.

[3] M. Kaya and R. Alhajj. Genetic algorithm based framework for mining fuzzy association rules. *Fuzzy Sets and Systems*, 152(3):587–601, 2005.

[4] J. Alcala-Fdez, R. Alcala, M.J. Gacto, and F. Herrera. Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms. *Fuzzy Sets and Systems*, 160(7):905–921, 2009.

[5] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. *Proc* 20^{th} *International Conf Very Large Databases. Santiago, Chile*, pages 478–499, 1994.

[6] M.J. Zaki and H.J. Hsiao. Efficient algorithms for mining closed itemsets and their lattice structure. *IEEE Transactions on Knowledge and Data Enginnering*, 17(4):462–478, 2005.

[7] V. Duquenne and J.L. Guigues. Famille minimale d'implications informatives resultant d'un tableau de donnes binaires. *Mathmatiques et sciences humaines*, 24(95):5–18, 1986.

[8] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Efficient mining of association rules using closed itemsets lattices. *Information Systems*, 24:25–46, 1999.

[9] B. Ganter and R. Wille. *Formal Concept Analysis*. Springer Mathematical Foundations, edn, 1999.

[10] R. Srikant and R. Agrawal. Mining quantitative association rules in large relational tables. *Proc. of the ACM SIGMOD Conf on Management of Data. Montreal Canada*, pages 1–12.

[11] D. Dubois, H. Prade, and T. Sudkamp. On the representation, measurement and discovery of fuzzy associations. *IEEE Transactions on Fuzzy Systems*, 13(2):250–262, 2005.

[12] T. Sudkamp. Examples, counterexamples, and measuring fuzzy associations. *Fuzzy Sets and Systems*, 149:57–71, 2005.

[13] Y. Djouadi., S. Redaoui, and K. Amroun. Mining association rules under imprecision and vagueness: towards a possibilistic approach. *FUZZ'IEEE 07, International Conference on Fuzzy Systems. London, UK*, pages 722–727, July 2007.

[14] M. Delgado, N. Marin, D. Sanchez, and M.A. Vila. Fuzzy association rules : general model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–255, 2005.

[15] W.H. Au and K.C.C. Chan. Mining fuzzy association rules in a bank-account databases. *IEEE Transactions on Fuzzy Systems*, 11(2):238–248, 2003.

[16] N. Messai, M.D. Devignes, A. Napoli, and M.S. Tabbone. Many-valued concept lattices for conceptual clustering and information retrieval. *ECAI'08,* 18^{th} *European Conference on Artificial Intelligence. Patras, Greece*, pages 722–727, July.

[17] R. Belohlavek and V. Vychodil. What is a fuzzy concept lattice. *CLA 05, Proc. Olomounc. Czech Republic*, pages 34–45, 2005.

[18] L.A. Zadeh. The concept of a linguistic and its application to approximate reasoning. *Information Science*, 8:199–249, 1975.

[19] M. Serrurier, D. Dubois, H. Prade, and T. Sudkamp. Learning fuzzy rules with their implication operators. *Data and Knowledge Engineering*, 60:71–89, 2007.

[20] U. Fayyad, G.P. Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *Artificial Intelligence Magazine*, pages 37–54, 1996.

[21] J.C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters. *Journal of Cybernetics*, 3:35–57, 1973.

[22] J.C. Bezdek, J. Keller, R. Krishnapuram, and N.R. Pal. *Fuzzy models and algorithms for pattern recognition and image processing*. Kluwer Academic Publisher, 1999.

[23] J.C. Bezdek and N.R. Pal. On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, 3(3):370–379, 1995.

[24] K.E. Wolff. Concepts in fuzzy scaling theory: order and granularity. *Fuzzy Sets and Systems*, 132:63–75, 2002.

[25] P. Bosc, D. Dubois, O. Pivert, H.Prade, and M. De Calmes. Fuzzy summarization of data using fuzzy cardinalities. *Proc. of the 9th Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 2002), Annecy, France*, pages 1553–1559, 2002.

[26] M. Serrurier and H. Prade. Improving expressivity of inductive logic programming by learning different kinds of fuzzy rules. *Soft Computing*, 11(5):459–466, 2007.

[27] P. Burmeister and R. Holzer. Treating incomplete knowledge in formal concepts analysis. *Lecture Notes in Computer Science 3626, Bernhard Ganter (Eds.)*, pages 114–126, 2005.

[28] Didier Dubois, Florence Dupin de Saint Cyr Bannay, and Henri Prade. A possibilty-theoretic view of formal concept analysis. *Fundamenta Informaticae*, 75(1-4):195–213, 2007.

ISBN: 978-989-95079-6-8

On possibilistic sequencing problems with fuzzy parameters

Adam Kasperski[1] Paweł Zieliński[2]

1.Institute of Industrial Engineering and Management, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370, Wrocław, Poland
2.Institute of Mathematics and Computer Science, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370, Wrocław, Poland
Email: adam.kasperski@pwr.wroc.pl, pawel.zielinski@pwr.wroc.pl

Abstract— *In this paper a wide class of sequencing problems with imprecise parameters is discussed. The imprecision is modeled by using closed intervals and fuzzy intervals, whose membership functions are regarded as possibility distributions for the values of unknown parameters. A possibilistic interpretation of fuzzy problems is provided, some solution concepts are proposed and some algorithms for computing the solutions are designed.*

Keywords— Sequencing, min-max regret, possibility theory, fuzzy interval, fuzzy optimization and design

1 Introduction

In a sequencing problem we wish to find a feasible order of elements, called jobs, to achieve some goal. This goal typically depends on job completion times and may also depend on some other job parameters such as due dates or weights. There are a lot of deterministic sequencing problems with different computational properties and a comprehensive description of them can be found for instance in [1]. Unfortunately, most of sequencing problems turned out to be NP-hard but there are also some important problems for which efficient polynomial algorithms exist.

Sequencing problems involve many parameters whose exact values are often unknown. For instance, a job processing time, which is a crucial parameter in all sequencing problems, is rarely precisely known. In order to model such a situation a stochastic approach can be applied (see e.g. [2]), where for every unknown parameter a probability distribution is specified and, typically, the expected cost of a solution is minimized. The stochastic approach has several drawbacks. Namely, it may be hard or expensive to estimate the probability distributions for the parameters. Also, minimizing the expected performance may be not reasonable if the obtained solution is used only once.

In practice, decision makers are often interested in minimizing the cost of a decision in the worst case, that is under the worst realization of parameters that may occur. In recent years a *robust* approach do discrete optimization has attracted a considerable attention. In this approach we specify a *scenario set* containing all possible realizations of parameters which may occur. No probability distribution over the scenario set is given. One of methods of determining the scenario set consists of specifying an interval of possible values for every unknown parameter. A scenario set is then the Cartesian product of all the uncertainty intervals. A natural criterion for choosing a solution under imprecision is the *maximal regret*, which expresses the maximum distance of a solution from optimality over all scenarios [3]. A deeper discussion on using the maximal regret criterion in decision making under uncertainty can be found in [4]. The class of min-max regret sequencing problem with interval parameters has been discussed in a number of papers, for instance in [5, 6, 7, 8, 9, 10, 11, 12]. We briefly describe the known results in this area later in this paper.

In recent years, theory of fuzzy sets has been applied to model the imprecision in optimization, in particular in sequencing problems. A good review of different concepts in fuzzy optimization can be found in [13]. In papers [14, 15, 16] some single machine sequencing problems with fuzzy processing times, fuzzy due dates and fuzzy precedence constraints have been discussed. In these papers a fuzzy due date expresses a degree of satisfaction with job completion time and a sequence is computed, which maximizes the minimum satisfaction or the sum of satisfactions over all jobs. In [17] an alternative approach has been proposed, where the possibility or necessity of job delays is minimized. Also, in [18] an optimality evaluation of sequences under fuzzy parameters has been investigated. We recall and extend this approach here.

In this paper we wish to propose a new approach to sequencing problems with fuzzy parameters. We generalize the min-max regret problems with interval parameters to the fuzzy case by extending the interval uncertainty representation to the fuzzy interval one. The fuzzy parameters induce a possibility distribution over the scenario set, which becomes then reacher in information than the classical scenario set. We provide a possibilistic interpretation of the obtained fuzzy problem and describe a solution concept, which is an adaptation of the elegant solution method proposed in [19] for fuzzy linear programming. This solution concept has been recently adopted in [20] for a class of combinatorial optimization problems. Apart from showing a general framework, we also point out some difficulties which arise when one try to solve a particular problem. Contrary to the class of combinatorial optimization problems described in [20], the sequencing problems are typically more complex to solve and there are very few general properties that are valid for all problems. However, as for the problems described in [20], the main computational difficulties are in the classical interval case and the algorithms known for the interval uncertainty representation can be generalized to the fuzzy case.

This paper is organized as follows. In Section 2 we recall the definition of the classical deterministic sequencing problem. In Section 3 we present the min-max regret approach to sequencing problems with interval parameters - we provide a

ISBN: 978-989-95079-6-8

general formulation and we recall some known results in this area. Finally, in Section 4 we introduce the class of sequencing problems with fuzzy parameters. We give a link between the fuzzy problems and the min-max regret ones. We construct some general methods of solving a fuzzy problem and we illustrate them using a sample problem.

2 Deterministic sequencing problems

We are given a set of *jobs* $J = \{J_1, \ldots, J_n\}$, which may be partially ordered by some *precedence constraints* of the form $i \rightarrow j$, where $i, j \in J$. For the simplicity of notations we will identify every job $J_i \in J$ with its index $i \in \{1, \ldots, n\}$. A solution is a *sequence* (permutation) $\sigma = (\sigma(1), \ldots, \sigma(n))$ of J and it represents an order in which the jobs are processed. A sequence σ is *feasible* if $i \rightarrow j$, $i \neq j$, implies that job j appears after i in σ. We will denote by $\boldsymbol{\mathcal{S}}$ the set of all feasible sequences. We use $C_i(\sigma)$ to denote the *completion time* of job i in sequence σ. In a *single machine* case, a processing time p_i is given for every job $i \in J$ and if $i = \sigma(k)$, then $C_i(\sigma) = \sum_{j=1}^{k} p_{\sigma(j)}$. If every job must be processed on $m > 1$ machines, first on machine 1 next on machine 2 and so on, then we get the *permutation flow shop problem*. In this case p_{ij} is a processing time of job i on machine j and $C_i(\sigma)$ is the time when job i is finished on the m-th machine (see Figure 1).

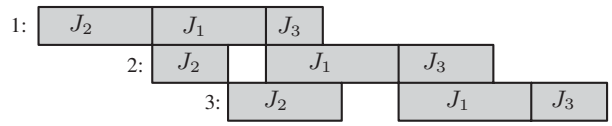


Figure 1: A permutation flow show problem with $m = 3$, $n = 3$ and the schedule that corresponds to the sequence $(2, 1, 3)$.

For every job $i \in J$, there is a cost function $f_i : R \rightarrow R$, which measures the cost of completing i at time t. The value of this function may depend on some other parameters associated with job i such as *due date* d_i or *weight* w_i. Finally, $F(\sigma)$ denotes a cost of sequence σ. There are two general types of the cost function $F(\sigma)$, namely a *bottleneck objective* $F(\sigma) = \max_{i=1,\ldots,n} f_i(C_i(\sigma))$ and a *sum objective* $F(\sigma) = \sum_{i=1}^{n} f_i(C_i(\sigma))$.

In a deterministic sequencing problem we seek a feasible sequence of the minimal cost, that is we solve the following optimization problem:

$$\min_{\sigma \in \boldsymbol{\mathcal{S}}} F(\sigma). \tag{1}$$

Our analysis in the next sections of this paper will be based on the concept of a *deviation*. In the deterministic case the deviation of a sequence σ expresses its distance to optimum and is defined as follows:

$$\delta_\sigma = F(\sigma) - \min_{\rho \in \boldsymbol{\mathcal{S}}} F(\rho). \tag{2}$$

Obviously, σ is optimal if and only if $\delta_\sigma = 0$. If the deviation is positive, then it indicates a distance of σ from optimality.

Sequencing problems are usually described by using a convenient Graham's notation. So, every sequencing problem can be denoted by a triple $\alpha|\beta|\gamma$, where α is the machine environment ($\alpha = 1$ for the single machine case), β specifies the job characteristic and γ describes the objective function. The following examples of sequencing problems are well known and will be explored later in this paper:

- $1|prec|L_{max}$. In this problem we seek a sequence, which minimizes the maximum *lateness*. So, the bottleneck objective is $\max_{i \in J}\{C_i(\sigma) - d_i\}$. We can generalize this problem by minimizing the maximal weighted lateness or modify it to minimize the maximal weighted *tardiness* (see [1]).

- $1||\sum C_i$. In this problem there are no precedence constraints between jobs and we seek a sequence for which the sum of completion times of all jobs, i.e. the *total flow time* $\sum_{i \in J} C_i(\sigma)$, is minimal.

- $1|p_i = 1|\sum w_i U_i$. In this problem there are no precedence constraints between jobs and all jobs have unit processing times. A job is *late* in σ if $C_i(\sigma) > d_i$. In this case we write $U_i(\sigma) = 1$ and $U_i(\sigma) = 0$ otherwise. The cost of σ is $\sum_{i \in J} w_i U_i(\sigma)$, so it expresses the weighted number of late jobs.

- $Fm||C_{max}$. This is the permutation flow shop problem with $m > 1$ machines. There are no precedence constraints between jobs and the bottleneck cost of σ is the *makespan* $\max_{i \in J} C_i(\sigma)$, that is the completion time of the last job on the last machine. The problem is polynomially solvable only if $m = 2$ and becomes NP-hard for $m \geq 3$ (see [1]).

The above examples illustrate a large variety of basic sequencing models. As we will see in the next section, they have quite different computational properties under imprecision and the state of the art in this area is still far from being complete.

3 Minmax regret sequencing problems

In practice, the exact values of parameters in a sequencing problem such as processing times, due dates or weights may be not precisely known. Assume that we have l parameters and the value of a parameter ξ_i, $i = 1, \ldots, l$, may fall within a closed interval $[\underline{\xi}_i, \overline{\xi}_i]$ independently of the values of the other parameters. A parameter ξ_i is *precise* if $\underline{\xi}_i = \overline{\xi}_i$. Every vector $S = (s_1, \ldots, s_l) \in R^l$ such that $s_i \in [\underline{\xi}_i, \overline{\xi}_i]$ is called a *scenario* and it expresses a possible state of the world, where $\xi_i = s_i$ for $i = 1, \ldots, l$. A scenario is called *extreme* if all parameters take the lower or upper bounds in their uncertainty intervals. We use Γ to denote the set of all possible scenarios. Hence Γ is the Cartesian product of all the uncertainty intervals. Now the cost of a sequence σ depends on scenario $S \in \Gamma$ and we will denote it as $F(\sigma, S)$. We will also denote by $F^*(S)$ the cost of an optimal sequence under scenario S. In order to obtain the value of $F^*(S)$ we need to solve problem (1) under the fixed realization of parameters S. It is clear that deviation of σ also depends on scenario S and we will denote it as $\delta_\sigma(S) = F(\sigma, S) - F^*(S)$.

Now the optimality of a sequence σ can be characterized by a *deviation interval* $[\underline{\delta}_\sigma, \overline{\delta}_\sigma]$, where $\underline{\delta}_\sigma = \min_{S \in \Gamma} \delta_\sigma(S)$ is the minimal deviation and $\overline{\delta}_\sigma = \max_{S \in \Gamma} \delta_\sigma(S)$ is the maximal deviation over all scenarios. The quantity $\overline{\delta}_\sigma$ is called in

ISBN: 978-989-95079-6-8

literature the *maximal regret* of σ and it expresses the largest distance of σ from optimum over the scenario set Γ. A scenario S_σ, for which the deviation of σ attains maximum, is called a *worst case scenario* for σ. So, under the interval uncertainty representation, we only know that $\delta_\sigma \in [\underline{\delta}_\sigma, \overline{\delta}_\sigma]$ and we can give the following characterization of a feasible sequence: a sequence σ is *possibly optimal* if $\underline{\delta}_\sigma = 0$ and it is *necessarily optimal* if $\overline{\delta}_\sigma = 0$. Notice that a sequence is possibly optimal if and only if it is optimal under some scenario $S \in \Gamma$ and it is necessarily optimal if and only if it is optimal for all scenarios $S \in \Gamma$.

Now the question arises which sequence of $\boldsymbol{\mathcal{S}}$ should be chosen. A necessarily optimal one is a natural choice. However, it rarely exists in most practical situations, because the necessary optimality is very strong criterion. In a more reasonable approach we can minimize the the maximal regret, that is the quantity $\overline{\delta}_\sigma$, which may be viewed as a distance to the necessary optimality. As a result we get the following *min-max regret sequencing problem*:

$$\min_{\sigma \in \boldsymbol{\mathcal{S}}} \overline{\delta}_\sigma. \tag{3}$$

We call an optimal solution to (3) a *min-max regret sequence*. In the next section we briefly review some known facts on problem (3).

3.1 Complexity of min-max regret sequencing problems

The first problem that arises while analyzing a particular min-max regret sequencing problem is the computation of the maximal regret for a given sequence σ, that is the quantity $\overline{\delta}_\sigma$. Unfortunately, contrary to the class of problems discussed in [20], there is no general method of performing this task. For the problems considered in [20], it is possible to find two extreme scenarios that maximize and minimize the deviation and, consequently, the computation of the maximal regret has the same complexity as the deterministic problem. For sequencing problems the situation is much more complex. First of all, for some problems there may be no extreme scenario that maximizes (minimizes) the deviation [21]. Furthermore, computing the maximal regret may be much more time consuming than solving a deterministic problem. For instance, in the min-max regret $1||\sum C_i$ problem with interval processing times, computing $\overline{\delta}_\sigma$ requires solving an assignment problem while an optimal sequence under a given scenario can be computed in $O(n \log n)$ time. An extreme case has been described in [9], where a permutation flow shop problem with m machines and with interval processing times and with only 2 jobs has been discussed. Since the solution set contains only two sequences, the total computational effort is focused on computing the maximal regret of a given sequence.

It is not surprising that min-max regret sequencing problems are typically hard to solve. There are only several problems that are known to be polynomially solvable. A polynomial algorithm for $1|prec|L_{max}$ with interval processing times and interval due dates has been constructed in [6]. A polynomial algorithm for $1|prec|\max w_i T_i$ with interval weights and precise processing times and precise due dates has been proposed in [8]. Apart from this two problems, only some very special cases are known to be polynomially solvable, for instance $Fm||C_{max}$ with interval processing times and only two jobs [9] and $1|p_i = 1|\sum w_i U_i$ with interval weights and a precise common due date ($d_1 = d_2 = \cdots = d_n$) [21]. This latter problem is equivalent to the min-max regret version of the selecting items problem, which is known to be polynomially solvable [21, 22].

Among the known negative results, the most important one has been obtained in [10], where it has been shown that the min-max regret $1||\sum C_i$ problem with interval processing times is NP-hard. This problem is also known to be approximable within 2 [12] and can be solved by using a mixed integer programming formulation proposed in [11]. However, the complexity status of a number of basic problems is still unknown. We do not know whether $F2||C_{max}$ with interval processing times is NP-hard. We can, however, compute the maximal regret of a given sequence and solve the problem by using a branch and bound algorithm [7, 3]. Also the problem $1|p_i = 1|\sum w_i U_i$ with interval weights and arbitrary precise due dates and the problem $1|prec|\max w_i L_i$ with interval processing times, interval due dates and precise weights are open. The former problem can be solved by a mixed integer programming formulation [21]. There are also a large number of different sequencing problems whose min-max regret versions have never been investigated and they should be the subjects of further research.

4 Fuzzy sequencing problems

In this section we propose an extension of the min-max regret approach to sequencing problems. We will apply fuzzy intervals to model the imprecise parameters and we will use possibility theory to define a solution concept. A comprehensive description of possibility theory can be found in [23].

4.1 Basic notions on possibility theory

A *fuzzy interval* $\tilde{A}$ is a fuzzy set in the space of reals whose membership function $\mu_{\tilde{A}}$ is normal, quasi concave and upper semicontinuous. It is typically assumed that the support of a fuzzy interval is compact. The main property of a fuzzy interval is the fact that all its λ-cuts, that is the sets $\tilde{A}^\lambda = \{x : \mu_{\tilde{A}}(x) \geq \lambda\}$, $\lambda \in (0, 1]$, are closed intervals. We will assume that $\tilde{A}^0$ is the smallest closed set containing the support of $\tilde{A}$. So, every fuzzy interval $\tilde{A}$ can be represented as a family of closed intervals $\tilde{A}^\lambda = [\underline{a}^\lambda, \overline{a}^\lambda]$, parametrized by the value of $\lambda \in [0, 1]$. In many practical applications, the class of *trapezoidal fuzzy intervals* is used. A trapezoidal fuzzy interval, denoted by a quadruple $(\underline{a}, \overline{a}, \alpha, \beta)$, can be represented as the family $[\underline{a} - \alpha(1 - \lambda), \overline{a} + \beta(1 - \lambda)]$ for $\lambda \in [0, 1]$. Notice that this representation contains classical intervals ($\alpha = \beta = 0$) and real numbers (additionally $\underline{a} = \overline{a}$) as special cases.

In this paper we adopt a possibilistic interpretation of a fuzzy interval [23]. Assume that for an imprecise real quantity ξ a fuzzy interval with membership function μ_ξ is given. This membership function expresses a *possibility distribution* for the values of ξ, namely $\Pi(\xi = x) = \mu_\xi(x)$ is the possibility of the event that ξ will take the value of x. It is easily seen that the closed interval $[\underline{\xi}^\lambda, \overline{\xi}^\lambda]$, $\lambda \in [0, 1]$, contains all values of ξ, whose possibility of occurrence is not less than λ. In particular, the interval $[\underline{\xi}^0, \overline{\xi}^0]$ should contain all possible values of ξ, while the interval $[\underline{\xi}^1, \overline{\xi}^1]$ should contain the most plausible ones. Some methods of obtaining the possibility distribution of an unknown quantity can be found in [23].

ISBN: 978-989-95079-6-8

Let $\tilde{G}$ be a fuzzy set in the space of reals with membership function $\mu_{\tilde{G}}$. Then $\xi \in \tilde{G}$ is a *fuzzy event* and the necessity that $\xi \in \tilde{G}$ holds is defined in the following way:

$$\begin{aligned} \mathrm{N}(\xi \in \tilde{G}) &= 1 - \Pi(\xi \notin \tilde{G}) = \\ &= 1 - \sup_{x \in R} \min\{\mu_{\xi}(x), 1 - \mu_{\tilde{G}}(x)\} \end{aligned} \tag{4}$$

where $1 - \mu_{\tilde{G}}(x)$ is the membership function of the complement of the fuzzy set $\tilde{G}$. It is not difficult to see that if $\tilde{G} = (0, \overline{g}, 0, \beta) = (\overline{g}, \beta)$, then the following equality is true:

$$\mathrm{N}(\xi \in \tilde{G}) = 1 - \inf_{\lambda \in [0,1]} \{\overline{\xi}^{\lambda} \leq \overline{g}^{1-\lambda}\} \tag{5}$$

and $\mathrm{N}(\xi \in \tilde{G}) = 0$ if $\overline{\xi}^{1} > \overline{g}^{0}$. Equality (5) is illustrated in Figure 2.

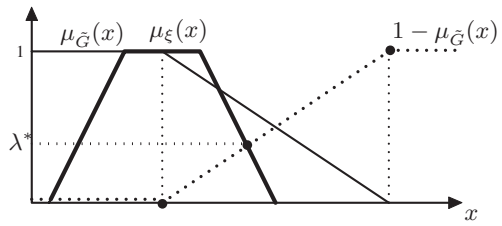


Figure 2: $\mathrm{N}(\xi \in \tilde{G}) = 1 - \lambda^*$.

4.2 Possibilistic sequencing problem

Assume that for every unknown parameter ξ_i, $i = 1, \ldots, l$, in a sequencing problem a fuzzy interval with membership function μ_{ξ_i} is specified. According to the interpretation given in the previous section μ_{ξ_i} is a possibility distribution for the values of ξ_i. Under the assumption that all parameters are unrelated, there is a possibility distribution over all scenarios $S = (s_1, \ldots, s_l) \in R^l$ defined as follows:

$$\begin{aligned} &\pi(S) = \Pi\left(\bigwedge_{i=1}^{l} [\xi_i = s_i]\right) = \min_{i=1,\ldots,l} \Pi(\xi_i = s_i) = \\ &= \min_{i=1,\ldots,l} \mu_{\xi_i}(s_i). \end{aligned} \tag{6}$$

So, the value of $\pi(S)$ express the possibility of the event that scenario S will occur. Notice that we generalize in this way the scenario set $\Gamma \subset R^l$ described in Section 3. Indeed, for the interval uncertainty representation $\pi(S) = 1$ if $S \in \Gamma$ and $\pi(S) = 0$ otherwise. In the fuzzy case $\pi(S)$ may take any value in the interval $[0, 1]$ so fuzzy intervals allow us to model the imprecision in a more sophisticated manner.

In the interval case the value of deviation δ_{σ} falls within a closed interval. Analogously, in the fuzzy case it falls within a fuzzy interval with membership function $\mu_{\delta_{\sigma}}$. Of course, $\mu_{\delta_{\sigma}}$ is a possibility distribution for the values of δ_{σ} and, according to possibility theory, it is defined as follows:

$$\mu_{\delta_{\sigma}}(x) = \Pi(\delta_{\sigma} = x) = \sup_{\{S:\ \delta_{\sigma}(S) = x\}} \pi(S). \tag{7}$$

Recall that the statement "'σ is optimal"' is equivalent to the assertion $\delta_{\sigma} = 0$. Consequently, we can define the degrees of possible and necessary optimality of a given sequence as follows:

$$\Pi(\sigma \text{ is optimal}) = \Pi(\delta_{\sigma} = 0) = \mu_{\delta_{\sigma}}(0), \tag{8}$$

$$\mathrm{N}(\sigma \text{ is optimal}) = \mathrm{N}(\delta_{\sigma} = 0) = 1 - \sup_{x > 0} \mu_{\delta_{\sigma}}(x). \tag{9}$$

As in the interval case, a question arises which sequence should be chosen. In order to perform this task we adopt a concept first applied to fuzzy linear programming in [19]. Assume that a decision maker knows his/her preference about the sequence deviation and expresses it by using a fuzzy interval $\tilde{G} = (\overline{g}, \beta)$. So, the values of deviation in $[0, \overline{g}]$ are fully accepted, the values in $[\overline{g} + \beta, \infty)$ are not at all accepted and the degree of acceptance decreases from 1 to 0 in $[\overline{g}, \overline{g} + \beta]$. Our aim is to compute a feasible sequence $\sigma \in \boldsymbol{\mathcal{S}}$, for which the necessity of the event $\delta_{\sigma} \in \tilde{G}$ is maximal, namely we wish to solve the following optimization problem:

$$\max_{\sigma \in \boldsymbol{\mathcal{S}}} \mathrm{N}(\delta_{\sigma} \in \tilde{G}). \tag{10}$$

This solution concept has been first proposed for linear programming problem with fuzzy objective function coefficients [19]. Observe that (10) can also be expressed as $\min_{\sigma \in \boldsymbol{\mathcal{S}}} \Pi(\delta_{\sigma} \in \tilde{G}^d)$, where $\tilde{G}^d$ is the complement of the fuzzy goal $\tilde{G}$ with membership function $1 - \mu_{\tilde{G}}(x)$. If we fix $\tilde{G} = (0, 0)$ in (10), then we get a special case, in which we seek a feasible sequence that maximizes the degree of necessary optimality:

$$\max_{\sigma \in \boldsymbol{\mathcal{S}}} \mathrm{N}(\delta_{\sigma} = 0) = \max_{\sigma \in \boldsymbol{\mathcal{S}}} \mathrm{N}(\sigma \text{ is optimal}). \tag{11}$$

We focus now on a method of solving the problem (10). According to equality (5), the problem (10) is equivalent to the following optimization problem:

$$\begin{aligned} &\min \lambda \\ &\overline{\delta}_{\sigma}^{\lambda} \leq \overline{g}^{1-\lambda} \\ &\sigma \in \boldsymbol{\mathcal{S}} \\ &\lambda \in [0, 1] \end{aligned} \tag{12}$$

If λ^* is the optimal objective value and σ^* is an optimal solution to (12), then $\mathrm{N}(\delta_{\sigma^*} \in \tilde{G}) = 1 - \lambda^*$. If (12) is infeasible, then $\mathrm{N}(\delta_{\sigma} \in \tilde{G}) = 0$ for all feasible sequences σ. Of course, if we replace expression $\overline{g}^{1-\lambda}$ with 0 in (12), then we get an equivalent formulation of the problem (11).

Let us focus now on the quantity $\overline{\delta}_{\sigma}^{\lambda}$. The closed interval $[\underline{\delta}_{\sigma}^{\lambda}, \overline{\delta}_{\sigma}^{\lambda}]$, $\lambda \in [0, 1]$, contains all values of deviation δ_{σ}, whose possibility of occurrence is not less than λ. So

$$\overline{\delta}_{\sigma}^{\lambda} = \sup_{\{S\,:\,\pi(S) \geq \lambda\}} \{F(\sigma, S) - F^*(S)\} \tag{13}$$

is the greatest deviation of σ, whose possibility of occurrence is not less than λ. From the definition of $\pi(S)$ it is easy to see that $\{S:\ \pi(S) \geq \lambda\} = [\underline{\xi}_1^{\lambda}, \overline{\xi}_1^{\lambda}] \times \cdots \times [\underline{\xi}_l^{\lambda}, \overline{\xi}_l^{\lambda}] = \Gamma^{\lambda}$. Consequently, the quantity $\overline{\delta}_{\sigma}^{\lambda}$ is the maximal regret of σ in the min-max regret version of the problem with scenario set Γ^{λ}. In particular, the condition $\overline{\delta}_{\sigma}^{\lambda} \leq 0$ means that σ is necessarily optimal under Γ^{λ}.

ISBN: 978-989-95079-6-8

Let us point out that the fuzzy problem is not simpler than the corresponding min-max regret one. Indeed, the interval uncertainty representation is a special case of the fuzzy one. If we additionally fix $\tilde{G} = (0, M)$ for a sufficiently large M, then the fuzzy problem is equivalent to the min-max regret one.

Observe that $\overline{\delta}_\sigma^\lambda$ is a nonincreasing and $\overline{g}^{1-\lambda}$ is a nondecreasing function of λ. In consequence, the problem (12) can be solved be a standard binary search technique if we only can decide somehow whether there is a feasible sequence $\sigma \in \boldsymbol{\mathcal{S}}$ fulfilling inequality $\overline{\delta}_\sigma^\lambda \leq \overline{g}^{1-\lambda}$ for a fixed $\lambda \in [0, 1]$. This task is easy if we have an algorithm for the corresponding min-max regret sequencing problem with interval parameters. Solving this problem for scenario set Γ^λ we get a min-max regret sequence σ^*. Then $\overline{\delta}_\sigma^\lambda \leq \overline{g}^{1-\lambda}$ for some $\sigma \in \boldsymbol{\mathcal{S}}$ if and only if $\overline{\delta}_{\sigma^*}^\lambda \leq \overline{g}^{1-\lambda}$. Notice that in the problem (11) it is enough to detect a necessarily optimal sequence for scenario set Γ^λ, which may be computationally easier than solving the min-max regret problem. The binary search algorithm is shown in Figure 3.

1: Find a min-max regret sequence σ under Γ^1
2: **if** $\overline{\delta}_\sigma^1 > \overline{g}^0$ **then return** $\emptyset$
3: $\lambda_1 \leftarrow 0.5$, $k \leftarrow 1$, $\lambda_2 \leftarrow 0$
4: **while** $|\lambda_1 - \lambda_2| \geq \epsilon$ **do**
5: $\quad\lambda_2 \leftarrow \lambda_1$
6: $\quad$Find a min-max regret sequence ρ under Γ^{λ_1}
7: $\quad$**if** $\overline{\delta}_\rho^{\lambda_1} \leq \overline{g}^{1-\lambda_1}$ **then**
8: $\quad\quad\lambda_1 \leftarrow \lambda_1 - 1/2^{k+1}$, $\sigma \leftarrow \rho$
9: $\quad$**else**
10: $\quad\quad\lambda_1 \leftarrow \lambda_1 + 1/2^{k+1}$
11: $\quad$**end if**
12: $\quad k \leftarrow k + 1$
13: **end while**
14: **return** σ

Figure 3: Solving the problem (10) with a given precision $\epsilon \in (0, 1)$. Algorithm returns $\emptyset$ if $\mathrm{N}(\delta_\sigma \in \tilde{G}) = 0$ for all $\sigma \in \boldsymbol{\mathcal{S}}$.

It is easy to check that if the min-max regret problem with interval parameters can be solved in $f(n)$ time, then the corresponding fuzzy problem is solvable in $O(f(n) \log \epsilon^{-1})$ time. W thus can see that the problem with fuzzy parameters boils down to solving a small number of min-max regret sequencing problems. So, every exact algorithm for the min-max regret problem can easily be adopted for the more general fuzzy case. In [6] a polynomial algorithm for the min-max regret $1|prec|L_{max}$ problem with interval processing times and interval due dates has been constructed. Consequently, the fuzzy $1|prec|L_{max}$ problem is solvable efficiently as well because we can use this algorithm as a subroutine in the binary search.

The binary search algorithm is the most general method of solving the fuzzy problem which, however, requires a given precision of calculations. The formulation (12) sometimes allows us to design an exact algorithm based on a mixed integer programming (MIP) formulation. The obtained MIP model can be then solved by using some standard packages such as CPLEX or GLPK. We will show an example in the next section.

4.3 The fuzzy problem $1||\sum C_i$

Assume that for every job $i \in J$, a fuzzy processing time with possibility distribution μ_{p_i} is given. A scenario in this problem is a particular realization of the processing times. The cost of a sequence σ under S is the total flow time, that is the sum of job completion times. Under a given processing times scenario S, an optimal sequence can be obtained in $O(n \log n)$ time by ordering the jobs with respect to nondecreasing processing times [1].

We first show that the problem of computing the most necessarily optimal sequence, i.e. the problem (11), is efficiently solvable. Let us fix $\lambda \in [0, 1]$ and consider the min-max regret $1||\sum C_i$ problem with interval processing times specified by scenario set $\Gamma^\lambda = [\underline{p}_i^\lambda, \overline{p}_i^\lambda] \times \cdots \times [\underline{p}_n^\lambda, \overline{p}_n^\lambda]$. Consider a scenario $S \in \Gamma^\lambda$ such that under this scenario the processing times are $\frac{1}{2}(\underline{p}_i^\lambda + \overline{p}_i^\lambda)$ for $i \in J$. We can compute in $O(n \log n)$ time and optimal sequence ρ under S. It turns out that $\overline{\delta}_\rho^\lambda \leq 2\overline{\delta}_\sigma^\lambda$ for all sequences σ (see [20]). In consequence, if there is a necessarily optimal sequence σ such that $\overline{\delta}_\sigma^\lambda = 0$, then ρ must also be necessarily optimal. We thus have an efficient method of detecting a necessarily optimal sequence and the problem (11) can be solved in $O(n \log n \log \epsilon^{-1})$ time by using the binary search shown in Figure 3.

We now focus on the more general problem (10). Unfortunately, this problem is NP-hard because the corresponding min-max regret problem with interval processing times is NP-hard [10]. Using the formulation (12) we will show that it is possible to design a mixed integer linear programming model to solve the fuzzy problem. Let $\mathcal{A}$ be the set of all binary vectors (x_{ij}), $i = 1, \ldots, n$, $j = 1 \ldots, n$, fulfilling the so called *assignment constraints*, that is $\sum_{i=1}^n x_{ij} = 1$ for all $j = 1, \ldots, n$ and $\sum_{j=1}^n x_{ij} = 1$ for all $i = 1, \ldots, n$. A vector $(x_{ij}) \in \mathcal{A}$ represents a sequence σ in which $x_{ij} = 1$ if job $i \in J$ occupies position j in σ. Obviously, there is one to one correspondence between the sequences of the set of jobs J and the vectors in $\mathcal{A}$ (recall that there are no precedence constraints in J). If $(x_{ij}) \in \mathcal{A}$ corresponds to sequence σ, then the maximal regret of σ under scenario set Γ^λ can be computed in the following way [3, 11]:

$$\overline{\delta}_\sigma^\lambda = \max_{(z_{ij}) \in \mathcal{A}} \sum_{i=1}^n \sum_{j=1}^n c_{ij}^\lambda z_{ij}, \tag{14}$$

where

$$c_{ij}^\lambda = \overline{p}_i^\lambda \sum_{k=1}^{j} (j-k) x_{ik} + \underline{p}_i^\lambda \sum_{k=j+1}^{n} (j-k) x_{ik}. \tag{15}$$

Observe that (14) is an assignment problem. We can construct the dual to (14) and it is well known that this dual has the same optimal objective function value as (14). So, it holds:

$$\begin{aligned} \overline{\delta}_\sigma^\lambda = \quad & \min \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \beta_i \\ & \alpha_i + \beta_j \geq c_{ij}^\lambda \text{ for } i, j = 1, \ldots, n \end{aligned} \tag{16}$$

where α_i and β_i, $i = 1, \ldots, n$, are unrestricted dual variables associated with the assignment constraints. Now, using formulation (12), we can represent the fuzzy problem in the fol-

ISBN: 978-989-95079-6-8

lowing way:

$$\begin{array}{l} \min \lambda \\ \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \beta_i \leq \overline{g}^{1-\lambda} \\ \alpha_i + \beta_j \geq c_{ij}^{\lambda} \text{ for } i,j = 1,\ldots,n \\ (x_{ij}) \in \mathcal{A} \\ \lambda \in [0,1] \end{array} \tag{17}$$

where c_{ij}^{λ} are given as (15). If trapezoidal fuzzy intervals $(\underline{p}_i, \overline{p}_i, \alpha_i, \beta_i)$, $i = 1,\ldots,n$, are used to model the processing times, then we can substitute $\overline{p}_i^{\lambda} = \overline{p}_i + (1-\lambda)\beta_i$ and $\underline{p}_i^{\lambda} = \underline{p}_i - (1-\lambda)\alpha_i$ in (15). The resulting model will be still not linear because some expressions of the form λx_{ij} will appear. However, we can easily linearize the model by substituting $t_{ij} = \lambda x_{ij}$ and adding additional constraints $t_{ij} - x_{ij} \leq 0$, $\lambda - t_{ij} + x_{ij} \leq 0$, $-\lambda + t_{ij} \leq 0$, $t_{ij} \geq 0$ for all $i,j = 1,\ldots,n$. Hence the resulting model is a mixed integer linear one and can be solved by using an available software. We refer the reader to [11] for some techniques that allow us to refine the formulation and speed up the computations.

5 Conclusions

In this paper we have adopted a general approach to sequencing problems with fuzzy parameters. We have applied possibility theory to model the imprecise parameters and to define a solution concept. We have extended the approach used in [19] and [20] to another class of problems. As in many fuzzy problems, the main computational difficultly is in the classical interval case. Namely, every exact algorithm that computes a min-max regret sequence under interval data can be easily adopted to solve the fuzzy problem. One can use a binary search method or try to design a mixed integer programming model.

Unfortunately, the min-max regret sequencing problems with interval data are mostly hard to solve. There are only few problems that are known to be polynomially solvable and the complexity of some important problems, such as the permutation flow shop on two machines, remains open. Contrary to the class of problems discussed in [20], there is also lack of general properties of the min-max regret sequencing problems. Specifically, there is no general method of computing the maximal regret of a given sequence and there are no general relationships between the optimality evaluation and the min-max regret sequences. In other words, every sequencing problem possesses its own properties and the area of research in this field is still not fully explored.

Acknowledgment

The second author was partially supported by Polish Committee for Scientific Research, grant N N111 1464 33.

References

[1] P. Brucker. *Scheduling Algorithms, 5th edition*. Springer Verlag, Hedielberg, 2007.

[2] X. Wu and X. Zhou. Stochastic scheduling to minimize expected maximum lateness. *European Journal of Operational Research*, 190(1):103–115, 2008.

[3] P. Kouvelis and G. Yu. *Robust discrete optimization and its applications*. Kluwer Academic Publishers, Boston, 1997.

[4] R. D. Luce and H. Raiffa. *Games and decisions. Introduction and critical survey*. Dover Publications, INC., New York, 1957.

[5] R.L. Daniels and P. Kouvelis. Robust scheduling to hedge against processing time uncertainty in single-stage production. *Management Science*, 41(2):363–376, 1995.

[6] A. Kasperski. Minimizing maximal regret in the single machine sequencing problem with maximum lateness criterion. *Oper. Res. Let.*, 33(4):431–436, 2005.

[7] P. Kouvelis, R.L. Daniels, and G. Vairaktarakis. Robust scheduling of a two-machine flow shop with uncertain processing times. *IIE Transactions*, 32:421–432, 2000.

[8] I. Averbakh. Minmax regret solutions for minimax optimization problems with uncertainty. *Oper. Res. Let.*, 27(2):57–65, 2001.

[9] I. Averbakh. The minmax regret permutation flow-shop problem with two jobs. *European Journal of Operational Research*, 169:761–766, 2006.

[10] V. Lebedev and I. Averbakh. Complexity of minimizing the total flow time with interval data and minmax regret criterion. *Discrete Applied Mathematics*, 154(15):2167–2177, 2006.

[11] R. Montemanni. A mixed integer programming formulation for the total flow time single machine robust scheduling problem with interval data. *Journal of Mathematical Modelling and Algorithms*, 6(2):287–296, 2007.

[12] A. Kasperski and P. Zieliński. A 2-approximation algorithm for interval data minmax regret sequencing problems with the total flow time criterion. *Oper. Res. Let.*, 36:343–344, 2008.

[13] M. Inuiguchi M. and J. Ramik. Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. *Fuzzy Sets and Systems*, 111(11):3–28, 2000.

[14] M. Vlach. Single machine scheduling under fuzziness. In R. Słowiński and M. Hapke, editors, *Scheduling under Fuzziness*, volume 37 of *Studies in Fuzziness and Soft Computing*, pages 223–245. Physica–Verlag, 2000.

[15] H. Ishii, M. Tada, and T. Masuda. Scheduling problem with fuzzy due dates. *Fuzzy Sets and Systems*, 46:339–347, 1992.

[16] S. Han, H. Ishii, and S. Fuji. One machine scheduling problem with fuzzy due dates. *European Journal of Operational Research*, 79:1–12, 1994.

[17] A. Kasperski. A possibilistic approach to sequencing problems with fuzzy parameters. *Fuzzy Sets and Systems*, 150(1):77–86, 2005.

[18] S. Chanas and A. Kasperski. Possible and necessary optimality of solutions in the single machine scheduling problem with fuzzy parameters. *Fuzzy Sets and Systems*, 142:359–371, 2004.

[19] M. Inuiguchi and M. Sakawa. Robust optimization under softness in a fuzzy linear programming problem. *International Journal of Approximate Reasonning*, 18:21–34, 1998.

[20] A. Kasperski and P. Zieliński. On possibilistic combinatorial optimization problems. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 324–329, Hong Kong, 2008.

[21] A. Kasperski. *Discrete optimization with interval data. Minmax regret and fuzzy approach*, volume 228 of *Studies in Fuzziness and Soft Computing*. Springer, Berlin Heidelberg, 2008.

[22] E. Conde. An improved algorithm for selecting p items with uncertain returns according to the minmax regret criterion. *Mathematical Programming*, 100:345–353, 2004.

[23] D. Dubois and H. Prade. *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, New York, 1988.

ISBN: 978-989-95079-6-8

The Bandler-Kohout Subproduct as a Suitable Inference Mechanism

Martin Štěpnička[1] Balasubramaniam Jayaram[2]

1. Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava 701 33, Czech Republic
2. Department of Mathematics, Indian Institute of Technology Madras
Chennai 600 036, India
Email: martin.stepnicka@osu.cz, jbala@iitm.ac.in,

***Abstract*— *The compositional rule of inference (CRI) introduced by Zadeh is widely used in approximate reasoning schemes using fuzzy sets. In this work we show that the Bandler-Kohout subproduct does possess all the important properties such as equivalent and reasonable conditions for their solvability, their interpolative properties and the preservation of the indistinguishability that may be inherent in the input fuzzy sets, as cited in favor of using CRI.*

***Keywords*—** Bandler-Kohout subproduct, compositional rule of inference, correctness and continuity of inference, fuzzy relation equations.

1 Introduction

Fuzzy systems are one of the best known applications of fuzzy logic. These are usually based on a set of fuzzy rules. Systems using fuzzy rules have been applied in a wide variety of applications, viz., automatic control, decision making, risk analysis, etc.

If X is a classical set, we denote the set of all fuzzy sets on X by $\mathcal{F}(X)$. Given two non-empty classical sets X, Y, a fuzzy rule is usually given in the following form:

$$\textbf{IF } x \text{ is } \mathcal{A} \textbf{ THEN } y \text{ is } \mathcal{B} \,, \tag{1}$$

where $\mathcal{A}, \mathcal{B}$ are membership predicates represented by fuzzy sets $A \in \mathcal{F}(X)$ and $B \in \mathcal{F}(Y)$, respectively. Given a fuzzy observation $A' \in \mathcal{F}(X)$ a corresponding output fuzzy set $B' \in \mathcal{F}(Y)$ is deduced using an inference mechanism.

1.1 Fuzzy Relational Inference

Fuzzy inference systems can be broadly classified as those inference mechanisms that are fuzzy relation based, i.e., those that interpret a given fuzzy rule base as a fuzzy relation, and those that are not. Similarity Based Reasoning (SBR) (see, for instance, TÜRKSEN [1]) and Inverse Truth-functional Modification of BALDWIN [2] are two of the representative examples of the latter type of inference mechanisms that do not use fuzzy relations and which are well established in the literature.

In this work, we focus only on those inference mechanisms that are fuzzy relation based. However, it should be mentioned that, under certain conditions, an equivalent fuzzy relation based description of some of these inference mechanisms can be given (see [3, 4]).

1.2 Compositional Rule of Inference

The earliest inference scheme, proposed by Zadeh himself and well established in the literature, is the Compositional Rule of Inference (CRI) [5]. In CRI, a fuzzy rule is interpreted by a fuzzy relation $R \in \mathcal{F}(X \times Y)$. Given a fuzzy observation $A' \in \mathcal{F}(X)$ an output fuzzy set $B' \in \mathcal{F}(Y)$ is deduced by the CRI based on the fuzzy relation R. From the fuzzy relational point of view the CRI is defined as follows:

$$B'(y) = \bigvee_{x \in X} (A'(x) * R(x, y)), \quad y \in Y \,, \tag{2}$$

where $*$ is a fuzzy conjunction, typically a t-norm (see KLEMENT et al. [6] for more details). We use the following notation to indicate the CRI scheme:

$$B' = A' \circ R \,. \tag{3}$$

1.3 Popularity of CRI

The CRI mechanism is most often and widely used for the following reasons:

(i) An important issue in the applicability of CRI scheme is to determine an appropriate fuzzy relation R based on the given fuzzy rules so as to obtain meaningful conclusions. This pertains to the solvability of the fuzzy relation equations. A necessary and sufficient condition for the solvability of fuzzy relation equations has been well established for long, see DI NOLA et al [7]. Moreover, recently, PERFILIEVA and LEHMKE [8] showed its solvability under conditions that are more favourable and prevalent in many applicational contexts.

(ii) In PERFILIEVA et al. [8, 9] the authors have dealt with the continuity of a fuzzy function adjoint to the CRI mechanism and a fuzzy relation interpreting fuzzy rules (8). The authors, defining continuity suitably, have shown that it is equivalent to the correctness of the model under consideration.

(iii) KLAWONN and CASTRO [10] have proven two important and interesting results about the CRI scheme using fuzzy rules and the indistinguishability inherent to the fuzzy sets considered. They showed the robustness of the CRI mechanism in scenarios where there can be slight discrepancies between the intended input and the actual input. They have also shown that the indistinguishability induced by the fuzzy set representing the linguistic expression in the premise of the rule cannot be overcome.

However, it should be noted that there are still some drawbacks that are usually cited about CRI. Recently, JAYARAM

ISBN: 978-989-95079-6-8

[11] has proposed a modified form of CRI called the Hierarchical CRI that addresses many of these drawbacks. The above properties will be dealt with in more detail in Sec. 3.

1.4 Motivation for this work: Bandler-Kohout Subproduct

Generally, the CRI - represented by $\circ$ - may be replaced by any appropriate image of a fuzzy set under the fuzzy relation R denoted by @, which fulfills some required properties. Besides the most often used CRI inference scheme, another scheme of inference is the Bandler-Kohout subproduct $\lhd$ (BK-Subproduct, for short) proposed by BANDLER and KOHOUT [12]. Considering the fuzzy rule as given in (1), for a given input $A' \in \mathcal{F}(X)$ the inference obtained using the BK-Subproduct is given as follows:

$$B'(y) = \bigwedge_{x \in X} (A'(x) \to R(x,y)), \quad y \in Y \ , \tag{4}$$

where $\to$ is a fuzzy implication (see Sec. 2 for more details), $B' \in \mathcal{F}(Y)$ and R is the fuzzy relation that interprets the fuzzy rule (1).

Later on, PEDRYCZ [13] has shown certain solvability conditions for the BK-Subproduct inference as given in (4). Buoyed by this important result on the BK-Subproduct, this paper is an attempt at studying the suitability of the BK-Subproduct as an inference mechanism, by showing that the BK-Subproduct does have all the above desirable properties possessed by CRI.

1.5 Organisation of the work

In Section 2 we recall the preliminaries required for the rest of the paper. Also the structure and inference in the Compositional Rule of Inference (CRI) mechanism is discussed. Following this, in Section 3 we discuss some of the important properties cited in favor of using CRI,e.g., equivalent and reasonable conditions for their solvability, their interpolative properties and the preservation of the indistinguishability that may be inherent in the input fuzzy sets. Section 4, which contains our main work, shows that all the above mentioned properties are valid even in the case of the BK-Subproduct.

2 Preliminaries

2.1 Fuzzy Inference Mechanisms

An inference mechanism acting upon fuzzy rules is usually based on fuzzy logic connectives. In this work, we restrict ourselves to inference schemes based on fuzzy logic connectives. Let us fix a complete residuated lattice $\mathcal{L} = ([0,1], \wedge, \vee, *, \to, 0, 1)$ (see, for example, NOVÁK et al. [14]) as the basic algebraic structure for the whole paper.

Since the set $[0,1]$ is totally ordered, $\mathcal{L}$ becomes an MTL-algebra and hence, in our context, $*$ becomes a left-continuous t-norm and $\to$ is the residual implication *uniquely* given as follows:

$$a \to b = \bigvee \{t \in \mathcal{L} | a * t \leq b\} \ , \tag{5}$$

which is also a fuzzy implication, for more details we refer the readers to [6, 14, 15].

From the two operations $\wedge, \to$ of $\mathcal{L}$ we can derive yet another operation known as the *biresiduum* which is defined as follows:

$$a \leftrightarrow b = (a \to b) \wedge (b \to a), \qquad a, b \in \mathcal{L} \ . \tag{6}$$

Finally, by an extension of an algebraic operation on $\mathcal{L}$ to an operation between fuzzy sets we mean the following:

$$(C \star D)(u,v) = C(u) \star D(v), \qquad u \in U, v \in V \ , \tag{7}$$

where $\star \in \{\wedge, \vee, *, \to, \leftrightarrow\}$ and where C, D are arbitrary fuzzy sets on arbitrary universes U, V, respectively.

2.2 Structure and Inference in CRI

The CRI inference mechanism given in (2) is for a fuzzy relation $R \in \mathcal{F}(X \times Y)$ that interprets a single fuzzy rule. However, usually multiple fuzzy rules are a necessity. Let us denote the multiple fuzzy rules as follows:

$$\textbf{IF } x \text{ is } \mathcal{A}_i \quad \textbf{THEN} \quad y \text{ is } \mathcal{B}_i \ , \tag{8}$$

where $\mathcal{A}_i, \mathcal{B}_i$ are membership predicates represented by fuzzy sets $A_i \in \mathcal{F}(X)$ and $B_i \in \mathcal{F}(Y)$, respectively, for $i = 1, \ldots, n$.

Then a fuzzy relation interpreting the whole *fuzzy rule base* (altern. *linguistic description*), composed of the n fuzzy rules (8), has to be constructed. basically, there are two main approaches to the construction of such a fuzzy relation.

The following fuzzy relation $\check{R} \in \mathcal{F}(X \times Y)$

$$\check{R}(x,y) = \bigvee_{i=1}^{n} (A_i(x) * B_i(y)) \ , \tag{9}$$

is used in most of the real world applications. This is mainly due to the successful applications of this, say Cartesian product approach, published by Mamdani and Assilian in [16], which was followed by a huge number of researchers and practitioners, see e.g. [17, 18].

Alternatively, to keep the conditional IF-THEN form of the fuzzy rules (8) fuzzy relation $\hat{R} \in \mathcal{F}(X \times Y)$ given as follows

$$\hat{R}(x,y) = \bigwedge_{i=1}^{n} (A_i(x) \to B_i(y)) \ , \tag{10}$$

can be chosen to interpret the fuzzy rule base. It deals with a mathematically correct extension of a classical implication.

To stress the difference between both the approaches, let us recall the work of DUBOIS et al. [19], where the authors state: "*In the view given by* (10)*, each piece of information (fuzzy rule) is viewed as a constraint. This view naturally leads to a conjunctive way of merging the individual pieces of information since the more information, the more constraints and the less possible values to satisfy them*." While the same authors describe the second approach proposed by Mamdani and Assilian as follows: "*It seems that fuzzy rules modelled by* $\check{R}$ *are not viewed as constraints but are considered as pieces of data. Then the maximum in (9) expresses accumulation of data*".

In other words, instead of interpreting fuzzy rules as logical implications, the approach given using $\check{R}$ builds up an input-output relation from smaller units, and those units are examples of fuzzy input-output pairs. And then Mamdani-Assilian approach is a method that interpolates between known input-output pairs [20]. For an extensive study of different fuzzy rules we refer to [21, 22].

ISBN: 978-989-95079-6-8

It should be stressed that both approaches have sound logical foundations but from different viewpoints, see e.g. [23, 24, 14]. However, only the approach using $\check{R}$ was widely used in applications although the implicational approach using $\hat{R}$ is probably as useful as the Mamdani-Assilian one, see [25]. Nevertheless, as we show in Section 4, the implicational approach using $\hat{R}$ does have an important role to play in the case of BK-Subproducts (see Theorem 4.7).

3 Some desirable properties of CRI

Fuzzy rules may be viewed as a partial mapping from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$ assigning $B_i \in \mathcal{F}(Y)$ to $A_i \in \mathcal{F}(X)$ for every $i = 1, \ldots, n$. The inference process then may be viewed as an extension of this partial mapping to a total one [26]. For better understanding, let us adopt the notation from [8] and consider the following structure

$$S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, @) \ ,$$

where $@ : \mathcal{F}(X) \times \mathcal{F}(X \times Y) \to \mathcal{F}(Y)$ is an image of a fuzzy set under a fuzzy relation. For instance, $@$ could be one of $\circ$ or $\triangleleft$. Now, by the choice of the fuzzy relation R interpreting the fuzzy rule base and by the choice of $@$, we define a fuzzy function $f_R(A) : \mathcal{F}(X) \to \mathcal{F}(Y)$ such that $f_R(A) = A@R$, for arbitrary $A \in \mathcal{F}(X)$.

3.1 Solvability of systems of fuzzy relation equations with the $\circ$ image

One of the fundamental properties of such a mapping is its interpolativity, i.e., $f_R(A_i) = B_i$. In this case, we say that R is a *correct model* of given fuzzy rules in the given structure S [8].

This leads us to deal with a system of fuzzy relation equations [27] where generally the system of equations

$$A_i@R = B_i, \quad i = 1, \ldots, n \ , \tag{11}$$

is solved with respect to the known $A_i \in \mathcal{F}(X), B_i \in \mathcal{F}(Y)$ and unknown $R \in \mathcal{F}(X \times Y)$. If R is a solution to (11) then the adjoint fuzzy function fulfills $f_R(A_i) = B_i$.

In the case of CRI, the above system of equations, viz., (11), reduces to the following:

$$A_i \circ R = B_i, \quad i = 1, \ldots, n \ . \tag{12}$$

Let us recall some main results which may be found, e.g., in [7, 27, 28, 29].

Theorem 3.1 *System (12) is solvable if and only if $\hat{R}$ is a solution of the system and moreover, $\hat{R}$ is the greatest solution of (12).*

On the one hand, Theorem 3.1 states the necessary and sufficient condition of the solvability of system (12) and it determines the solution. Moreover, it ensures that the given solution is the greatest one. On the other hand, we still do not know, when $\hat{R}$ is the solution, i.e., how to ensure the solvability.

Theorem 3.2 [29] *Let A_i for $i = 1, \ldots, n$ be normal. Then $\check{R}$ is a solution of (12) if and only if the following condition*

$$\bigvee_{x \in X} (A_i(x) * A_j(x)) \leq \bigwedge_{y \in Y} (B_i(y) \leftrightarrow B_j(y)) \ , \tag{13}$$

holds for arbitrary $i, j \in \{1, \ldots, n\}$.

Theorem 3.2 specifies a sufficient condition under which the system is solvable and moreover, it ensures that not only $\hat{R}$ but also even $\check{R}$ is a solution of system (12).

It is worth mentioning that condition (13) appearing in Theorem 3.2 is from the practical point of view not very convenient. However, if the antecedent fuzzy sets form the so called $*$-semi-partition it forces fulfilling the discussed condition in advance, see DE BAETS and MESIAR [30] for more details.

Another sufficient condition for solvability of the systems with a high practical importance was published in [31].

Theorem 3.3 (Theorem 7 in [31]) *Let A_i for $i = 1, \ldots, n$ be normal and fulfill the Ruspini condition*

$$\sum_{i=1}^{n} A_i(x) = 1, \quad x \in X. \tag{14}$$

Then the system (12) is solvable.

3.2 Continuity of a model of fuzzy rules

In [8, 9] the authors have dealt with the continuity of a fuzzy function adjoint to the CRI mechanism and a fuzzy relation interpreting fuzzy rules (8). They have defined continuity suitably and have shown that it is equivalent to the correctness of the model under consideration.

Although the original definition in [8] of a continuous model was given for the particular inference mechanism CRI, i.e., for $@ \equiv \circ$, the particular image plays absolutely no role in the proof of Theorem 3.6 explaining the nature of the definiton and hence can be generalized for an arbitrary image of a fuzzy set under a fuzzy relation.

Definition 3.4 A fuzzy relation $R \in \mathcal{F}(X \times Y)$ is said to be a *continuous model of fuzzy rules* (8) *in a structure* $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, @)$ if for each $i \in I$ and for each $A \in \mathcal{F}(X)$ the following inequality holds:

$$\bigwedge_{y \in Y} (B_i(y) \leftrightarrow (A@R)(y)) \geq \bigwedge_{x \in X} (A_i(x) \leftrightarrow A(x)) \ . \tag{15}$$

Inequality (15) can be rewritten in terms of the adjoint fuzzy function f_R as follows:

$$\bigwedge_{y \in Y} (B_i(y) \leftrightarrow (f_R(A))(y)) \geq \bigwedge_{x \in X} (A_i(x) \leftrightarrow A(x)). \tag{16}$$

Remark 3.5 Let us explain why formula (15) expresses the continuity. The closeness between fuzzy sets is measured by the biresiduation operation $\leftrightarrow$, i.e. it is a dual information to the metric one. Let us consider a continuous Archimedean t-norm $*$ with an additive generator $g : [0, 1] \to [0, +\infty]$. Then the biresiduum may be written in the form

$$a \leftrightarrow b = g^{-1}(|g(a) - g(b)|) \tag{17}$$

where $g^{-1} : [0, \infty] \to [0, 1]$ is the inverse function and where in the case of $g(0) = \infty$ we define $g(0) - g(0) = 0$. Now, for an arbitrary non-empty universe X it is possible to define a metric D_g on $\mathcal{F}(X)$ generated by g as follows:

$$D_g(A, B) = \bigvee_{x \in X} |g(A(x)) - g(B(x))|. \tag{18}$$

The following theorem justifies the use of the notion of continuity in Definition 3.4. For more details, see PERFILIEVA and LEHMKE [8].

ISBN: 978-989-95079-6-8

Theorem 3.6 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, @)$ be a structure for fuzzy rules* (8) *such that $\mathcal{L}$ be a residuated lattice on $[0,1]$ with a continuous Archimedean t-norm $*$ having a continuous additive generator g. A fuzzy relation $R \in \mathcal{F}(X \times Y)$ is a continuous model of the fuzzy rules in the given structure S if and only if*

$$D_g(B_i, (A@R)) \leq D_g(A_i, A), \quad i = 1, \ldots, n \tag{19}$$

for each fuzzy set $A \in \mathcal{F}(X)$.

The following lemma was crucial for further results published in [8].

Lemma 3.7 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \circ)$ be a structure for fuzzy rules* (8) *and let $R \in \mathcal{F}(X \times Y)$. Then for any $A \in \mathcal{F}(X)$ and all $i = 1, \ldots, n$ and $y \in Y$ it is true that*

$$B_i(y) \leftrightarrow (A \circ R)(y) \geq \delta_{R,i}(y) * \bigwedge_{x \in X} (A_i(x) \leftrightarrow A(x)) , \tag{20}$$

where $\delta_{R,i}(y) = B_i(y) \leftrightarrow (A_i \circ R)(y)$.

Finally, we recall here the main result in PERFILIEVA et al. [8, 9] concerning the relationship of the above mentioned continuity and the interpolativity for the CRI.

Theorem 3.8 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \circ)$ be a structure for fuzzy rules* (8)*. A fuzzy relation $R \in \mathcal{F}(X \times Y)$ is a correct model of fuzzy rules* (8) *in the given structure S if and only if it is a continuous model of these rules in S.*

3.3 CRI and the Indistinguishability of Premises

Let X be a classical set and let $\sim$ be an equivalence relation defined on X, i.e., $\sim$ is reflexive, symmetric and transitive. Immediately, $\sim$ partitions X into equivalence classes. It is well-known then that an $M \subseteq X$ belongs to this partition if, and only if, whenever $x \in M$ and $x \sim y$ for some $y \in X$ then $y \in M$. In a sense, the elements of M are indistinguishable and can be represented mathematically as follows:

$$x \in M \text{ and } x \sim y \text{ implies } y \in M .$$

A similar relation between fuzzy equivalence relations and fuzzy sets on X was introduced by KLAWONN and CASTRO [10]. Once again the operation $*$ comes from the residuated lattice $\mathcal{L}$.

Definition 3.9 A fuzzy subset E of the Cartesian product X^2 is called a *fuzzy equivalence* relation on X if the following properties are satisfied for all $x, y, z \in X$:

$$\begin{aligned} &\text{(Reflexivity)} && E(x,x) = 1, && \text{(ER)} \\ &\text{(Symmetry)} && E(x,y) = E(y,x), && \text{(ES)} \\ &\text{(Transitivity)} && E(x,z) \geq E(x,y) * E(y,z) . && \text{(ET)} \end{aligned}$$

Definition 3.10 A fuzzy set $\mu \in \mathcal{F}(X)$ is called *extensional* with respect to a fuzzy equivalence relation E on X if

$$\mu(x) * E(x,y) \leq \mu(y) , \qquad x, y \in X . \tag{21}$$

If a fuzzy set μ is not extensional with respect to the considered fuzzy equivalence relation E, instead one considers the smallest fuzzy set that is extensional with respect to E and contains μ.

Definition 3.11 Let $\mu \in \mathcal{F}(X)$ and let E be a fuzzy equivalence relation on X. The fuzzy set

$$\widetilde{\mu}(x) = \bigwedge \{\nu \mid \mu \leq \nu \text{ and } \nu \text{ is extensional with respect to } E\} , \tag{22}$$

is called the *extensional hull* of μ. Note that by $\mu \leq \nu$ we mean that for all $x \in X$, $\mu(x) \leq \nu(x)$, i.e., we mean ordering in the sense of inclusion, not in the sense of ordering fuzzy quantities.

Proposition 3.12 [10] *Let $\mu \in \mathcal{F}(X)$ and let E be a fuzzy equivalence relation on X.*

(i) $\widetilde{\mu}(x) = \bigvee \{\mu(y) * E(x,y) \mid y \in X\}$,

(ii) $\widetilde{\mu}$ *is extensional with respect to E,*

(iii) $\widetilde{\widetilde{\mu}} = \widetilde{\mu}$.

KLAWONN and CASTRO [10] have proven the following two important and interesting results about the CRI scheme using fuzzy rules and the indistinguishability inherent to the fuzzy sets considered.

Theorem 3.13 *Let $S = (X, Y, \{A, B\}, \mathcal{L}, \circ)$ be a structure for fuzzy rule* (1)*. Let E be a fuzzy equivalence relation on X with respect to which A is extensional. Let $A' \in \mathcal{F}(X)$ be any fuzzy set, then*

$$A' \circ \hat{R} = \widetilde{A'} \circ \hat{R},$$
$$A' \circ \check{R} = \widetilde{A'} \circ \check{R}.$$

The following interpretation of the above result is given in [10]: *The output obtained from CRI for a given fuzzy rule and an input fuzzy set A' does not change if we substitute A' by its extensional hull $\widetilde{A'}$. The indistinguishability inherent in the fuzzy set A cannot be avoided even if the input fuzzy set A' stands for a crisp value*. Further, *a fuzzified input does not change the outcome of a rule as long as the fuzzy set obtained by the fuzzification is contained in the extensional hull of the original crisp input value.* They finally conclude that *it does not make sense to measure more exactly than the indistinguishability admits.*

In other words, this shows the robustness of the inference in scenarios where there can be slight discrepancies between the intended input and the actual input.

It is immediate now, as already observed in [10], that *the indistinguishability induced by the fuzzy set representing the linguistic expression in the premise of the rule cannot be overcome.*

Let us stress that the proof of Theorem 3.13 may be easily modified in order to capture the situation with n fuzzy rules. It can be observed that only the Mamdani-Assilian approach $\check{R}$ generally works in the combination with the CRI.

Theorem 3.14 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \circ)$ be a structure for fuzzy rules (8). Let E be a fuzzy equivalence relation on X with respect to which A_i is extensional for arbitrary $i = 1, \ldots, n$. Let $A' \in \mathcal{F}(X)$ be any fuzzy set, then*

$$A' \circ \check{R} = \widetilde{A'} \circ \check{R}.$$

ISBN: 978-989-95079-6-8

4 BK-Subproduct and some desirable properties

4.1 Solvability of systems of fuzzy relation equations with the $\triangleleft$ image

In the case of BK-Subproduct, the system of equations (11), reduces to the following:

$$A_i \triangleleft R = B_i, \quad i = 1, \ldots, n \, . \tag{23}$$

Concerning system (23), let us recall the following two basic theorems .

Theorem 4.1 [13] *System (23) is solvable if and only if $\check{R}$ is a solution of the system and moreover, $\check{R}$ is the least solution of system (23).*

Theorem 4.2 [32] *Let A_i for $i = 1, \ldots, n$ be normal. Then $\hat{R}$ is a solution of (23) if and only if the condition* (13) *holds for arbitrary $i, j \in \{1, \ldots, n\}$.*

Again, condition (13) to which Theorem 4.2 refers to, is not very convenient from a practical point of view and may be a priori fulfilled by using antecedent fuzzy sets which form the $*$-semi-partition.

Fortunately, that the sufficient condition for solvability of the systems with a high practical importance stated in Theorem 3.3 is valid even for system (23).

Theorem 4.3 (Theorem 7 in [31]) *Let A_i for $i = 1, \ldots, n$ be normal and fulfill the Ruspini condition (14). Then the system (23) is solvable.*

4.2 Continuity of BK-Subproduct

The BK-Subproduct, unlike the CRI, was not motivated by approximate reasoning, see BANDLER and KOHOUT [12]. However, as mentioned in GOTTWALD [33], it is not necessary to insist on purely logical foundations for an inference mechanism, and it can simply be a mapping from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$ fulfilling certain properties. Let us also recall that it was PEDRYCZ [13] who firstly proposed the BK-Subproduct as an inference scheme.

We may again state an analogous lemma to Lemma 3.7 also for the case of the BK-Subproduct.

Lemma 4.4 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \triangleleft)$ be a structure for fuzzy rules* (8) *and let $R \in \mathcal{F}(X \times Y)$. Then for any $A \in \mathcal{F}(X)$ and all $i = 1, \ldots, n$ and $y \in Y$ it is true that*

$$B_i(y) \leftrightarrow (A \triangleleft R)(y) \geq \delta_{R,i}(y) * \bigwedge_{x \in X} (A_i(x) \leftrightarrow A(x)) \, , \tag{24}$$

where $\delta_{R,i}(y) = B_i(y) \leftrightarrow (A_i \triangleleft R)(y)$.

Due to Lemma 4.4, we may state the following analogous theorem to Theorem 3.8 which again shows that the BK subproduct as an inference mechanism carries the same property as the CRI.

Theorem 4.5 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \triangleleft)$ be a structure for fuzzy rules* (8)*. A fuzzy relation $R \in \mathcal{F}(X \times Y)$ is a correct model of fuzzy rules* (8) *in the given structure S if and only if it is a continuous model of these rules in S.*

4.3 BK-Subproduct and the Indistinguishability of Premises

In this subsection, we show the robustness of the BK-Subproduct inference mechanism along similar lines as KLAWONN and CASTRO [10]. Once again the employed operations come from the residuated lattice $\mathcal{L}$. Firstly note that if a fuzzy set $\mu \in \mathcal{F}(X)$ is extensional with respect to a fuzzy equivalence relation E on X then

$$E(x, y) \to \mu(y) \geq \mu(x), \; x, y \in X \, . \tag{25}$$

Proposition 4.6 *Let $\mu \in \mathcal{F}(X)$ and E a fuzzy equivalence relation on X. Then*

$$\widetilde{\mu}(x) = \bigwedge \{E(x, y) \to \mu(y) \mid y \in X\} \, . \tag{26}$$

Now, we present the main result analogous to Theorem 3.13.

Theorem 4.7 *Let $S = (X, Y, \{A, B\}, \mathcal{L}, \triangleleft)$ be a structure for fuzzy rule* (1)*. Let E be a fuzzy equivalence relation on X with respect to which A is extensional. Let $A' \in \mathcal{F}(X)$ be any fuzzy set, then*

$$A' \triangleleft \hat{R} = \widetilde{A'} \triangleleft \hat{R},$$
$$A' \triangleleft \check{R} = \widetilde{A'} \triangleleft \check{R}.$$

The above result, as already noted in the case of CRI, shows the robustness of the BK-Subproduct inference in scenarios where there can be slight discrepancies between the intended input and the actual input and reinforces the fact that even in the case of BK-Subproduct the indistinguishability induced by the fuzzy set representing the linguistic expression in the premise of the rule cannot be overcome.

And again as in case of the CRI, we may generalize the result concerning the indistinguishability of the premises for an arbitrary finite number of rules. Note that in the case of the BK-Subproduct the $\hat{R}$ is plays the main role.

Theorem 4.8 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \triangleleft)$ be a structure for fuzzy rules (8). Let E be a fuzzy equivalence relation on X with respect to which each A_i is extensional, for arbitrary $i = 1, \ldots, n$. Let $A' \in \mathcal{F}(X)$ be any fuzzy set, then*

$$A' \triangleleft \hat{R} = \widetilde{A'} \triangleleft \hat{R}.$$

5 Conclusions

In this work, after recalling some of the properties that are usually cited in favor of using the Compositional Rule of Inference (CRI) introduced by ZADEH [5], viz., equivalent and reasonable conditions for their solvability, their interpolative properties and the preservation of the indistinguishability that may be inherent in the input fuzzy sets, we have shown that the Bandler-Kohout subproduct introduced in [12] does possess all the above properties and hence is equally suitable for consideration when reasoning with a system of fuzzy rules. Towards this end some new but equivalent results on indistinguishability operations has also been presented.

Acknowledgment

This work has been supported by projects MSM6198898701 of the MŠMT ČR and by IAA108270902 of the GA AV ČR.

ISBN: 978-989-95079-6-8

References

[1] I. B. Turksen and Z. Zhong. An approximate reasoning schema based on similarity measures and interval valued fuzzy sets. *IEEE Trans. Syst. Man Cybern.*, 18:1049–1056, 1988.

[2] J. F. Baldwin. A new approach to approximate reasoning using fuzzy logic. *Fuzzy Sets and Systems*, 2:309–325, 1979.

[3] V. Cross and T. Sudkamp. Patterns of fuzzy rule-based inference. *Int. J. of Approximate Reasoning*, 11:235–255, 1994.

[4] B. Bouchon-Meunier, R. Mesiar, C. Marsala, and M. Rifqi. Compositional rule of inference as an analogical scheme. *Fuzzy Sets and Systems*, 138:56–65, 2003.

[5] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC-3:28–44, 1973.

[6] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer, Dordrecht, 2000.

[7] Di Nola A., S. Sessa, W. Pedrycz, and E. Sanchez. *Fuzzy Relation Equations and Their Applications to Knowledge Engineering*. Kluwer, Boston, 1989.

[8] I. Perfilieva and S. Lehmke. Correct models of fuzzy if-then rules are continuous. *Fuzzy Sets and Systems*, 157:3188–3197, 2006.

[9] I. Perfilieva and V. Novák. System of fuzzy relation equations as a continuous model of IF-THEN rules. *Information Sciences*, 177:3218–3227, 2007.

[10] F. Klawonn and J. L. Castro. Similarity in fuzzy reasoning. *Mathware & Soft Computing*, 2:197–228, 1995.

[11] B. Jayaram. On the law of importation $(x \wedge y) \rightarrow z \equiv (x \rightarrow (y \rightarrow z))$ in fuzzy logic. *IEEE Trans. Fuzzy Syst.*, 16:130–144, 2008.

[12] W. Bandler and L. J. Kohout. Semantics of implication operators and fuzzy relational products. *Int. J. Man-Mach. Stud.*, 12:89–116, 1979.

[13] W. Pedrycz. Applications of fuzzy relational equations for methods of reasoning in presence of fuzzy data. *Fuzzy Sets and Systems*, 16:163–175, 1985.

[14] V. Novák, I. Perfilieva, and J. Močkoř. *Mathematical Principles of Fuzzy Logic*. Kluwer, Boston, 1999.

[15] M. Baczyński and B. Jayaram. *Fuzzy Implications (Studies in Fuzziness and Soft Computing series)*. Springer-Verlag, Heidelberg, Germany, 2008.

[16] E.H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. of Man-Machine Studies*, 7:1–13, 1975.

[17] W. J. M. Kickert and E. H. Mamdani. Analysis of a fuzzy logic controller. *Fuzzy Sets and Systems*, 1:29–44, 1978.

[18] P. M. Larsen. Industrial applications of fuzzy logic control. *Int. J. Man-Mach. Stud.*, 12:3–10, 1980.

[19] D. Dubois, H. Prade, and L. Ughetto. Checking the coherence and redundancy of fuzzy knowledge bases. *IEEE Trans. Fuzzy Systems*, 5(6):398–417, 1997.

[20] F. Klawonn and V. Novák. The relation between inference and interpolation in the framework of fuzzy systems. *Fuzzy Sets and Systems*, 81(3):331–354, 1996.

[21] D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185, 1996.

[22] V. Novák and S. Lehmke. Logical structure of fuzzy IF-THEN rules. *Fuzzy Sets and Systems*, 157:2003–2029, 2006.

[23] L. Godo and P. Hájek. A note on fuzzy inference as deduction. In D. Dubois, E. P. Klement, and H. Prade, editors, *Fuzzy Sets, Logics and Reasoning about Knowledge (Applied Logic series Vol.15)*, pages 237–241. Kluwer Academic Publishers, Dordrecht, Netherlands, 1999.

[24] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer, Dordrecht, 1998.

[25] U. Bodenhofer, M. Daňková, M. Štěpnička, and V. Novák. A plea for the usefulness of the deductive interpretation of fuzzy rules in engineering applications. In *Proc. of FUZZ-IEEE'07*, pages 1567–1572, London, 2007.

[26] T. Vetterlein and M. Štěpnička. Completing fuzzy if-then rule bases by means of smoothing splines. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Syst.*, 14:235–244, 2006.

[27] E. Sanchez. Resolution of composite fuzzy relation equations. *Inform. and Control*, 30:38–48, 1976.

[28] B. De Baets. Analytical solution methods for fuzzy relation equations. In D. Dubois and H. Prade, editors, *The Handbook of Fuzzy Set Series Vol. 1*, volume 15. 2000.

[29] F. Klawonn. Fuzzy points, fuzzy relations and fuzzy functions. In V. Novák and I. Perfilieva, editors, *Discovering the World with Fuzzy Logic*, pages 431–453. Springer, Berlin, 2000.

[30] B. De Baets and R. Mesiar. T-partitions. *Fuzzy Sets and Systems*, 97:211–223, 1998.

[31] M. Štěpnička, B. De Baets, and L. Nosková. On additive and multiplicative fuzzy models. In M. Štěpnička, V. Novák, and U. Bodenhofer, editors, *New Dimensions in Fuzzy Logic and Related Technologies, Int. Conference EUSFLAT 2007, Ostrava, Proceedings*, volume 2, pages 95–102, Ostrava, 2007.

[32] L. Nosková. Systems of fuzzy relation equation with inf$\rightarrow$ composition: solvability and solutions. *J. of Electrical Eng.*, 12(s):69–72, 2005.

[33] S. Gottwald. An abstract approach toward the evaluation of fuzzy rule systems. In P.P. Wang, D. Ruan, and E.E. Kerre, editors, Fuzzy Logic. *A Spectrum of Theoretical & Practical Issues (Studies in Fuzziness and Soft Computing 215)*, pages 281–297. Springer-Verlag, 2007.

ISBN: 978-989-95079-6-8

Early Start Can Inhibit Learning: Towards A New Explanation

Olga Kosheleva

Department of Teacher Education
University in the Texas at El Paso
El Paso, TX 79968, USA
Email: olgak@utep.edu

Abstract— *The age at which we teach different topics change. If it turns out that students do not learn, say, reading by the time they should, a natural idea is to start teaching them earlier. Several decades ago, reading and writing started in the first grade, now they start at kindergarten and even earlier. At first glance, the earlier we start, the better the students will learn. Yes, they may play less, their childhood may be not as careless as it used to be, but at least they will learn better. In practice, however, this does not always work this way: early start often inhibits learning.*

In this paper, we propose a simple model that explains why this can happen, and how we can avoid such situations and enhance the student learning. In particular, we explain why, contrary to the traditional concrete-first way of teaching abstracts concepts of school mathematics, abstract-first approach is often beneficial.

Keywords— Abstract-first, early start, education, optimal education, uncertainty

1 Formulation of the Problem

1.1 Early Start: A Seemingly Natural Idea

The age at which we teach different topics change. If it turns out that students do not learn, say, reading by the time they should, a natural idea is to start teaching them earlier.

Several decades ago, reading and writing started in the first grade, now they start at kindergarten and even earlier. At first glance, the earlier we start, the better the students will learn.

1.2 Early Start: Known Side Effects

With the early start, children may play less, their childhood may be not as careless as it used to be – but a usual expectation is that with an early start, children will learn better.

1.3 Early Start: Serious Problems

In practice, however, early start does not always help: often, early start inhibits learning.

For example, according to [18],

- human infants who started learning to turn their heads to specific sounds at the age of 31 days mastered this task, on average, at the age of 71 days, while
- infants who started learning this task at birth mastered this task, on average, at the age of 128 days.

This phenomenon is not limited to human infants: according to [8], an early start in training rhesus monkeys to discriminate objects decreased their peak performance level.

Numerous examples when an early start inhibits learning are presented and discussed in [2, 3, 4, 6, 19].

1.4 Natural Questions

The empirical fact that an early start often inhibits learning leads to the following natural question: how do take this phenomenon into account when enhancing student learning?

To be able to take this phenomenon into account in the learning process, we must be able to understand this phenomenon – and ideally, understand on the quantitative level.

1.5 These Questions Are Still Largely Open

In [2, 3, 4, 6], an attempt is made to understand why early start can inhibit learning. However,

- the existing understanding is still mostly on the qualitative level, and
- even on this level, the proposed explanations are still not fully satisfactory; see, e.g., [19].

1.6 More General Questions

The above questions about the efficiency of the early start can be viewed as a particular case of more general questions: what is the best order of presenting the material, the order that leads to the best possible learning?

1.7 These More General Questions Are Often Very Important

Many empirical studies have shown that a change in the order in which different parts of the material are presented often drastically changes the learning efficiency; see, e.g., [5, 9, 10, 13, 14, 16, 21, 22].

This is not only about using common sense: sometimes, the empirical results are counter-intuitive. For example:

- it is usually assumed that most students learn mathematical concepts better if
  - they are first presented with concrete examples of these concepts, and
  - they only learn abstract ideas later on;
- however, it turns out that empirically, the abstract-first approach for presenting the material often enhances learning; see, e.g., [9, 10, 14, 21].

1.8 What We Do in This Paper

In this paper, we propose a simple model that attempts:

- to explain the negative effect of early start and,

ISBN: 978-989-95079-6-8

- more generally, to explain the reasons why a change in presentation order can drastically change the efficiency of learning.

We then show how this explanation can be used to avoid inhibition of learning – and to enhance the student learning.

2 Towards A Theoretical Explanation

2.1 Learning: A Natural Geometric Representation

To facilitate reasoning about learning, let us start with a simple geometric representation of learning.

The process of learning means that we change the state of a student:

- from a state in which the student did not know the material (or does not have the required skill)
- to a state in which the student has (some) knowledge of the required material (or has the required skill).

Let s_0 denote the original state of a student, and let S denote the set of all the states corresponding to the required knowledge or skill.

- We start with a state which is not in the set S ($s_0 \notin S$), and
- we end up in a state s which is in the set S.

On the set of all possible states, it is natural to define a metric $d(s, s')$ as the difficulty (time, effort, etc.) needed to go from state s to state s'. Our objective is to help the students learn in the easiest (fastest, etc.) way. In terms of the metric d, this means that we want to go from the original state $s_0 \notin S$ to the state $s \in S$ for which the effort $d(s_0, s)$ is the smallest possible.

In geometric terms, the smallest possible effort means the shortest possible distance. Thus, our objective is to find the state $s \in S$ which is the closest to s_0. Such closest state is called the *projection* of the original state s_0 on the set S.

2.2 Learning Complex Material: Geometric Interpretation

The above geometric description of learning as a transition from the original state s_0 to its projection on the desired set S describes learning *as a whole*. Our objective is to find out which order of presenting information is the best. Thus, our objective is to analyze the *process* of learning, i.e., learning as a multi-stage phenomenon. For this analysis, we must explicitly take into account that the material to be learned consists of several pieces.

Let S_i, $1 \le i \le n$, denote the set of states in which a student has learned the i-th part of the material. Our ultimate objective is to make sure that the student learns all the parts of the material. In terms of states, learning the i-th part of the material means belonging to the set S_i. Thus, in terms of states, our objective means that the student should end up in a state which belongs to all the sets $S_1, \ldots, S_n$ – i.e., in other words, in a state which belongs to the intersection

$$S \stackrel{\text{def}}{=} S_1 \cap \ldots \cap S_n \tag{1}$$

of the corresponding sets S_i.

In these terms, if we present the material in the order S_1, S_2, …, S_n, this means that:

- we first project the original state s_0 onto the set S_1, resulting is a state $s_1 \in S_1$ which is the closest to s_0;
- then, we project the state s_1 onto the set S_2, resulting is a state $s_2 \in S_2$ which is the closest to s_1;
- …
- at the last stage of the cycle, we project the state s_{n-1} onto the set S_n, resulting is a state $s_n \in S_n$ which is the closest to s_{n-1}.

In some cases, we end up learning all the material – i.e., in a state $s_n \in S_1 \cap \ldots \cap S_n$. However, often, by the time the students have learned S_n, they have somewhat forgotten the material that they learned in the beginning. So, it is necessary to repeat this material again (and again). Thus, starting from the state s_n, we again sequentially project onto the sets S_1, S_2, etc.

2.3 The Above Geometric Interpretation Makes Computational Sense

The above "sequential projections" algorithm is actually actively used in many applications; see, e.g., [7, 12, 20]. In the case when all the sets S_i are convex, the resulting Projections on Convex Sets (POCS) method actually guarantees (under certain reasonable conditions) that the corresponding projections converge to a point from the intersection $S_1 \cap \ldots \cap S_n$ – i.e., in our terms, that the students will eventually learn all parts of the necessary material.

In the more general non-convex case, the convergence is not always guaranteed – but the method is still efficiently used, and often converges.

2.4 The Simplest Case: Two-Part Knowledge

Let us start with the simplest case when knowledge consists of two parts. In this simplest case, there are only two options:

- The first option is that:
  - we begin by studying S_1;
  - then, we study S_2,
  - then, if needed, we study S_1 again, etc.
- The second option is that:
  - we begin by studying S_2;
  - then, we study S_1,
  - then, if needed, we study S_2 again, etc.

We want to get from the original state s_0 to the state $\widetilde{s} \in S_1 \cap S_2$ which is the closest to s_0. The effectiveness of learning is determined by how close we get to the desired set $S = S_2 \cap S_2$ in a given number of iterations.

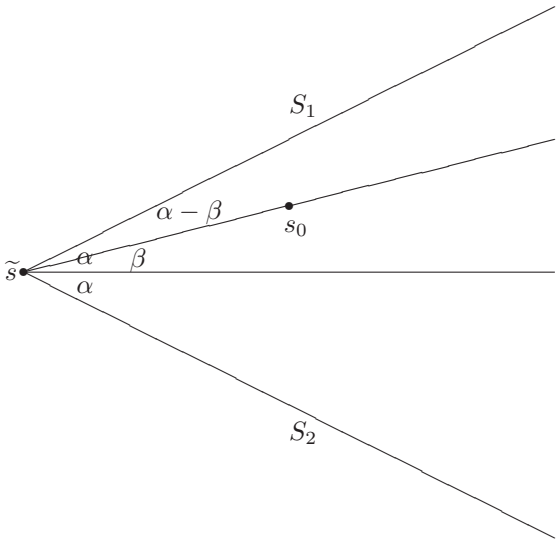
In the case of two-part knowledge, it is natural to conclude that the amount of this knowledge is reasonably small – otherwise, we would have divided into a larger number of easier-to-learn pieces.

In geometric terms, this means that the original state s_0 is close to the desired intersection set $S_1 \cap S_2$, i.e., that the distance $d_0 \stackrel{\text{def}}{=} d(s_0, \widetilde{s})$ is reasonably small.

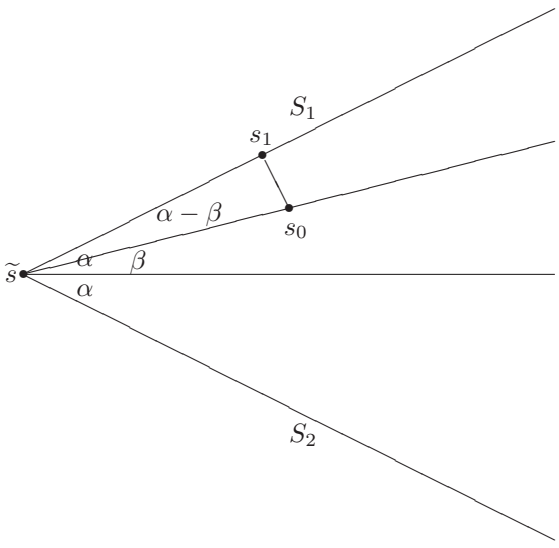
ISBN: 978-989-95079-6-8

Since all the states are close to each other, in the vicinity of the state $\widetilde{s}$, we can therefore expand the formulas describing the borders of the sets S_i into Taylor series and keep only terms which are linear in the (coordinates of the) difference $s - \widetilde{s}$. Thus, it is reasonable to assume that the border of each of the two sets S_i is described by a linear equation – and is hence a (hyper-)plane: a line in 2-D space, a plane in 3-D space, etc.

As a result, we arrive at the following configuration. Let 2α denote the angle between the borders of the sets S_1 and S_2, so that the angles between each of these borders and the midline is exactly α. Let β denote the angle between the direction from $\widetilde{s}$ to s_0 and the midline. In this case, the angle between the border of S_1 and the midline is equal to $\alpha - \beta$. So, we arrive at the following configuration:



In the first option, we first project s_0 onto the set S_1. As a result, we get the following configuration:



Here, the projection line $s_0 s_1$ is orthogonal to the border of S_1. From the right triangle $\triangle \widetilde{s} s_0 s_1$, we therefore conclude that the distance $d_1 \stackrel{\text{def}}{=} d(\widetilde{s}, s_1)$ from the projection point s_1 to the desired point $\widetilde{s}$ is equal to

$$d_1 = d_0 \cdot \cos(\alpha - \beta). \quad (2)$$

On the next step, we project the point s_1 from S_1 onto the line S_2 which is located at the angle 2α from S_1. Thus, for the projection result s_2, we will have

$$d_2 = d(s_2, \widetilde{s}) = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos(2\alpha). \quad (3)$$

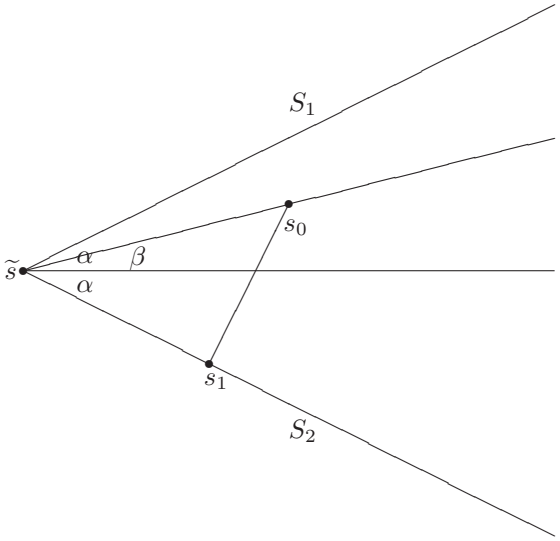
After this, we may again project onto S_2, then again project onto S_1, etc. For each of these projections, the angle is equal to 2α, so after each of them, the distance from the desired point $\widetilde{s}$ is multiplied the same factor $\cos(2\alpha)$.

As a result, after k projection steps, we get a point s_k at a distance

$$d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha) \quad (4)$$

from the desired state $\widetilde{s}$.

In the second option, we start with teaching S_2, i.e., if we first project the state s_0 into the set S_2. In this option, we get the following configuration:



Here, we have

$$d_1 = d_0 \cdot \cos(\alpha + \beta), \quad (5)$$

$$d_2 = d(s_2, \widetilde{s}) = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha + \beta) \cdot \cos(2\alpha), \quad (6)$$

$$\ldots$$

$$d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha + \beta) \cdot \cos^{k-1}(2\alpha). \quad (7)$$

Since, in general, $\cos(\alpha - \beta) \neq \cos(\alpha + \beta)$, we can see that a change in the presentation order can indeed drastically change the success of the learning procedure.

2.5 Conclusion: Dependence Explained

Thus, our simple geometric model explains why the effectiveness of learning depends on the order in which the material is presented.

3 Specific Recommendations: Case of Two-Part Knowledge

3.1 Analysis

Let us extract more specific recommendations from our model. According to the above formulas, starting with S_1

ISBN: 978-989-95079-6-8

leads to a more effective learning than starting with S_2 if and only if

$$d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha) < d_0 \cdot \cos(\alpha + \beta) \cdot \cos^{k-1}(2\alpha), \quad (8)$$

i.e., equivalently, if and only if

$$\cos(\alpha - \beta) < \cos(\alpha + \beta). \quad (9)$$

Since for the angles $x \in [0, \pi]$, the cosine $\cos(x)$ is a decreasing function, we conclude that projection of S_1 is better if and only if

$$\alpha - \beta > \alpha + \beta. \quad (10)$$

Thus, we arrive at the following recommendation:

3.2 Recommendation

To make learning more efficient, we should start with studying the material which is further away from the current state of knowledge. In other words, we should start with a material that we know the least.

This ties in nicely with a natural commonsense recommendation that to perfect oneself, one should concentrate on one's deficiencies.

This recommendation is also in a very good accordance:

- with the seemingly counter-intuitive conclusion from [9, 10, 14, 21], that studying more difficult (abstract) ideas first enhances learning, and
- with the human infant studies [18] according to which a concentration on teaching, to human infants, skills that they can easily learn is detrimental in the long run.

4 General Case

4.1 Analysis of the Problem

What happens in the general case, when instead of only two knowledge components, we have a large number of different components? In the beginning, it still makes sense to project to the set S_{i_1} which is the farthest from the original state s_0.

After this original projection, in the general case, we still have a choice. We can project to any set S_{i_2}, $i_2 \neq i_1$, in which case the current distance d_1 to the desired state is multiplied by the cosine $\cos(\alpha_{i_1 i_2})$ of the angle between the corresponding sets S_{i_1} and S_{i_2}. After k steps, we get the original distance multiplied by the product of the corresponding cosines.

Our objective is to find the best order, i.e., the sequence S_{i_1}, S_{i_2}, ..., S_{i_n} that covers all n sets $S_1, \ldots, S_n$ and for which the corresponding product

$$\cos(\alpha_{i_1 i_2}) \cdot \cos(\alpha_{i_1 i_2}) \cdot \ldots \cdot \cos(\alpha_{i_n i_1}) \quad (11)$$

attains the smallest possible value.

Usually, it is easier to deal with the sums than with the products. To transform the product into a sum, we can use the fact that:

- minimizing the product is equivalent to minimizing its logarithm, and
- the logarithm of the product is equal to the sum of the logarithms.

Thus, minimizing the product (11) is equivalent to minimizing the sum

$$D(i_1, i_2) + \ldots + D(i_n, i_1), \quad (12)$$

where

$$D(i, j) \stackrel{\text{def}}{=} \log(\cos(\alpha_{ij})). \quad (13)$$

In other words, we arrive at the following problem:

- we have n objects with known distances

$$D(i, j), \ \ 1 \leq i, j \leq n;$$

- we must find a way to traverse all the objects and come back in such a way that the overall traveled distance is the smallest possible.

This is a well-known problem called a *traveling salesman* problem.

It is known that

- in general, this problem is NP-hard (see, e.g., [17]), and
- in many cases, there exist reasonable algorithms for solving this problem; see, e.g., [1].

4.2 Recommendations and the Need for Expert (Fuzzy) Knowledge

Based on the above analysis, we can make the following recommendations:

- To find the optimal order of presenting the material, we must solve the corresponding instance of the traveling salesman problem, with the distances determined by the formula (13).
- Since in general, the traveling salesman problem is computationally difficult (NP-hard), to efficiently solve this problem, we must use expert knowledge – the knowledge for which fuzzy technique have been invented; see, e.g., [11, 15].

Acknowledgment

The author is thankful to Mourat Tchoshanov for valuable discussions, and to the anonymous referees for useful suggestions.

References

[1] D. L. Applegate, R. M. Bixby, V. Chvátal, and W. J. Cook, *The Traveling Salesman Problem*, Princeton University Press, Princeton, New Jersey, 2006.

[2] D. F. Bjorklund, *Why Youth is Not Wasted on the Young: Immaturity in Human Development*, Blackwell Publishing, 2007.

[3] D. F. Bjorklund and B. L. Green, The adaptive nature of cognitive immaturity, *American Psychologist*, 47:46–54, 1992.

[4] D. F. Bjorklund and A. Pellegrini, *The Origins of Human Nature: Evolutionary Developmental Psychology*, American Psychological Association, Washington, DC, 2002.

ISBN: 978-989-95079-6-8

[5] V. Davydov, *Types of Generalizations in Instruction: Logical and Psychological Problems in the Structuring of Schhol Curricular*, NCTM, Reston, Virginia, 1990.

[6] B. J. Ellis and D. F. Bjorklund (eds.), *Origins of the Social Mind: Evolutionary Psychology and Child Development*, Guilford Press, New York, 2005.

[7] L. G. Gubin, B. T. Polyak, and E. V. Raik, The method of projections for finding the comon point of convex sets, *USSR Computational Mathematics and Mathematical Physics*, 7:1–24, 1967.

[8] H. F. Harlow, The development of learning in the rhesus monkey, *American Scientist*, 47:459–479, 1959.

[9] J. A. Kaminski, V. M. Sloutsky, and A. F. Heckler, Do children need concrete instantiations to learn an abstract concept?, In: R. Sun and N. Miyake (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society, Vancouver, BC, July 26–29, 2006*, Lawrence Erlbaum, Mahwah, NJ, 2006, pp. 411–416.

[10] J. A. Kaminski, V. M. Sloutsky, and A. F. Heckler, The advantage of abstract examples in learning math, *Science*, 320:454–455, April 25, 2008.

[11] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, New Jersey, 1995.

[12] E. J. Kontoghiorghes, *Handbook of Parallel Computing and Statistics*, CRC Press, Boca Raton, Florida, 2006.

[13] L. M. Lesser and M. A. Tchoshanov, The effect of representation and representational sequence on students' understanding, In: G. M. Lloyd, M. Wilson, J. M. L. Wilkins, and S. L. Behm (eds.), *Proceedings of the 27th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education PME-NA'2005*, Roanoke, Virginia, October 20–23, 2005.

[14] L. M. Lesser and M. A. Tchoshanov, Selecting representations, *Texas Mathematics Teacher*, 53(2):20–26, Fall 2006.

[15] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida, 2005.

[16] S. Paper and M. A. Tchoshanov, The role of representation(s) in developing mathematical understanding, *Theory into Practice*, 40(2):118–127, 2001.

[17] C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, 1994.

[18] H. Papousek and M. Papousek, Mothering and the cognitive head-start: Psychobiological considerations. In H. R. Schaffer (Ed.), *Studies in Mother-Infant Interaction*, Academic Press, London, 1977.

[19] E. Remmel, The benefits of a long childhood, *American Scientist*, 96(3):250–251, May-June 2008.

[20] H. Stark and Y. Yang, *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets and Optics*, Wiley, New York, 1988.

[21] M. A. Tchoshanov, *Visual Mathematics*, ABAK Publ., Kazan, 1997 (in Russian).

[22] J. Van Patten, C. Chao, and C. Reigeluth, A review of strategies for sequencing and synthesizing instruction, *Review of Educational Research*, 56(4):437–471, 1986.

ISBN: 978-989-95079-6-8

Intensity assessment of pedestrian collisions in city of Mashhad based on fuzzy probabilities

Farzaneh Tatari and Mohammad-R. Akbarzadeh-T.

Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
Email: fa_tatari@yahoo.com, akbarzadeh@ieee.org

Abstract- *Fuzzy probabilities have been widely used in the areas of risk assessment and decision making. Here, we propose a system based on fuzzy probabilities for assessing the intensity of pedestrian collisions and levels of injury. The proposed approach uses possibility probability distributions for demonstrating fuzzy probabilities. A composition of possibility probability distributions (fuzzy probabilities) of collision intensities coming from different environmental variables yields the final fuzzy probability of intensity with higher reliability. The proposed approach is then applied to an actual database of pedestrian collisions from the City of Mashhad for over a period of five years.*

Keywords_ *Collision Intensity, Collision Variables, Fuzzy Probability, Possibility Probability Distribution.*

1 Introduction

Walking is a healthy and natural mean of transposition; it is therefore desirable that it also be safe. However, with the modern society and its rising crowd of transportation vehicles, this aspect of transposition has been sidelined and pedestrian collisions have caused significant social and financial losses every year. The safety of pedestrians during daily transposition is influenced by various variables. An appropriate paradigm for handling the effect of these variables can help insurance companies as well as city planners to better evaluate the safety of pedestrians in different situations and to plan the city environment for their improved safety. Unfortunately, recognizing all of the effective variables in a collision and their precise amount of effectiveness is non-trivial. The available information is highly imprecise and can lead to uncertain conclusions. With respect to the existing uncertainty in the collected information, the first source of data is known as "soft" data [17] that can be represented in linguistic form. This data is from expert opinions based on their background knowledge about the collision's intensity by considering related variables. Due to the linguistic form of this information and imprecision of data in this application, soft computing based methods, and in particular fuzzy-logic based analysis, offer a promising solution paradigm [18] to handle this type of uncertainty. In addition to soft data, the statistical historical data of pedestrian collisions are another valid data source which can be helpful in this case. The statistical data of pedestrian collisions may be incomplete when the number of observed data is not considerably large. Hence the statistical data are also accompanied by uncertainty which is caused by sparsity and insufficiency of data. We have found fuzzy probability framework as a suitable approach which enables us to enhance the reliability of our assessments by employing the both databases of soft data and statistical data at the same time. Fuzzy probabilities were first introduced by Zadeh (1984). Fuzzy probability theory [2] is a fuzzy approach to probability theory and is a generalized form of probability theory. In fuzzy probability, probability theory is complemented with an extra dimension of uncertainty provided by fuzzy set theory [16]. Generally we can divide the applications of fuzzy probabilities to two different main areas. The first is the area of reliability and risk assessment. In this area fuzzy probability has been widely applied in fuzzy fault trees [3,4,8] to assess the fault risk, reliability assessment for pressure piping [20], risk assessment of natural hazards [7] and reliability enhancement by combining expert opinions [14]. The second area is the area of the decision making. In this area, the fuzzy probabilities have been employed for decision making in perception-based theory [19], optimal decision fusion [11,12], inference by aggregation [9], information retrieval [6] and inventory control [1].

At any collision, there are two states for the pedestrian who has been hit by a car, in one state he would be just injured and in the other state he would die. Here, we employ fuzzy probabilities to handle the imprecision in expert opinions as the first dataset to assess the probability of injury and death for the pedestrians of Mashhad. We use fuzzy numbers to show these fuzzy probabilities. The second database is the statistical numeric data of pedestrian's collisions in Mashhad [10]. We combine these two insufficient databases to provide a more reliable result by the use of possibility probability distribution (PPD) which can model fuzzy probabilities in a suitable way.

2 Possibility probability distribution

The idea of representing assessments with PPD has been originally proposed by Haung [5]. This idea was improved by Karimi and Hullermeier [7] which was a new concept of fuzzy probability expressed by PPD. As we mentioned, on one hand, the statistical data in our type of application are typically insufficient, on the other hand background knowledge is often accompanied by imprecision. Therefore we apply the PPD to combine these two databases based on Bayesian approach. In the PPD approach, the prior knowledge (soft data) which is called prior distribution can be specified by a fuzzy number. $\overline{\theta}^{(i)}$ is an estimated probability by an expert for the state of death or injury in any collision where i is the variable that influences this probability and $p_i(\theta)$ is the prior distribution modeled in the form of symmetric, triangular fuzzy number, whose center is given by $\overline{\theta}^{(i)}$ (θ is the variable representing any probability in unit interval of [0,1]). As stated in [7] the support of the fuzzy number is of the form

ISBN: 978-989-95079-6-8

$$[\bar{\theta}^{(i)} - \bar{\theta}^{(i)}(1-\bar{\theta}^{(i)}).s \ , \bar{\theta}^{(i)} + \bar{\theta}^{(i)}(1-\bar{\theta}^{(i)}).s] \quad (1)$$

$$0 \le s \le \min(1/\bar{\theta}^{(i)}, 1/(1-\bar{\theta}^{(i)})) \quad (2)$$

the so-called uncertainty factor, is a constant that specifies the reliability of the point estimation. For combining prior knowledge with empirical data which is the available statistical numeric data, we first divide the unit interval of probabilities into m points, $\theta_j, j = 1,...,m$, and then calculate the posterior probability for each point θ_j as follows [7].

$$p'_i(\theta_j) = \binom{n}{n_i} \theta_j^{n_i} (1-\theta_j)^{n-n_i} . p_i(\theta_j) \approx$$

$$\frac{1}{c}(\gamma_j . q_j + (1-\gamma_j).q_{j+1}),$$

$$q_j = \theta_j^{n_i} (1-\theta_j)^{n-n_i} \, p_i(\theta_j) \quad (3)$$

where n_i is the number of observations that satisfies the features of the ith variable among n observations. $\gamma_j \in [0,1)$ is a unique value such that

$$\theta_j = \gamma_j \theta_j + (1-\gamma_j).q_{j+1} \quad (4)$$

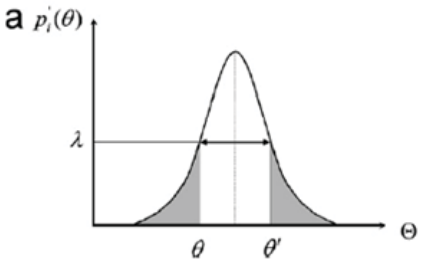
and c is a normalizing constant that guarantees

$$\int_0^1 p_i'(\theta) d\theta = 1 \quad (5)$$

Now is the time of transforming probability distribution to possibility distribution. Suppose a probability measure P on a set X is obtained via some statistical experiment. This probability function is a very rich piece of information, if the number of statistical experiments supporting it, is high enough. According to widely accepted consistence principle, the possibility measures should dominate P in the sense that $P(A) \le \pi(A)$ for all events $A \subseteq X$ and the maximally specific possibility distribution that exists must be unique [7]. As illustrated in [7] the most specific possibility distribution that approximates P from above (in the sense that $P \le \pi$) can be derived quite easily: for obtaining a possibility degree for any θ, $\pi_i(\theta)$, first $\lambda = p_i(\theta)$ is computed. Next, the second boundary point θ' that satisfies $p_i(\theta') = \lambda$ is found. Finally, the possibility degree $\pi_i(\theta)$ can be obtained as:

$$\pi_i(\theta) = \pi_i(\theta') = 1 - \left| \int_\theta^{\theta'} p'_i(x) dx \right|. \quad (6)$$

This possibility is equal to the area shaded grey in Fig. 1.a.



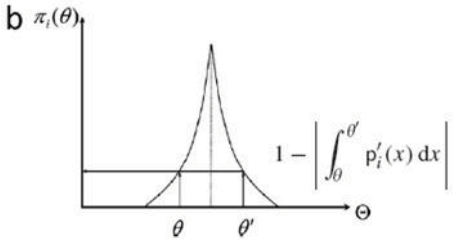


Figure 1: a. probability distribution to be transformed into a possibility distribution; b. transformed possibility distribution

3 Combining experts opinions and statistical data

We collected expert opinions (based on their own background knowledge and their reviews of traffic statics of death and injury in middle east countries [10]) about the percentage of collisions intensity (injury and death) by considering just one certain significant variable. The results of expert opinions are shown in Tables 1~5.

Table 1: Significant variable: hit organ

Injury probability if	Death probability if
Head and neck is hit, is approximately 0.09	Head and neck is hit, is approximately 0.55
Hand and arm is hit, is approximately 0.01	Hand and arm is hit, is approximately 0.001
Chest and stomach is hit, is approximately 0.2	Chest and stomach is hit, is approximately 0.2
Legs and pelvis is hit, is approximately 0.3	Legs and pelvis is hit, is approximately 0.03
More than one organ are hit, is approximately 0.5	More than one organ are hit, is approximately 0.3

Table 2: Significant variable: agility (crossing ability)

Injury probability for a pedestrian	Death probability for a pedestrian
Able to cross is 0.3	Able to cross is 0.4
Unable to cross is 0.6	Unable to cross is 0.6

Table 3: Significant variable: vehicle front shape

Injury probability if vehicle	Death probability if vehicle
Has bonnet is 0.7	Has bonnet is 0.4
Does not have bonnet is 0.2	Does not have bonnet is 0.2

Table 4: Significant variable: collision location

Injury probability if location is	Death probability if location is
In the city is 0.6	In the city is 0.3
Out of city is 0.04	Out of city is 0.05
In the suburb is 0.2	In the suburb is 0.15

Table 5: Significant variable: type of road

Injury probability if the road is a	Death probability if the road is a
Highway is 0.2	Highway is 0.6
Boulevard is 0.5	Boulevard is 0.6
Two_way road is 0.3	Two_way road is 0.2
incidental way is 0.1	incidental way is 0.05

With respect to the resulting experts ideas we obtained the fuzzy numbers for the proposed probability based on the approach illustrated in Section 2 where s is chosen as

$$s = 1/2\,(\min(1/\bar{\theta}^{(i)}, 1/(1-\bar{\theta}^{(i)}))) \quad (7)$$

It should be mentioned that we normalized $P(\bar{\theta}^{(i)})$ of all

ISBN: 978-989-95079-6-8

the prior probability distributions by factor N, to convert the prior probability distributions to fuzzy numbers.

$$N = \begin{cases} 2/(\overline{\theta}^{(i)}) & \overline{\theta}^{(i)} \le 0.5 \\ 2/(1-\overline{\theta}^{(i)}) & \overline{\theta}^{(i)} \ge 0.5 \end{cases} \quad (8)$$

As an example, through Figs. 2~3 we have shown the result of fuzzy probabilities of injury and death based on hit organ variable.

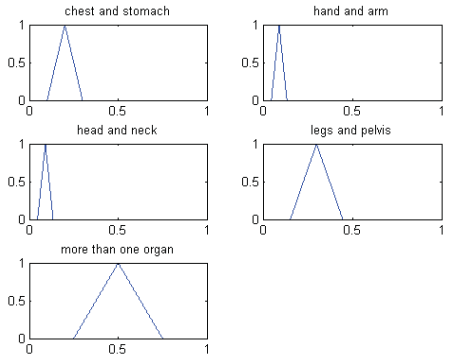


Figure 2: Injury fuzzy probability considering hit organ

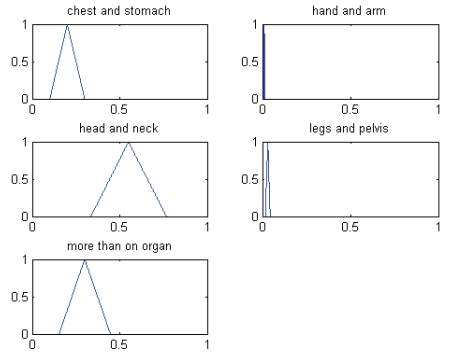


Figure 3: Death fuzzy probability considering hit organ

In the next step we apply (3) to obtain the posterior distribution based on statistical data which was gathered during one month in Shahidkamyab and Emamreza hospitals. In our database among 131 victims 99 of them had just injured and 32 of them had died. While injury and death are two independent states, we have considered the injuries which ends to fatality in the death statistical society. Therefore n=32 for dead pedestrians statistical society and n=99 for injured statistical society, these numbers are distributed through any variable partitions based on the collected data which is available in Tables 6 ~ 10. Among 18 factors which consist of human, vehicle, road and environmental factors, we could find 5 variables (hit organ, agility (crossing ability), vehicle front shape, location of collision and type of road) that had significant relations with collision intensity by the 5 percentage of error using SPSS.

Table 6: Distribution of statistical database on Significant variable of hit organ

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Head and neck = 10	Head and neck = 17
Hand and arm = 0	Hand and arm = 4
Chest and stomach = 5	Chest and stomach = 2
Legs and pelvis = 33	Legs and pelvis = 1
More than one point = 51	More than one point = 8

Table 7: Distribution of statistical database on Significant variable of agility (crossing ability)

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Able = 97	Able = 24
Unable =2	Unable = 8

Table 8: Distribution of statistical database on Significant variable of vehicle of front shape

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Has bonnet = 86	Has bonnet = 23
Does not have bonnet = 13	Does not have bonnet = 9

Table 9: Distribution of statistical database on Significant variable of location of the collision

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
In the city =60	In the city =14
Out of city =3	Out of city =6
Suburb =36	Suburb =12

Table 10: Distribution of statistical database on Significant variable vehicle of type of road

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Highway = 4	Highway = 7
Boulevard = 51	Boulevard = 10
Two way road = 34	Two way road = 11
Incidental way = 10	Incidental way = 4

The resulting posterior distributions of injury and death for the hit organ variable are as below:

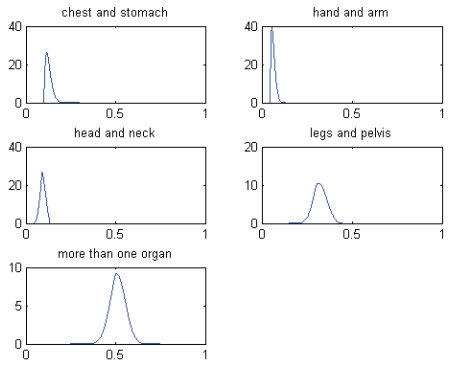


Figure 4: Posterior distribution of injury considering hit organ

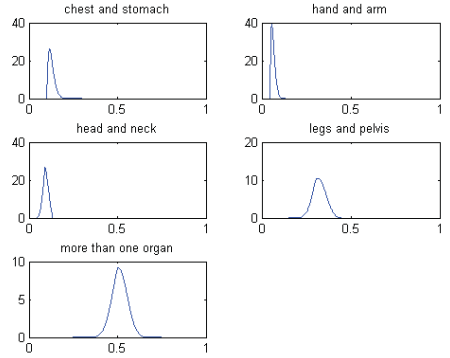


Figure 5: Posterior distribution of death considering hit organ

ISBN: 978-989-95079-6-8

Based on the procedure mentioned for transforming probability distribution to possibility distribution in Section 2, we obtained the possibility distributions of probability distributions. By comparing the prior fuzzy probabilities and possibility-probability distributions (Figs. 6~15), It can be easily interpreted that by combining the information of two databases less ambiguous, i.e. more certain and reliable results, are obtained in PPDs. It is also obvious that the PPD (fuzzy probability) can easily handle the existing uncertainty in the results.

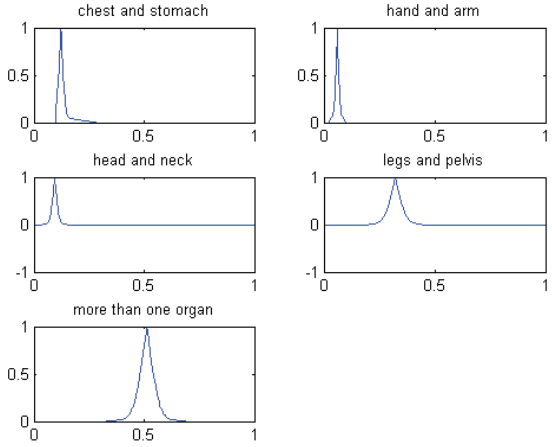


Figure 6: Possibility-probability distribution of injury considering hit organ

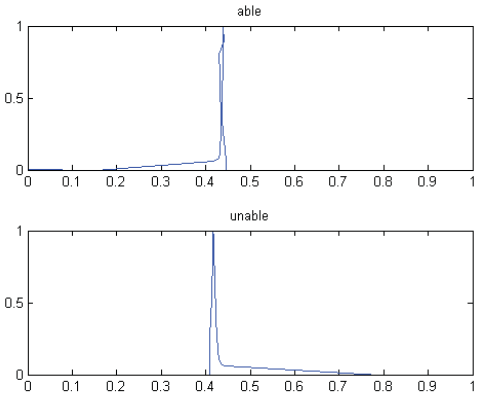


Figure 7: Possibility-probability distribution of injury knowing agility (crossing ability)

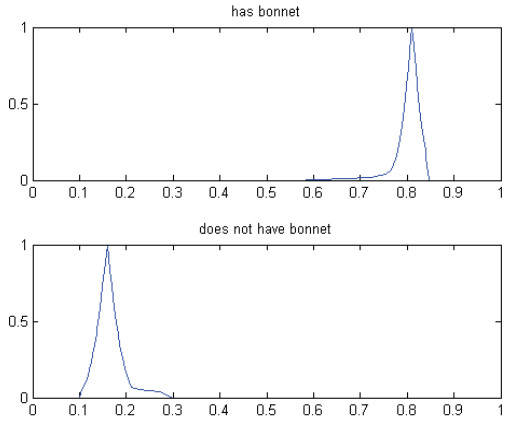


Figure 8: Possibility_probability distribution of injury considering vehicle front shape

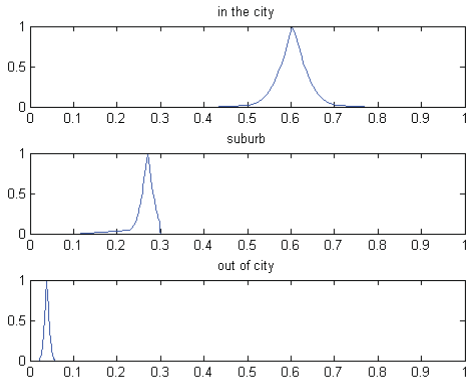


Figure 9: Possibility_probability distribution of injury considering location of collision

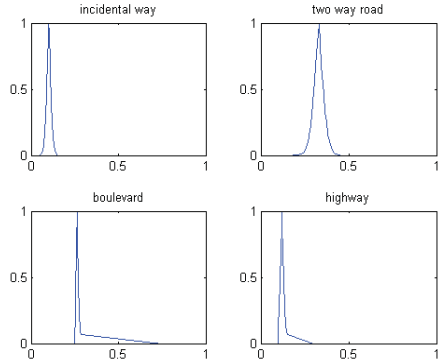


Figure 10: Possibility_probability distribution of injury considering type of road

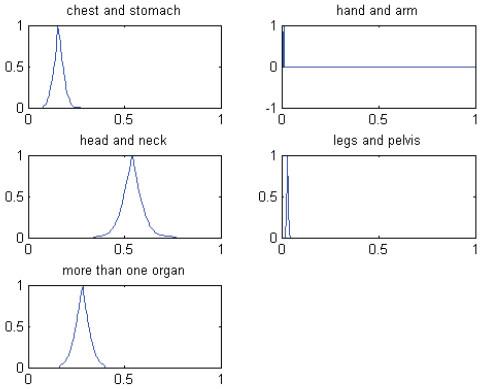


Figure 11: Possibility_probability distribution of death considering hit organ

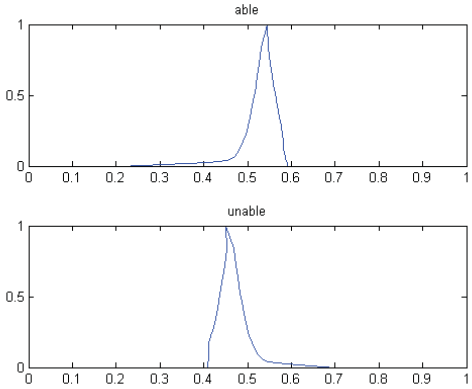


Figure 12: Possibility_probability distribution of death considering agility (crossing ability)

ISBN: 978-989-95079-6-8

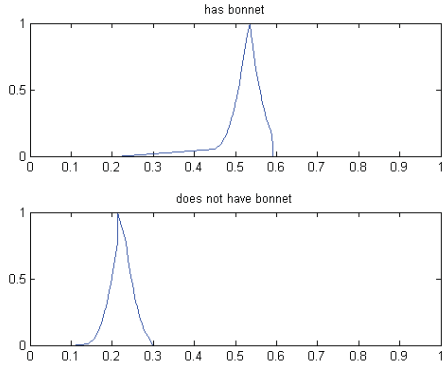


Figure 13: Possibility_probability distribution of death considering vehicle front shape

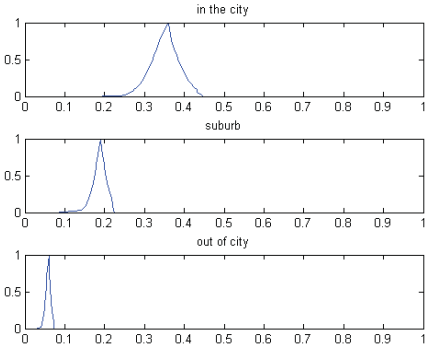


Figure 14: Possibility_probability distribution of death considering location of collision

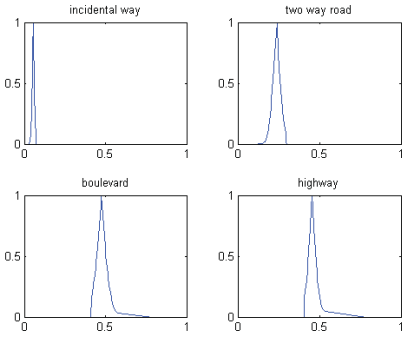


Figure 15: Possibility-probability distribution of death considering type of road

4 Collision intensity assessment

As it was explained before there are various variables which influence the intensity of collisions. If we assess the collision intensity based on one or few known variables the uncertainty of the assessment is so higher in comparison with the time that the assessment is based on more known variables. Therefore we can assess the fuzzy probability of injury or death of the pedestrians based on the combination of all significant variables PPDs. It should be mentioned that the employed method for combining fuzzy probabilities is based on fuzzy arithmetic [2] not Bayesian approach.

Assume that $\overline{A}$ and $\overline{B}$ are fuzzy probabilities of X. By having $\overline{A}$ and $\overline{B}$ which have been calculated based on different variables of X , we want to obtain $\overline{C}$ as a more reliable fuzzy probability of X by considering both variables, as the following

$$P[X(a)] \text{ is } \overline{A} \text{ and } P[X(b)] \text{ is } \overline{B} \equiv P[X(a,b)] \text{ is } \overline{C}.$$

Where a and b are related variables of X which effect assessing the fuzzy probability of X. One of the most common ways to aggregate fuzzy probabilities such as $\overline{A}$ and $\overline{B}$ is to obtain the α _cuts of $\overline{C}$ for any $\alpha \in [0,1]$ as below

$$\overline{A}[\alpha] = [a_1(\alpha), a_2(\alpha)], \overline{B}[\alpha] = [b_1(\alpha), b_2(\alpha)] \Rightarrow$$

$$\overline{C}(\alpha) = [\min(a_1(\alpha), b_1(\alpha)), \max(a_2(\alpha), b_2(\alpha))] \quad (9)$$

We can extend this procedure to more than two fuzzy probabilities with different variables to obtain a more reliable fuzzy probability of X, $\overline{C}$. It is also notable to know that any α _cut in any fuzzy probability is a sub interval of the unit interval of probabilities. However the mentioned method may exhibit shortcomings, first the max and min operations are noninteractive so they tend to lead to the results that are getting close to 1 and 0 respectively. Second, the result of combination at any level of α is entirely dependent upon the extreme bounds. To overcome these shortcomings we propose to calculate the α _cuts of $\overline{C}$ based on fuzzy arithmetic [2].

$$\overline{A}[\alpha] = [a_1(\alpha), a_2(\alpha)], \overline{B}[\alpha] = [b_1(\alpha), b_2(\alpha)] \Rightarrow$$

$$\overline{C}(\alpha) = [a_1(\alpha) \times b_1(\alpha) \ , \ a_2(\alpha) \times b_2(\alpha)] \quad (10)$$

The only drawback of fuzzy arithmetic method is that it results in low fuzzy probabilities which do not conform the reality (statistical data) but we can solve this problem by normalizing the obtained fuzzy probabilities of all the existing states. This normalization is in the way that for each $\alpha -$ cut , $\alpha \in [0,1]$, the summation of $\alpha -$ cut boundary points (probabilities) must be equal to one. In the specific application that we focus on it, we have two states of "X= death, Y= injury" in any collision, therefore we must normalize the final fuzzy probabilities in the way that

$$P'_{X1\alpha} = \frac{P_{X1\alpha}}{(P_{X1\alpha} + P_{Y2\alpha})}, \quad P'_{Y2\alpha} = \frac{P_{Y2\alpha}}{(P_{X1\alpha} + P_{Y2\alpha})},$$

$$P'_{X2\alpha} = \frac{P_{X2\alpha}}{(P_{X2\alpha} + P_{Y1\alpha})}, \quad P'_{Y1\alpha} = \frac{P_{Y1\alpha}}{(P_{X2\alpha} + P_{Y1\alpha})}$$

$$\text{then } P'_{X1\alpha} + P'_{Y2\alpha} = 1 \text{ and } P'_{X2\alpha} + P'_{Y1\alpha} = 1 \ \forall \alpha \in [0,1] \quad (11)$$

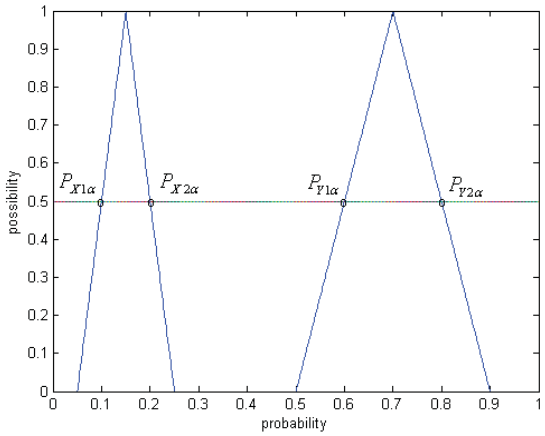


Figure 16: The probabilities which must be normalized at any α level

ISBN: 978-989-95079-6-8

We continue by an example where we used the fuzzy arithmetic procedure to obtain a more reliable fuzzy probability of injury and death based on the obtained results. In the following example for combining the fuzzy probabilities of injury and death which were assessed based on 5 significant variables, we use seven α _cuts (α = 0, 0.1, 0.3, 0.5, 0.7, 0.8, 1) of all fuzzy probabilities (PPDs) to do the combination.

Example: A pedestrian who does not have the agility (crossing ability) is crashed by a mini-bus while crossing a two-way road in city of Mashhad. His chest and stomach are hit by the vehicle. (a) What is the fuzzy probability of his injury? (b) What is the fuzzy probability of his death?

Hit organ: chest and stomach, agility (crossing ability): unable to cross, Vehicle front shape: does not have bonnet, Collision location: in the city, Type of road: two-way road. By combining the PPDs of the related variables we can obtain the final fuzzy probability of injury as below:

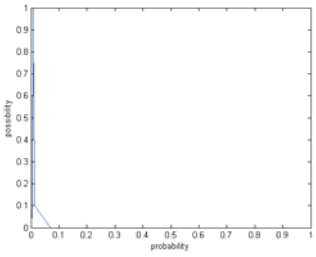


Figure 17: Fuzzy probability of injury before normalization

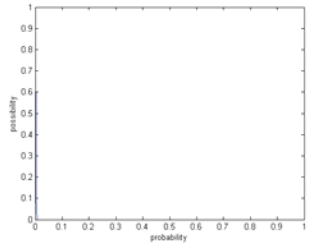


Figure 18: Fuzzy probability of death before normalization

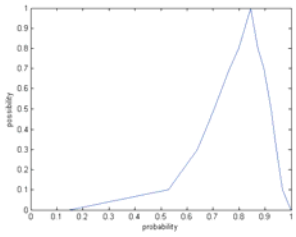


Figure 19: Normalized fuzzy probability of injury

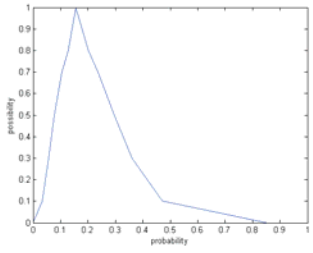


Figure 20: Normalized fuzzy probability of death

The normalized fuzzy probabilities are closer to statistical data and reality and the fuzzy probability of death and injury are complement of each other. The obtained result is more reliable while it has considered 5 significant variables at the same time instead of one.

5 Conclusion

Each year, too many pedestrians die all over the world with significant social and financial losses. The intensity of these collisions is of great importance for city planners and insurance companies alike. Here, we propose the use of fuzzy set theory to complement the probability theory with an additional dimension of uncertainty which leads to fuzzy probability characterized in terms of possibility probability distribution. We assessed the intensity of pedestrian collisions in Mashhad based on the variables that had significant relation with the intensity by use of fuzzy probability. The results indicate that by considering more significant variables in our assessment, we can decrease the amount of uncertainty and reach a more reliable result.

References

[1] J.J.Buckley et al., Fuzzy Mathematics in Economics and Engineering, Physica-Verlag, Heidelberg, Germany, 2002.
[2] J.J.Buckley, Fuzzy probabilities, new approach and applications, Springer, 2005.
[3] James P. Dunyak and Donald Wunsch, Fuzzy Probability for System Reliability, Proceedings of the 37th IEEE Conference on Decision & Control Tampa, 1998.
[4] James P. Dunyak et al., A Theory of Independent Fuzzy Probability for System Reliability, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 7, NO. 2, 1999.
[5] C. Huang, Fuzzy risk assessment of urban natural hazards, Fuzzy Sets and Systems 83 , 271–282, 1995.
[6] Z. Huang and T. D. Gedeon, Information retrieval estimation via fuzzy probability, World Automation Congress (WAC), 2006.
[7] I. Karimi and E. Hüllermeier, Risk assessment system of natural hazards: A new approach based on fuzzy probability, Fuzzy Sets and Systems 158, 987 – 999, 2007.
[8] R. Kenarangui, Event-Tree Analysis by Fuzzy Probability, IEEE TRANSACTIONS ON RELIABILITY, VOL. 40, NO. 1, 1991.
[9] A. L. Ralescu , D. A. Ralescu, Yoko Yamakata , Inference by aggregation of evidence with applications to fuzzy probabilities, Information Sciences 177, 378–387, 2007.
[10] M. Rezaei, estimation of collisions intensity for pedestrians in Mashhad using logistic regression, pp.92_107, 2004.
[11] G. Wang , S. Mao and Y. He, Optimal Decision Fusion With Fuzzy Priori Probabilities, Proceedings of IBP, 1998.
[12] G.Wang, S.Mao and Y. He, Distributed detection fusion with fuzzy priori probabilities and fuzzy cost function, IEEE INTERNATIONAL RADAR CONFERENCE, 2000.
[13] Ronald R. Yager, Decision Making with Fuzzy Probability Assessments, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 7, NO. 4, 1999.
[14] D. Yu and W. S. Park, Combination and evaluation of expert opinions characterized in terms of fuzzy probabilities, Annals of Nuclear Energy 27, 713-726, 2000.
[15] L. A. Zadeh, Fuzzy probabilities, Information processing and management 20(3), pp. 363_372, 198.
[16] L.A. Zadeh, Fuzzy sets , Information and control, 8(3), pp. 338_353, 1965.
[17] L.A. Zadeh, Posibility theory and soft data analysis, In: Cobb, L. and R. M. Thrall, eds., mathematical frontiers of the social and policy sciences. Westview press, Boulder, Colorado, pp.69_129, 1981.
[18] L.A. Zadeh, Soft computing and fuzzy logic, IEEE Softw are, 11 (6), pp.48_56, 1994.
[19] L. A. Zadeh, Toward a perception_based theory of probabilistic reasoning with imprecise probabilities, Journal of statistical planning and inference 105, 233-264, 2002.
[20] J. Zhou, Reliability assessment method for pressure piping containing circumferential defects based on fuzzy probability, International Journal of Pressure Vessels and Piping 82, 669–678, 2005.

ISBN: 978-989-95079-6-8

Number-free Mathematics Based on T-norm Fuzzy Logic

Libor Běhounek

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic
Email: behounek@cs.cas.cz

Abstract—*The paper presents a form of rendering classical mathematical notions by formal theories over suitable t-norm fuzzy logics in such a way that references to real numbers are eliminated from definitions and theorems, being removed to the standard semantics of fuzzy logic. Several examples demonstrate how this move conceptually simplifies the theory in exchange for non-classical reasoning, facilitates certain generalizations, and puts the concepts into a different perspective. The formal framework employed for the number-free formalization of mathematical concepts is that of higher-order fuzzy logic, also known as Fuzzy Class Theory.*

Keywords—Fuzzy Class Theory, fuzzy mathematics, fuzzy set, real-valued function, similarity-based limit, t-norm fuzzy logic.

1 Introduction

In the standard semantics of t-norm fuzzy logics [1], truth values are represented by real numbers from the unit interval $[0, 1]$; truth functions of n-ary propositional connectives are certain real-valued functions on $[0, 1]^n$; and the quantifiers $\exists, \forall$ are interpreted as the suprema and infima of sets of truth values. Read inversely, the apparatus of t-norm fuzzy logics expresses, by means of their standard semantics, certain first- and second-order constructions over real numbers. The axioms and rules of t-norm logics are designed to capture basic properties of such constructions and ensure the soundness of formally derivable theorems expressible in the language.

A suitable formal theory in first-order fuzzy logic thus can, by means of its standard semantics, express facts about classical mathematical notions and make them derivable by logical deductions in fuzzy logic. Since real numbers occur in the standard semantics of such a theory, they need no longer be explicitly mentioned by the theory itself. A notion of classical mathematics then becomes represented as the standard model of another notion of formal fuzzy mathematics that makes no explicit reference to real numbers: the reals are only implicitly present behind the logical axioms that govern reasoning in formal fuzzy mathematics.

This way of eliminating real numbers from the theory in favor of reasoning by means of t-norm fuzzy logic will here be called the *number-free* approach.

Number-free formalization of mathematical concepts is not new and has implicitly been around since the beginning of the theory of fuzzy sets: in fact, the notion of fuzzy set itself can be understood as a number-free rendering of the notion of real-valued function (see Section 2). However, it was only after the advancement of t-norm fuzzy logics, mostly in the past decade, that number-free notions could be treated rigorously in formal theories over first-order fuzzy logics. An early example of the number-free treatment of a classical notion is the formalization of finitely additive probability as a modality *Probable* in Łukasiewicz logic (see Section 3). Nevertheless, number-free rendering of more advanced mathematical notions requires more complex concepts of formal fuzzy mathematics—esp. higher-order set-constructions and a formal theory of fuzzy relations. The latter prerequisites have only recently been developed in the framework of higher-order fuzzy logic [2, 3, 4], which made it possible to apply the number-free approach systematically to various classical mathematical notions.

The profit we gain under the number-free approach in exchange for having to use non-classical rules of reasoning is, in the first place, conceptual simplification (roughly speaking, we get 'a set for a function'). Secondly, the number-free rendering often reveals a new perspective upon the notion, exposing the gradual quality of the classical construct, treating its gradualness as a primitive rather than derivative feature, and telling a different story about the concepts than is told by classical mathematics. Thirdly, many theorems of classical mathematics are under this approach detected as provable by simple (often, propositional) logical derivations in a suitable fuzzy logic, instead of complex classical proofs involving arithmetic, infima, functions, etc. Furthermore, adopting a non-standard semantics (e.g., using Chang's MV-algebra instead of the standard real numbers) or a different interpretation of the logical symbols involved (e.g., taking another t-norm for conjunction) yields an effortless generalization that might be harder to find (and motivate) in the classical language of crisp mathematics. Finally, the many-valuedness of all formulae in fuzzy logic makes it possible to consider another kind of graded generalization, by admitting partial satisfaction of the axioms for the represented notion (e.g., a metric to degree .99, cf. [5, 6]).

The particular formal framework in which number-free formalization of classical mathematical concepts is carried out in this paper is that of higher-order fuzzy logic, also known as Fuzzy Class Theory (FCT), over a suitable propositional fuzzy logic at least as strong as $\mathrm{MTL}_\triangle$. A working knowledge of FCT will be assumed throughout; for an introduction to the theory and more information see [2, 5]. For reference, the definitions used in the present paper are repeated below in the Appendix.

The aim of this paper is only to introduce the number-free approach as a distinct paradigm of formalization, rather than to develop particular number-free theories in depth. Therefore it only gives definitions of and a few observations on several number-free notions and discusses the merits of such a formalization; a detailed investigation of number-free theories is left for future work. A slightly more eloquent version of this paper has been made available as a research report [7].

ISBN: 978-989-95079-6-8

2 Real-valued functions

The very first notion of fuzzy mathematics, namely Zadeh's notion of fuzzy set [8], can be regarded as a number-free representation of classical $[0,1]$-valued functions by non-classical (namely, fuzzy) sets. Even though the formal apparatus of first-order fuzzy logic, which makes it possible to cast fuzzy sets as a primitive notion instead of representing them by classical real-valued functions, was developed years later, the tendency of regarding fuzzy sets and relations as a number-free rendition of real-valued functions has partly been present since the very beginning of the fuzzy set theory, as witnessed by the vocabulary and notation employed. E.g., the function $x \mapsto \min(A(x), B(x))$ is in the traditional fuzzy set theory denoted by $A \cap B$ and called the *intersection* of A and B: that is, the functions $A, B\colon X \to [0,1]$ are regarded as (non-classical) *sets* rather than real-valued functions (as the intersection of real-valued functions is a different thing). Similarly, such notions as fuzzy relational composition or the image under a fuzzy relation would make little sense if the n-ary functions involved were not regarded as (a kind of non-classical) relations. The terminological shift towards the number-free discourse is expressed by the very term "fuzzy set" and by its informal motivation of an unsharp collection of elements.

A certain part of the talk about *real-valued functions* and their properties was thus replaced by a talk about *sets* and *relations* that behave non-classically (e.g., do not follow the rule of excluded middle). This move eliminated references to numbers at least from the wording of some theorems, giving them compact forms and new conceptual meanings.[1] Clearly, the original number-free notion of fuzzy set has proved immensely fruitful even in its semi-formal form of traditional fuzzy mathematics. The formal apparatus of logic-based fuzzy mathematics has provided means for accomplishing the long-present idea and developing a fully fledged number-free approach to fuzzy sets and fuzzy relations.[2]

3 Finitely additive probability measures

Another number-free representation, already based on formal fuzzy logic, was the axiomatization of finitely additive probability measures as models of a fuzzy modality *Probable* over propositional Łukasiewicz logic by Hájek, Godo, and Esteva [10]. Later it was elaborated in a series of papers by Flaminio, Marchioni, Montagna, and the authors of [10]. We shall briefly recapitulate the original axiomatization (adapted from [11]) as another illustration of the number-free approach.

Consider a classical probability space $(\Omega, \mathcal{B}, \pi)$, where Ω is a set of elementary events, $\mathcal{B}$ a Boolean algebra of subsets of Ω, and π a finitely additive probability measure on $\mathcal{B}$, i.e., a function $\pi\colon \mathcal{B} \to [0,1]$ satisfying the following conditions:

$$\pi(\Omega) = 1 \tag{1}$$

$$\text{If } A \subseteq B, \text{ then } \pi(A) \leq \pi(B) \tag{2}$$

$$\text{If } A \cap B = \emptyset, \text{ then } \pi(A \cup B) = \pi(A) + \pi(B) \tag{3}$$

A number-free representation of π draws on the fact that a $[0,1]$-valued function on a Boolean algebra can be understood as a standard model of a fuzzy modality P over an algebra of crisp propositions. The above conditions on π can be transformed into the axioms for P, which (due to the additivity) are expressible in Łukasiewicz logic:[3]

Definition 3.1. The axioms and rules of the logic $\mathrm{FP}(\mathbf{Ł})$ are those of Łukasiewicz propositional logic plus the following axioms and rules, for non-modal φ, ψ:

$$\varphi \vee \neg\varphi \tag{4}$$

$$\text{From } \varphi \text{ infer } P\varphi \tag{5}$$

$$P\varphi, \text{ for all Boolean tautologies } \varphi \tag{6}$$

$$P(\varphi \to \psi) \to (P\varphi \to P\psi) \tag{7}$$

$$P(\neg\varphi) \leftrightarrow \neg P\varphi \tag{8}$$

$$P(\varphi \vee \psi) \leftrightarrow ((P\varphi \to P(\varphi \wedge \psi)) \to P\psi) \tag{9}$$

The axioms and rules of $\mathrm{FP}(\mathbf{Ł})$ ensure the following representation theorem (adapted from [10]):

Theorem 3.2. *Any probability space $(\Omega, \mathcal{B}, \pi)$ is a standard model of $\mathrm{FP}(\mathbf{Ł})$. Vice versa, all standard models of $\mathrm{FP}(\mathbf{Ł})$ are probability spaces.*

The representation theorem shows that the number-free theory faithfully captures the original notion of finitely additive probability measure. Moreover, by the completeness theorem of $\mathrm{FP}(\mathbf{Ł})$ w.r.t. probability spaces, proved in [11], all valid laws of finitely additive probability that are expressible in the language of $\mathrm{FP}(\mathbf{Ł})$ can in $\mathrm{FP}(\mathbf{Ł})$ be also (number-freely) proved.

In a given probability space $(\Omega, \mathcal{B}, \pi)$, i.e., a standard model of $\mathrm{FP}(\mathbf{Ł})$ with $\|P\| = \pi$, the truth value of $P\varphi$ is the probability of the event φ: $\|P\varphi\| = \pi(\|\varphi\|)$; the formula $P\varphi$ can therefore be understood as "φ is probable". Numerical calculations with probabilities are thus in $\mathrm{FP}(\mathbf{Ł})$ replaced by *logical derivations* with the modality "is probable". The key difference is that the latter represent inference *salva probabilitate* (i.e., salvo *probabilitatis gradu,* in the sense of P): e.g., it can be observed that numberless probability is transmitted by modus ponens, as $P\varphi \,\&\, P(\varphi \to \psi) \to P\psi$, i.e., "if φ is probable and $\varphi \to \psi$ is probable, then ψ is probable", is a theorem of $\mathrm{FP}(\mathbf{Ł})$.

The number-free approach to probability facilitates several kinds of generalization. First, generalizations to models over non-standard algebras for Łukasiewicz logic: thus we can have, e.g., probability valued in Chang's MV-algebra, or in non-standard reals (as in [12]). Second, a generalization to measures with only partially satisfied additivity (by a many-valued interpretation of the axioms, see [6]). And third, a generalization to the probability of fuzzy events, where one discards the axiom $\varphi \vee \neg\varphi$ for events and adapts the finite additivity axiom to work well with fuzzy events (e.g., as $P(\varphi \oplus \psi) \leftrightarrow ((P\varphi \to P(\varphi \,\&\, \psi)) \to P\psi)$).[4]

[1] The elimination of numbers also from proofs would have required a consistent use of first-order fuzzy logic. This approach was not embraced in the early works on fuzzy set theory, even though particular first-order fuzzy logics already existed, e.g., [9].

[2] Cf. [2, 3], where the formal theory of fuzzy sets and relations is developed without making any reference to real numbers in definitions, theorems, or proofs (only in explanatory semantic examples).

[3] Ł with rational truth constants is used in [10], but the truth constants are inessential for our account. The language is two-layered, admitting only non-modal formulae and propositional combinations of non-nested modal formulae.

[4] This approach has been taken in [13], though only over finitely-valued events, as the authors strove for the completeness of the logic; in [12] this was generalized to infinitely-valued events, with completeness w.r.t. non-standard reals.

ISBN: 978-989-95079-6-8

4 Distribution functions

Classical distribution functions present a special way how to define a probabilistic measure on Borel sets, i.e., on the σ-algebra $\mathcal{B}$ of subsets of the real line generated by all intervals $(-\infty, x]$. A function $f\colon \mathbb{R} \to [0,1]$ defines a measure on $\mathcal{B}$ with $\mu(-\infty, x] = f(x)$ and $\mu(\mathbb{R}) = 1$ iff it satisfies the following conditions, which can thus be taken as the axioms for distribution functions:

1. *Monotony:* if $x \leq y$ then $f(x) \leq f(y)$, for all $x, y \in \mathbb{R}$
2. *Margin conditions:* $\lim_{x\to-\infty} f(x) = 0$, $\lim_{x\to+\infty} f(x) = 1$
3. *Right-continuity:* $\lim_{x\to x_0^+} f(x) = f(x_0)$ for all $x_0 \in \mathbb{R}$

Let us translate these conditions into the number-free language.[5] The function $f\colon \mathbb{R} \to [0,1]$ represents a standard fuzzy set of reals, i.e., in FCT over any expansion of MTL_Δ, the standard model of a predicate A on reals.[6] By the standard semantics of MTL_Δ, the above conditions translate into the following axioms on the predicate A:[7]

$$(\forall xy)(x \leq y \to (Ax \to Ay)) \tag{10}$$

$$\neg(\forall x)Ax,\ (\exists x)Ax \tag{11}$$

$$(\forall x_0)[(\forall x > x_0)Ax \to Ax_0] \tag{12}$$

In FCT, these conditions express, respectively, the *upperness* of A in $\mathbb{R}$, the *null plinth* and *full height* of A, and the *left-closedness* of the fuzzy upper class A. Consequently, number-free distribution functions are left-closed upper sets in $\mathbb{R}$ with full height and null plinth, i.e., (weakly bounded) *fuzzy Dedekind cuts on* $\mathbb{R}$. The number-free rendering of distribution functions as fuzzy Dedekind cuts corresponds to the known fact that distribution functions represent Hutton fuzzy reals (cf., e.g., [15]). A use of fuzzy Dedekind cuts for the development of a logic-based theory of fuzzy intervals (or fuzzy numbers) is hinted at in [16].

5 Continuous functions on $\mathbb{R}$

In Section 4 we abused the presence of monotony for the number-free rendering of right-continuity. Yet, if we aim at a graded theory where monotony can be satisfied to partial degrees, we need a different number-free characterization of continuity that does not rely on monotony. We shall work with functions $\mathbb{R} \to [0,1]$ only, even though various generalizations are easy to obtain.

For a number-free rendering of right-continuity, we shall use the following classical characterization. A function $f\colon \mathbb{R} \to [0,1]$ is right-continuous in x_0 iff

$$\limsup_{x\to x_0^+} f(x) = \liminf_{x\to x_0^+} f(x) = f(x_0), \tag{13}$$

where

$$\limsup_{x\to x_0^+} f(x) = \inf_{x_1>x_0} \sup_{x_1>x>x_0} f(x) \tag{14}$$

$$\liminf_{x\to x_0^+} f(x) = \sup_{x_1>x_0} \inf_{x_1>x>x_0} f(x). \tag{15}$$

This translates into the following number-free definitions in FCT over MTL_Δ:[8]

$$\mathrm{LimSup}^+(A, x_0) \equiv_{\mathrm{df}} (\forall x_1 > x_0)(\exists x \in (x_0, x_1))Ax \tag{16}$$

$$\mathrm{LimInf}^+(A, x_0) \equiv_{\mathrm{df}} (\exists x_1 > x_0)(\forall x \in (x_0, x_1))Ax \tag{17}$$

$$\Delta\,\mathrm{Cont}^+(A, x_0) \equiv_{\mathrm{df}} (Ax_0 = \mathrm{LimSup}^+(A, x_0))\ \&\ (Ax_0 = \mathrm{LimInf}^+(A, x_0)) \tag{18}$$

The left-sided predicates LimSup^- and LimInf^- are defined dually (with $<$ for $>$), and the both-sided ones as

$$\mathrm{LimSup}(A, x_0) \equiv_{\mathrm{df}} \mathrm{LimSup}^-(A, x_0) \vee \mathrm{LimSup}^+(A, x_0) \tag{19}$$

$$\mathrm{LimInf}(A, x_0) \equiv_{\mathrm{df}} \mathrm{LimInf}^-(A, x_0) \wedge \mathrm{LimInf}^+(A, x_0). \tag{20}$$

These definitions reconstruct the classical notions in a number-free way in the framework of FCT; the representation theorems follow directly from the above considerations and the standard semantics of MTL_Δ. The following observation shows that many properties of the classical notions can be reconstructed in FCT as well.

Observation 5.1. By shifts of crisp relativized quantifiers valid in first-order MTL_Δ, the following theorems are easily provable in FCT:[9]

1. $\mathrm{LimInf}(A, x_0) \leq \mathrm{LimSup}(A, x_0)$
2. $A \subseteq B \to (\mathrm{LimSup}(A, x_0) \to \mathrm{LimSup}(B, x_0))$ and analogously for LimInf.
3. $\mathrm{LimInf}(\smallsetminus A, x_0) \leq \neg\,\mathrm{LimSup}(A, x_0)$
   $\mathrm{LimSup}(\smallsetminus A, x_0) \geq \neg\,\mathrm{LimInf}(A, x_0)$
   (Equality holds in logics with involutive negation, but not generally in MTL_Δ.)
4. $\mathrm{LimInf}(A \sqcap B, x_0) = \mathrm{LimInf}(A, x_0) \wedge \mathrm{LimInf}(B, x_0)$
   $\mathrm{LimInf}(A \sqcup B, x_0) = \mathrm{LimInf}(A, x_0) \vee \mathrm{LimInf}(B, x_0)$
   and analogously for LimSup.

Since the definitions reconstruct classical notions, we have retained the classical terminology and notation referring to limits and continuity, even though these regard membership functions (i.e., semantic *models* of fuzzy classes), rather than fuzzy classes themselves. In FCT, the fuzzy predicate $\mathrm{LimSup}^+(A, x_0)$ actually expresses the condition that x_0 is a *right-limit point* of the fuzzy class A, and $\mathrm{LimInf}^+(A, x_0)$ that x_0 is an *interior point* of $A \cup \{x_0\}$ in the left half-open interval topology,[10] as these are the properties expressed by the defining formulae if all sets involved are crisp.

[5] Of course, the 'translation' is by no means unique: we always select one which is sufficiently straightforward and which results in *meaningful* concepts of fuzzy mathematics.

[6] Recall that $\mathbb{R}$ as well as other crisp mathematical structures are available in FCT by means of the Δ-interpretation, see [2, §7] and [14, §4]. In this and the next section, we shall understand all first-order quantifications relativized to $\mathbb{R}$, unless specified otherwise.

[7] We use the fact that due to the monotony assumed, the margin conditions reduce to $\inf_x f(x) = 0$ and $\sup_x f(x) = 1$, and the right-continuity to $f(x_0) \geq \inf_{x>x_0} f(x)$. The representation theorem is then immediate by the standard semantics of MTL_Δ. Observe that as the axioms are required to degree 1, they are (due to the crispness of $\leq$) independent of the particular left-continuous t-norm used.

[8] Again it can be observed that the definitions are independent of a particular t-norm and are the same in all expansions of MTL_Δ.

[9] The theorems are stated for both-sided limits only, but hold equally well for one-sided limits.

[10] I.e., the topology with the open base of all half-open intervals $[a, b)$, also known as the lower-limit topology or the Sorgenfrey line.

ISBN: 978-989-95079-6-8

Consequently, the formulae $(\forall x_0 \in A)\,\mathrm{LimInf}(A, x_0)$ and $(\forall x_0 \in A)\,\mathrm{LimSup}(A, x_0)$ express the notions of *openness* resp. *closedness* of A in a fuzzy interval topology on $\mathbb{R}$. The study of this fuzzy topology and its relationship to the fuzzy interval topologies of [17, 18] is left for future work.

Like the definitions of LimInf and LimSup, the theorems of Observation 5.1 have double meanings, too. On the one hand they can be understood as number-free reconstructions and graded generalizations of the classical theorems on (membership) functions. On the other hand, they can be interpreted as fuzzy-mathematical theorems on (fuzzy) sets, under the above fuzzy interval topology on reals. In particular, 1. says that an interior point of a fuzzy set of reals is also its limit point; 2. that a limit point of a fuzzy set is also a limit point of a larger fuzzy set (and dually for interior points); 3. that an interior point of the complement of a fuzzy set A is not a limit point of A (and vice versa); and 4. that x_0 is a limit point of $A \sqcap B$ exactly to the degree it is a limit point of A *and* ($\wedge$) a limit point of B (and dually for $\sqcup$). (The theorems are graded, thus 'is' represents fuzzy implication $\rightarrow$.)

6 Operations on $\mathbb{R}$

In the previous examples, only the codomain of real-valued functions of reals was rendered numberless. Obviously, the domain $\mathbb{R}$ (or more conveniently, $\mathbb{R} \cup \{\pm\infty\}$) can be re-scaled into $[0, 1]$ and regarded as the standard set of truth degrees as well. The functions $\mathbb{R}^n \rightarrow \mathbb{R}$ then become n-ary functions from truth values to truth values, i.e., truth functions of fuzzy propositional connectives.

An apparatus for internalizing truth values and logical connectives in FCT was developed in [4, §3]. As shown there, the truth values can be internalized as the elements of the crisp class $\mathrm{L} = \mathrm{Ker}\,\mathrm{Pow}\{a\}$, i.e., subclasses of a fixed crisp singleton. The class L of internalized truth values is ordered by crisp inclusion $\subseteq^{\Delta}$, and the correspondence between internal and semantical truth values is given as follows:[11] $\alpha \in \mathrm{L}$ represents the semantic truth value of $\emptyset \in \alpha$, and the semantic truth value of φ is represented by the class $\overline{\varphi} =_{\mathrm{df}} \{a \mid \varphi\}$; the correspondences $\varphi \leftrightarrow (a \in \overline{\varphi})$ and $\overline{\varphi} \subseteq \overline{\psi} \leftrightarrow (\varphi \rightarrow \psi)$ then hold.

Logical connectives are then internalized by crisp functions $\mathbf{c}\colon \mathrm{L}^n \rightarrow \mathrm{L}$ (which can be called *internal, inner,* or *formal* connectives). In particular, definable connectives c of the logic are represented by the corresponding class operations $\overline{c} = \{x \in \mathrm{L} \mid c(x \in X_1, \ldots, x \in X_n)\}$ on L (e.g., & by $\cap$; $\vee$ by $\sqcup$; etc.). Since n-ary internal connectives are crisp functions valued in L, they can as well be regarded as fuzzy subsets of L^n, i.e., n-ary fuzzy relations on L. Usual fuzzy class operations and graded predicates, e.g., the graded inclusion

$$\mathbf{c} \subseteq \mathbf{d} \equiv_{\mathrm{df}} (\forall x_1 \ldots x_n)(\mathbf{c}x_1 \ldots x_n \rightarrow \mathbf{d}x_1 \ldots x_n), \quad (21)$$

then apply to them and make their theory graded.

The number-free theory of functions $\mathbb{R}^n \rightarrow \mathbb{R}$ is thus in fact the fuzzy-logical theory of internal connectives, i.e., of fuzzy relations on internal truth values. An elaboration of the theory of unary and binary internal connectives has been sketched in [19, 20]. These preliminary papers focus on the defining properties of *t-norms* (i.e., monotony, commutativity, associativity, and the unit) and the relation of domination between internal connectives, making them graded by reinterpretation of their defining formulae in fuzzy logic (cf. [2, §7], [5, §2.3], or [6, §4]) and studying their graded properties. A full paper on the topic (by the authors of [20]) is currently under construction.

7 Metrics

Recall that a *pseudometric*[12] on a set X is a function $d\colon X^2 \rightarrow [0, +\infty]$ such that

$$d(x, x) = 0 \quad (22)$$
$$d(x, y) = d(y, x) \quad (23)$$
$$d(x, z) \leq d(x, y) + d(y, z). \quad (24)$$

The numberless reduction will first need to normalize the range of pseudometrics from $[0, +\infty]$ to $[0, 1]$, e.g., by setting

$$c(x, y) = 2^{-d(x,y)}. \quad (25)$$

The defining conditions on pseudometric then become the following equivalent conditions on c:[13]

$$c(x, x) = 1 \quad (26)$$
$$c(x, y) = c(y, x) \quad (27)$$
$$c(x, z) \geq c(x, y) \cdot c(y, z) \quad (28)$$

These conditions are nothing else but the defining conditions of *fuzzy equivalences,* also known as *similarity relations* [21], in the standard semantics of product fuzzy logic [1]. We can thus equate number-free pseudometrics with *product similarities,* i.e., standard models of the following axioms in product fuzzy logic:

$$Cxx \quad (29)$$
$$Cxy \rightarrow Cyx \quad (30)$$
$$Cxy \,\&\, Cyz \rightarrow Cxz \quad (31)$$

The definition of a *metric* strengthens the first condition to $d(x, y) = 0$ iff $x = y$, which is equivalent to c being a *fuzzy equality* (also called *separated* similarity), i.e., $c(x, y) = 1$ iff $x = y$, thus replacing the first axiom by $Cxy \leftrightarrow x = y$.

Using a different left-continuous t-norm represents a different way of combining the distances $d(x, y)$ and $d(y, z)$ in the triangle inequality: e.g., with the minimum t-norm, c represents a *(pseudo)ultrametrics* under the same transformation (25), while with the Łukasiewicz t-norm, c represents a *bounded* pseudometric d under a different transformation $c(x, y) = (1 - d(x, y))/d_{\max}$, where $d_{\max} < +\infty$ is an upper bound on the distances. (The obvious both-way representation theorems are left to the reader.) Since furthermore many theorems on number-free metrics hold generally over MTL_{Δ}, it is quite natural to generalize the notion of number-free metrics to any similarity, not only the product one.

Various notions based on such (generalized) number-free metrics can be defined and their properties investigated in the framework of FCT (over MTL_{Δ} or stronger). Only a few

[11] However, see [4, Rem. 3.3] for certain metamathematical qualifications regarding this correspondence.

[12] For simplicity, we shall work with *extended* pseudometrics, allowing the value $+\infty$.

[13] Notice that since the function 2^{-x} reverses the order, the fuzzy relation $c\colon X^2 \rightarrow [0, 1]$ expresses *closeness* rather than distance.

ISBN: 978-989-95079-6-8

observations on number-free limits are given here as a concluding illustration of the numberless approach.

Fix a metric d rendered in the numberless way by a closeness predicate C under the transformation (25). The *limit* $\lim_{n\to\infty} x_n$ of a sequence $\{x_n\}_{n\in\mathbb{N}}$ (abbreviated $\vec{x}$) under C can be defined as follows:[14]

$$\Delta\mathrm{Lim}_C(\vec{x},x) \equiv_{\mathrm{df}} \Delta(\exists n_0)(\forall n > n_0)Cxx_n \qquad (32)$$

Theorem 7.1. *Standard models over product logic validate* $\Delta\mathrm{Lim}_C(\vec{x},x)$ *iff* $x = \lim_{n\to\infty} x_n$ *under* d.

Proof: $\lim x_n = x$ under d iff $\limsup d(x,x_n) = 0$, iff $\liminf 2^{-d(x,x_n)} = 1$, iff $\sup_{n_0} \inf_{n>n_0} c(x,x_n) = 1$, which is the semantics of $\Delta\mathrm{Lim}_C(\vec{x},x)$. □

As noted above, the meaning of $\Delta\mathrm{Lim}_C$ is natural not only in product logic, but also in other t-norm logics, esp. if the relation C is interpreted as *indistinguishability* rather than mere closeness:[15] then (32) expresses the condition that *from some index on, x_n is indistinguishable from x*. Similarly, Theorem 7.2 below expresses the fact that all limits of $\vec{x}$ are indistinguishable (to the degree the indistinguishability relation is symmetric and transitive).

Discarding the Δ in (32) yields a graded number-free notion of limit:[16]

$$\mathrm{Lim}_C(\vec{x},x) \equiv_{\mathrm{df}} (\exists n_0)(\forall n > n_0)Cxx_n \qquad (33)$$

Interestingly, $\mathrm{Lim}_C(\vec{x},x)$ coincides with G. Soylu's notion of *similarity-based fuzzy limit* [23]. Even without employing explicitly the formalism of t-norm fuzzy logic, Soylu was able to prove graded theorems such as [23, Prop. 3.5],

$$\mathrm{Lim}_C(\vec{x},x)\ \&\ \mathrm{Lim}_C(\vec{y},y) \to \mathrm{Lim}_C(\vec{x}+\vec{y},x+y). \qquad (34)$$

With the apparatus of FCT, the gradedness of Soylu's results can be extended even further by not requiring the full satisfaction of the defining properties of the similarity C (this conforms to the standard methodology of constructing graded theories [6, §7]). An example of such graded results is the following theorem on the fuzzy uniqueness of the limit:

Theorem 7.2. FCT *over* MTL_Δ *proves:*

$$\mathrm{Sym}\,C\ \&\ \mathrm{Trans}\,C\ \&\ \mathrm{Lim}_C(\vec{x},x)\ \&\ \mathrm{Lim}_C(\vec{x},x') \to Cxx' \qquad (35)$$

Proof. By $\mathrm{Trans}\,C$ we obtain $Cxx_n\ \&\ Cx_nx' \to Cxx'$; thus $Cx_nx\ \&\ Cx_nx' \to Cxx'$ by $\mathrm{Sym}\,C$, whence

$$((n > n_0) \to Cx_nx)\ \&\ ((n > n_0) \to Cx_nx') \to Cxx' \qquad (36)$$

follows propositionally. By generalization on n and distribution of the quantifier,

$$(\forall n > n_0)Cx_nx\ \&\ (\forall n > n_0)Cx_nx' \to Cxx' \qquad (37)$$

is obtained (as in the consequent the quantification is void). Generalization on n_0 and the shift of the quantifier to the antecedent (as $\exists$) then yields the required formula. □

[14] The Δ in $\Delta\mathrm{Lim}_C$ refers to the Δ in the defining formula, which will later be dropped.

[15] Cf. Menger's probabilistic indistinguishability relations [22].

[16] Observe that Lim_C is a Σ_2-formula: compare it with the classical Π_3-definition and the Π_1-definition in non-standard analysis.

A more detailed investigation of convergence based on fuzzy indistinguishability in the formal framework of FCT exceeds the scope of the present paper, and is therefore left for future research.

Appendix: Fuzzy Class Theory

Fuzzy Class Theory (FCT) is a formal theory over a given first-order fuzzy logic aiming at axiomatic approximation of Zadeh's fuzzy sets of all orders over a fixed crisp domain. It can be characterized as Henkin-style higher-order fuzzy logic, or fuzzified Russell-style simple type theory. FCT can be regarded as a foundational theory for fuzzy mathematics [24], since other axiomatic mathematical theories over fuzzy logic can be formalized within its framework. For more details on FCT see [2, 5]; the relevant definitions of [5] are briefly repeated here for reference.

The reader's familiarity with the logic MTL_Δ and its main extensions is assumed; for details on these logics see [1, 25]. Here we only recapitulate its standard real-valued semantics, which is crucial for number-free mathematics:

$\&$	...	a left-continuous t-norm $*$
$\to$	...	the residuum $\Rightarrow$ of $*$, defined as $x \Rightarrow y =_{\mathrm{df}} \sup\{z \mid z * x \le y\}$
$\wedge, \vee$	...	$\min, \max$
$\neg$	...	$x \Rightarrow 0$
$\leftrightarrow$	...	the bi-residuum: $\min(x \Rightarrow y, y \Rightarrow x)$
Δ	...	$\Delta x = 1 - \mathrm{sgn}(1-x)$
$\forall, \exists$	...	$\inf, \sup$

Łukasiewicz logic further specifies $x * y = (x + y - 1) \vee 0$, product logic sets $x * y = x \cdot y$, and Gödel logic sets $x * y = x \wedge y$.

Fuzzy Class Theory FCT is a formal theory over a given multi-sorted first-order fuzzy logic $\boldsymbol{L}$ (at least as strong as MTL_Δ), with sorts of variables for: atomic objects (lowercase letters $x, y, \ldots$), fuzzy classes of atomic objects (uppercase letters $A, B, \ldots$), fuzzy classes of fuzzy classes of atomic objects (calligraphic letters $\mathcal{A}, \mathcal{B}, \ldots$), etc., in general for fuzzy classes of the n-th order $(X^{(n)}, Y^{(n)}, \ldots)$.

Besides the crisp identity predicate $=$, the language of FCT contains:

- The membership predicate $\in$ between objects of successive sorts
- The class terms $\{x \mid \varphi\}$ of order $n+1$, for any variable x of any order n and any formula φ
- The symbols $\langle x_1, \ldots, x_k\rangle$ for k-tuples of individuals $x_1, \ldots, x_k$ of any order

In formulae of FCT we employ usual abbreviations and defined notions known from classical mathematics or traditional fuzzy mathematics, including those listed in Table 1, for all orders of fuzzy classes.

FCT has the following axioms, for all formulae φ and variables of all orders:

- The logical axioms of multi-sorted first-order logic $\boldsymbol{L}$
- The axioms of crisp identity: $x = x$; $x = y\ \&\ \varphi(x) \to \varphi(y)$; and $\langle x_1, \ldots, x_k\rangle = \langle y_1, \ldots, y_k\rangle \to x_i = y_i$

ISBN: 978-989-95079-6-8

Table 1: Abbreviations and defined notions of FCT

$\varphi = \psi$	$\equiv_{\mathrm{df}}$	$\Delta(\varphi \leftrightarrow \psi)$
$\varphi \leq \psi$	$\equiv_{\mathrm{df}}$	$\Delta(\varphi \rightarrow \psi)$
Ax	$\equiv_{\mathrm{df}}$	$x \in A$
$x_1 \ldots x_k$	$=_{\mathrm{df}}$	$\langle x_1, \ldots, x_k \rangle$
$(\forall x \in A)\varphi$	$\equiv_{\mathrm{df}}$	$(\forall x)(x \in A \rightarrow \varphi)$
$(\exists x \in A)\varphi$	$\equiv_{\mathrm{df}}$	$(\exists x)(x \in A \,\&\, \varphi)$
$\{x \in A \mid \varphi\}$	$=_{\mathrm{df}}$	$\{x \mid x \in A \,\&\, \varphi\}$
$\emptyset$	$=_{\mathrm{df}}$	$\{x \mid 0\}$
$\mathrm{Ker}\, A$	$=_{\mathrm{df}}$	$\{x \mid \Delta Ax\}$
$\smallsetminus A$	$=_{\mathrm{df}}$	$\{x \mid \neg Ax\}$
$A \cap B$	$=_{\mathrm{df}}$	$\{x \mid Ax \,\&\, Bx\}$
$A \sqcap B$	$=_{\mathrm{df}}$	$\{x \mid Ax \wedge Bx\}$
$A \sqcup B$	$=_{\mathrm{df}}$	$\{x \mid Ax \vee Bx\}$
$\mathrm{Pow}\, A$	$=_{\mathrm{df}}$	$\{X \mid X \subseteq A\}$
$\mathrm{Hgt}\, A$	$\equiv_{\mathrm{df}}$	$(\exists x)Ax$
$\mathrm{Plt}\, A$	$\equiv_{\mathrm{df}}$	$(\forall x)Ax$
$\mathrm{Crisp}\, A$	$\equiv_{\mathrm{df}}$	$(\forall x)\Delta(Ax \vee \neg Ax)$
$\mathrm{Sym}\, R$	$\equiv_{\mathrm{df}}$	$(\forall xy)(Rxy \rightarrow Ryx)$
$\mathrm{Trans}\, R$	$\equiv_{\mathrm{df}}$	$(\forall xyz)(Rxy \,\&\, Ryz \rightarrow Rxz)$
$\mathrm{Fnc}\, R$	$\equiv_{\mathrm{df}}$	$(\forall xyy')(Rxy \,\&\, Rxy' \rightarrow y = y')$
$A \subseteq B$	$\equiv_{\mathrm{df}}$	$(\forall x)(Ax \rightarrow Bx)$
$A \subseteq^{\Delta} B$	$\equiv_{\mathrm{df}}$	$(\forall x)(Ax \leq Bx)$

- The comprehension axioms: $y \in \{x \mid \varphi(x)\} \leftrightarrow \varphi(y)$
- The extensionality axioms: $(\forall x)(Ax = Bx) \rightarrow A = B$

Models of FCT are systems of fuzzy sets (and fuzzy relations) of all orders over a crisp universe of discourse, with truth degrees taking values in an $\boldsymbol{L}$-chain $\mathbf{L}$ (e.g., the interval $[0, 1]$ equipped with a left-continuous t-norm); thus all theorems on fuzzy classes provable in FCT are true statements about $\mathbf{L}$-valued fuzzy sets. Note, however, that the theorems of FCT are derived from its axioms by the rules of the fuzzy logic $\boldsymbol{L}$ rather than classical Boolean logic. For details on proving theorems in FCT see [5] or [26].

Acknowledgment

This work was supported by grant No. A100300503 of GA AV ČR and Institutional Research Plan AV0Z10300504.

References

[1] Petr Hájek. *Metamathematics of Fuzzy Logic*, volume 4 of *Trends in Logic*. Kluwer, Dordercht, 1998.

[2] Libor Běhounek and Petr Cintula. Fuzzy class theory. *Fuzzy Sets and Systems*, 154(1):34–55, 2005.

[3] Libor Běhounek, Ulrich Bodenhofer, and Petr Cintula. Relations in Fuzzy Class Theory: Initial steps. *Fuzzy Sets and Systems*, 159:1729–1772, 2008.

[4] Libor Běhounek and Martina Daňková. Relational compositions in Fuzzy Class Theory. To appear in *Fuzzy Sets and Systems*, doi:10.1016/j.fss.2008.06.013.

[5] Libor Běhounek and Petr Cintula. Fuzzy Class Theory: A primer v1.0. Technical Report V–939, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2006.

[6] Libor Běhounek and Petr Cintula. Features of mathematical theories in formal fuzzy logic. *Lecture Notes in Artificial Intelligence*, 4529:523–532, 2007. (Proc. IFSA Congr. 2007.)

[7] Libor Běhounek. Number-free reductions in logic-based fuzzy mathematics. Technical Report V–1032, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2008.

[8] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[9] Louise Schmir Hay. Axiomatization of the infinite-valued predicate calculus. *Journal of Symbolic Logic*, 28:77–86, 1963.

[10] Petr Hájek, Lluís Godo, and Francesc Esteva. Fuzzy logic and probability. *Proc. 11th UAI*, pages 237–244, Montreal, 1995.

[11] Lluís Godo, Francesc Esteva, and Petr Hájek. Reasoning about probability using fuzzy logic. *Neural Network World*, 10:811–823, 2000. (Spec. issue SOFSEM 2000.)

[12] Tommaso Flaminio. Strong non-standard completeness for fuzzy logics. *Soft Computing*, 12:321–333, 2008.

[13] Tommaso Flaminio and Lluís Godo. A logic for reasoning about the probability of fuzzy events. *Fuzzy Sets and Systems*, 158:625–638, 2006.

[14] Petr Cintula and Rostislav Horčík. Fuzzy Class Theory: Some advanced topics. *Proc. 5th EUSFLAT Conf.*, volume I, pages 137–143. University of Ostrava, 2007.

[15] Ulrich Höhle. Fuzzy real numbers as Dedekind cuts with respect to a multiple-valued logic. *Fuzzy Sets and Systems*, 24(3):263–278, 1987.

[16] Libor Běhounek. Towards a formal theory of fuzzy Dedekind reals. *Proc. Joint 4th Conf. of EUSFLAT and 11th LFA*, pages 946–954, Barcelona, 2005.

[17] Libor Běhounek and Tomáš Kroupa. Topology in Fuzzy Class Theory: Basic notions. *Lecture Notes in Artificial Intelligence*, 4529:513–522, 2007. (Proc. IFSA Congr. 2007.)

[18] Libor Běhounek and Tomáš Kroupa. Interior-based topology in Fuzzy Class Theory. *Proc. 5th EUSFLAT Conf.*, volume I, pages 145–151. University of Ostrava, 2007.

[19] Libor Běhounek, Ulrich Bodenhofer, Petr Cintula, and Susanne Saminger-Platz. Graded dominance. *Abstracts of the 29th Linz Seminar on Fuzzy Set Theory*, pages 11–14, Linz, 2008.

[20] Libor Běhounek, Ulrich Bodenhofer, Petr Cintula, Peter Sarkoci, and Susanne Saminger-Platz. Graded properties of binary fuzzy connectives. Technical Report V–1042, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2008.

[21] Lotfi A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[22] Karl Menger. Probabilistic theories of relations. *Proceedings of the National Academy of Sciences*, 37:178–180, 1951.

[23] Gültekin Soylu. Similarity-based fuzzy limits. *Fuzzy Sets and Systems*, 159:3380–3387, 2008.

[24] Libor Běhounek and Petr Cintula. From fuzzy logic to fuzzy mathematics: A methodological manifesto. *Fuzzy Sets and Systems*, 157(5):642–646, 2006.

[25] Francesc Esteva and Lluís Godo. Monoidal t-norm based logic: Towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124(3):271–288, 2001.

[26] Libor Běhounek and Petr Cintula. Fuzzy class theory as foundations for fuzzy mathematics. *Proc. 11th IFSA World Congress*, volume 2, pages 1233–1238, Beijing, 2005. Tsinghua University Press & Springer.

ISBN: 978-989-95079-6-8

Uncertainty Can Decrease Privacy: An Observation

Karen Villaverde[1] Olga Kosheleva[2]

1. Department of Computer Science
New Mexico State University
Las Cruces, NM 88003, USA
2. Department of Teacher Education
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
Email: kvillave@cs.nmsu.edu, olgak@utep.edu

Abstract— *In many application areas, e.g., in medical applications, it is important to be able to get statistical information about the data without disclosing individual cases. One of the methods to preserve privacy in such statistical databases is to introduce uncertainty, i.e., to replace the exact values with intervals of possible values. In this paper, we show that while the introduction of uncertainty can often enhance privacy, sometimes, the opposite effect occurs: adding uncertainty can decrease privacy.*

Keywords— Fuzzy uncertainty, interval uncertainty, privacy, statistical databases

1 Statistical Databases – Need to Preserve Privacy

1.1 Need for Collecting Data

In many practical situations, it is very useful to collect large amounts of data.

For example, from the data that we collect during a census, we can extract a lot of information about health, mortality, employment in different regions – for different age ranges, and for people from different genders and of different ethnic groups. By analyzing these statistics, we can reveal troubling spots and allocate (usually limited) resources so that the help goes first to social groups that need it most.

Similarly, by gathering data about people's health in a large medical database, we can extract a lot of useful information on how the geographic location, age, and gender affect a person's health. Then, we can improve the public health by appropriate appropriate public health measures to different portions of the population.

Finally, a large database of purchases can help find out what people are looking for, make shopping easier for customers and at the same time, decrease the stores' expenses related to storing unnecessary items.

1.2 Need for Privacy

Privacy is an important issue in the statistical analysis of human-related data. For example, to check whether in a certain geographic area, there is a gender-based discrimination, one can use the census data to check, e.g., whether for all people from this area who have the same level of education, there is a correlation between salary and gender. One can think of numerous possible questions of this type related to different sociological, political, medical, economic, and other questions. From this viewpoint, it is desirable to give researchers *ability to perform* whatever *statistical analysis* of this data that is reasonable for their specific research.

On the other hand, we may not want to give the researchers direct access to the raw census data, because a large part of the census data is *confidential*. For example, for most people (those who work in the private sector) salary information is confidential. Suppose that a corporation is deciding where to built a new plant and has not yet decided between two possible areas. This corporation would benefit from knowing the average salary of people of needed education level in these two areas, because this information would help them estimate how much it will cost to bring local people on board. However, since salary information is confidential, the company should not be able to know the exact salaries of different potential workers.

The need for privacy is also extremely important for *medical* experiments, where we should be able to make statistical conclusions about, e.g., the efficiency of a new medicine without disclosing any potentially embarrassing details from the individual medical records.

Such databases in which the outside users cannot access individual records but can solicit statistical information are often called *statistical databases*.

1.3 Maintaining Privacy is Not Easy

Maintaining privacy in statistical databases is not easy. Clerks who set up policies on access to statistical databases sometimes erroneously assume that once the records are made anonymous, we have achieved perfect privacy. Alas, the situation is not so easy: even when we keep all the records anonymous, we can still extract confidential information by asking appropriate questions.

Many examples of such extraction can be found in a book by D. Denning [1]. For example, suppose that we are interested in the salary of Dr. X who works for a local company. Dr. X's mailing address can be usually taken from the phone book; from the company's webpage, we can often get his photo and thus find out his race and approximate age. Then, to determine Dr. X's salary, all we need is to ask what is the average salary of all people with a Ph.D. of certain age brackets who live in a small geographical area around his actual home address – if the area is small enough, then Dr. X will be the

ISBN: 978-989-95079-6-8

only person falling under all these categories.

Even if we only allow statistical information about salaries $s_1, \ldots, s_q$ when there are at least a certain amount n_0 people within a requested range, we will still be able to reconstruct the exact salaries of all these people. Indeed, for example, we can ask for the number and average salary of all the people who live on Robinson street at houses 1 through 1001, and then we can ask the same question about all the people who live in houses from 1 to 1002. By comparing the two numbers, we get the average salary of the family living at 1002 Robinson – in other words, we gain the private information that we tried to protect.

In general, we can ask for the average

$$\frac{s_1 + \ldots + s_q}{q},$$

and for several moments of salary (variance, third moment, etc):

- if we know the values v_j of at least q different functions $f_j(s_1, \ldots, s_q)$ of s_i,
- then we can, in general, reconstruct all these values from the corresponding system of q equations with q unknowns:

$$f_1(s_1 \ldots, s_q) = v_1,$$

$$\ldots,$$

$$f_q(s_1, \ldots, s_q) = v_q.$$

At first glance, moments are natural and legitimate statistical characteristics, so researchers would be able to request these characteristics. On the other hand, we do not want the researchers to be able to extract the exact up-to-cent salaries of all the people leaving in a certain geographical area.

What restriction should we impose on possible statistical queries that would guarantee privacy but restrict research in the least possible way?

These are anecdotal examples of poorly designed privacy and security, but, as we have mentioned, the problem is indeed difficult: Many seemingly well-designed privacy schemes later turn out to have unexpected privacy and security problems, and it is known that the problem of finding a privacy-preserving scheme is, in general, NP-hard [1].

Different aspects of the problem of privacy in statistical databases, different proposed solutions to this problem, and their drawbacks, are described in [1, 11, 12] (see also references therein).

2 Known Fact: Uncertainty Can Enhance Privacy

A reasonable way to avoid privacy violations is to store ranges (intervals) of values instead of the actual values. For example, instead of keeping the exact age, we only record whether the age is between 0 and 10, 10 and 20, 20 and 30, etc.

In this case, no matter what statistics we allow, the worst that can happen is that the corresponding ranges will be disclosed. However, in this situation, we do not disclose the original exact values – since these values are not stored in the database in the first place; see, e.g., [5, 6, 7].

3 New Observation: Uncertainty Can Decrease Privacy

3.1 What We Do in This Section

The successful use of uncertainty to enhance privacy in statistical databases may lead to an impression that uncertainty *always* enhances privacy. In this paper, we show that uncertainty can actually *decrease* privacy in a statistical database.

3.2 Simplest Possible Case: 1-D Databases

We will show that the privacy decrease phenomenon occurs in the simplest 1-dimensional case, when each record in the statistical database consists of a single value x_i. In this case, the database consists of n values $x_1, \ldots, x_n$.

3.3 1-D Case: Choice of Statistical Characteristics

Among the most important statistical characteristics are the mean values of certain quantities.

In the 1-D case, each record contains the value x_i of a single quantity x. So, in the 1-dimensional case, we are interested in estimating the mean values of different characteristics $u(x)$, i.e., in estimating the values

$$E[u] \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} u(x_i). \tag{1}$$

3.4 Smooth Characteristics

Among different statistical characteristics, an important class is formed by characteristics for which the corresponding function $u(x)$ is smooth.

Usually, most values x_i corresponding to different individuals do not differ much. In other words, most values x_i belong to a small-size region. In this region, we can expand the function $u(x)$ in Taylor series and ignore higher-order terms. In the first approximation, we thus approximate the function $u(x)$ by a linear expression

$$u(x) \approx u_0 + u_1 \cdot x. \tag{2}$$

In the next approximation, we approximate an arbitrary smooth function $u(x)$ by a quadratic expression:

$$u(x) \approx u_0 + u_1 \cdot x + u_2 \cdot x^2. \tag{3}$$

In this paper, we will show that the phenomenon of decreasing privacy occurs already for such quadratic characteristics.

3.5 For Exactly Known Values, the Computation of Quadratic Characteristics Preserves Privacy

In applications, we may be interested in the mean values of different statistical characteristics. If we restrict ourselves to quadratic characteristics, this means that we must be able, given an arbitrary quadratic characteristic (3), to estimate the average value (1) of this characteristic.

Substituting the general expression (3) into the formula (1), we can conclude that the desired average value has the form

$$E[u] = u_0 + u_1 \cdot E[x] + u_2 \cdot E[x^2], \tag{4}$$

where

$$E[x] \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{5}$$

ISBN: 978-989-95079-6-8

and

$$E[x^2] \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2. \tag{6}$$

Thus, in effect, even when we know the mean value $E[u]$ for all possible quadratic characteristics, all we know about the values x_i are the two combinations: the mean $E[x]$ and the second moment $E[x^2]$.

Based on the knowledge of these two values, we cannot reconstruct n different values $x_1, \ldots, x_n$ and thus, privacy is preserved.

3.6 Case of Interval Uncertainty

Let us now consider what happens if instead of the exact value of x_i, we only know the *interval* $[\underline{x}_i, \overline{x}_i]$ of possible values of x_i.

Often, these intervals come from the fact that we know an approximate value $\widetilde{x}_i$, and we know the upper bound Δ_i on the approximation accuracy $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$: $|\Delta x_i| \le \Delta_i$. In this case, the only information that we have about the actual (unknown) value x_i is that x_i belongs to the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. So, we get an interval $[\underline{x}_i, \overline{x}_i]$ with $\underline{x}_i = \widetilde{x}_i - \Delta_i$ and $\overline{x}_i = \widetilde{x}_i + \Delta_i$.

In general, every interval $[\underline{x}_i, \overline{x}_i]$ can be represented in the form $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$: it is sufficient to take:

- as $\widetilde{x}_i$, the midpoint of the interval, i.e., the value

$$\widetilde{x}_i = \frac{\underline{x}_i + \overline{x}_i}{2}; \tag{7}$$

  and

- as Δ_i, the radius (half-width) of the interval, i.e., the value

$$\Delta_i = \frac{\overline{x}_i - \underline{x}_i}{2}. \tag{8}$$

3.7 Estimating Statistical Characteristics under Interval Uncertainty: Formulation of the Problem

Suppose that we are interested in the mean value $E[u]$ of a statistical characteristic $u(x)$. In the ideal case when we know the exact values $x_1, \ldots, x_n$, this mean value is simply equal to the arithmetic average of the corresponding values $u(x_i)$.

Under interval uncertainty, we can have different values $x_i \in [\underline{x}_i, \overline{x}_i]$. In general, different values x_i can lead to different values of the mean. We are interested in computing the *range* $[\underline{u}, \overline{u}]$ of possible values of this mean, i.e., the interval

$$[\underline{u}, \overline{u}] \stackrel{\text{def}}{=} \left\{ \frac{1}{n} \cdot \sum_{i=1}^{n} u(x_i) : x_i \in [\underline{x}_i, \overline{x}_i] \right\}. \tag{9}$$

3.8 Estimating Quadratic Characteristics under Interval Uncertainty: Algorithm

How can we compute the range of the mean

$$E[u] = \frac{1}{n} \cdot \sum_{i=1}^{n} u(x_i)?$$

Each of the variables x_i can take any value from the corresponding interval $[\underline{x}_i, \overline{x}_i]$. Thus,

- the mean $E[u]$ takes the *largest* value if and only if each term $u(x_i)$ takes the *largest* possible value on the interval $[\underline{x}_i, \overline{x}_i]$; we will denote this largest value by $\overline{u}_i$;
- similarly, the mean $E[u]$ takes the *smallest* value if and only if each term $u(x_i)$ takes the *smallest* possible value on the interval $[\underline{x}_i, \overline{x}_i]$; we will denote this smallest value by $\underline{u}_i$.

For a given quadratic function $u(x_i)$ on a given interval $[\underline{x}_i, \overline{x}_i]$, computing its smallest value $\underline{u}_i$ and its largest value $\overline{u}_i$ is an easy computational task. Indeed, according to calculus, if a function attains its minimum or maximum at some point inside an interval, then its derivative is 0 at this point. Thus, to find the minimum and the maximum of a given function $u(x_i)$ on a given interval $[\underline{x}_i, \overline{x}_i]$, it is sufficient to compute the values of this function

- at the endpoints $\underline{x}_i$ and $\overline{x}_i$ of this interval and
- at the point(s) (if any) where the derivative $u'(x)$ of the function $u(x)$ is equal to 0.

Then,

- the smallest of these values is the desired minimum $\underline{u}_i$, and
- the largest of these values is the desired maximum $\overline{u}_i$.

For a quadratic function $u(x) = u_0 + u_1 \cdot x + u_2 \cdot x^2$, the derivative is a linear function $u'(x) = u_1 + 2 \cdot u_2 \cdot x$, so computing the point where the derivative is 0 is a straightforward task: $u_1 + 2 \cdot u_2 \cdot x = 0$ implies $x = -u_1/(2u_2)$.

Once we compute the values $\underline{u}_i$ and $\overline{u}_i$, we can now find the desired range $[\underline{u}, \overline{u}]$ as

$$\underline{u} = \frac{1}{n} \cdot \sum_{i=1}^{n} \underline{u}_i, \ \ \overline{u} = \frac{1}{n} \cdot \sum_{i=1}^{n} \overline{u}_i. \tag{10}$$

3.9 Under Interval Uncertainty, Privacy Is No Longer Preserved: Result

Let us show that in this case, privacy is no longer preserved. To be more precise, we assume that we are given a 1-D database, i.e., a collection of intervals $[\underline{x}_1, \overline{x}_1], \ldots, [\underline{x}_n, \overline{x}_n]$. We do not have an explicit access to intervals from this collection, but for every quadratic function $u(x)$, we can generate the range of the mean value $E[u]$ over these intervals.

We will show that in almost all cases, based on these ranges, we can actually reconstruct all the original intervals $[\underline{x}_i, \overline{x}_i]$. In other words, adding uncertainty leads to a loss of privacy.

Comment. By "almost all" cases, we mean all cases in which all n midpoints $\widetilde{x}_i$ are different. Situations when two midpoints coincide are indeed degenerate, since a minor modification of the original data leads to $\widetilde{x}_i \neq \widetilde{x}_j$.

3.10 Under Interval Uncertainty, Privacy Is No Longer Preserved: Proof

In our proof, for every real number a, we consider the quadratic function $u(x) = (x - a)^2$. For this function, the derivative is equal to 0 at the minimum point $x = a$. Thus, this function attains its largest value on the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ at one of the endpoints $\widetilde{x}_i - \Delta_i$ or $\widetilde{x}_i + \Delta_i$. One can easily check that:

ISBN: 978-989-95079-6-8

- when $a \le \widetilde{x}_i$, then the largest possible value $\overline{u}_i$ of $u(x)$ on the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ is attained when $x_i = \overline{x}_i = \widetilde{x}_i + \Delta_i$ and is equal to $\overline{u}_i = (\overline{x}_i - a)^2$;
- when $a \ge \widetilde{x}_i$, then the largest possible value $\overline{u}_i$ of $u(x)$ on the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ is attained when $x_i = \underline{x}_i = \widetilde{x}_i - \Delta_i$ and is equal to $\overline{u}_i = (\underline{x}_i - a)^2$.

Let us use this fact to describe the dependence of $\overline{u}$ on the parameter a.

When $a \ne \widetilde{x}_i$, the value $\overline{u}$ is the average of n smooth expressions.

At each point $a = \widetilde{x}_i$, all the terms $\overline{u}_j$ in the sum $\overline{u}$ are smooth except for the term $\overline{u}_i$ that turns from $(\overline{x}_i - a)^2$ to $(\underline{x}_i - a)^2$. The derivative of $\overline{u}_i$ with respect to a changes from $2 \cdot (a - \overline{x}_i)$ to $2 \cdot (a - \underline{x}_i)$, i.e., increases by

$$2 \cdot (a - \underline{x}_i) - 2 \cdot (a - \overline{x}_i) = 2 \cdot (\overline{x}_i - \underline{x}_i) = 4 \cdot \Delta_i. \quad (11)$$

Since all the other components $\overline{u}_j$ are smooth at $a = \widetilde{x}_i$, at $a = \widetilde{x}_i$, the derivative of the average $\overline{u}(a)$ increases by $\frac{4}{n} \cdot \Delta_i$.

Thus, once we know the value $\overline{u}$ for all a,

- we can find the values $\widetilde{x}_i$ as the values at which the derivative is discontinuous; and
- we can find each value Δ_i as $n/4$ times the increase of the derivative at the corresponding point $\widetilde{x}_i$.

The statement is proven.

3.11 Case of Fuzzy Uncertainty

In many practical situations, the estimates $\widetilde{x}_i$ come from experts. Experts often describe the inaccuracy of their estimates in terms of imprecise words from natural language, such as "approximately 0.1", etc. A natural way to formalize such words is to use special techniques developed for formalizing this type of estimates – specifically, the technique of fuzzy logic; see, e.g., [4, 10].

In this technique, for each possible value of $x_i \in [\underline{x}_i, \overline{x}_i]$, we describe the degree $\mu_i(x_i)$ to which this value is possible. For each degree of certainty α, we can determine the set of values of x_i that are possible with at least this degree of certainty – the *α-cut* $\mathbf{x}_i(\alpha) = \{x \,|\, \mu_i(x) \ge \alpha\}$ of the original fuzzy set. Vice versa, if we know α-cuts for every α, then, for each object x, we can determine the degree of possibility that x belongs to the original fuzzy set [2, 4, 8, 9, 10]. A fuzzy set can be thus viewed as a nested family of its (interval) α-cuts.

3.12 From the Computational Viewpoint, Fuzzy Uncertainty Can Be Reduced to the Interval One

Once we know how to propagate interval uncertainty, i.e., how to compute the range

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_n) : x_i \in \mathbf{x}_i\} \quad (12)$$

of a given function $f(x_1, \ldots, x_n)$ over given intervals $\mathbf{x}_i$, then, to propagate the fuzzy uncertainty, we can consider, for each α, the fuzzy set y with the α-cuts

$$\mathbf{y}(\alpha) = f(\mathbf{x}_1(\alpha), \ldots, \mathbf{x}_n(\alpha)); \quad (13)$$

see, e.g., [2, 4, 8, 9, 10]. This is equivalent to using Zadeh's extension principle.

So, from the computational viewpoint, the problem of propagating fuzzy uncertainty can be reduced to several interval propagation problems.

For example, the fuzzy value of $E[u]$ can be described as follows: for each α, the corresponding α-cut is equal to the range of $E[u]$ when for each i, all values x_i belong to the corresponding α-cuts: $x_i \in \mathbf{x}_i(\alpha)$.

3.13 Under Fuzzy Uncertainty, Privacy Is Also Not Preserved: Result

Let us assume that we are given a 1-D database, i.e., a collection of fuzzy numbers $\mu_1(x_1), \ldots, \mu_n(x_n)$. We do not have an explicit access to fuzzy numbers from this collection, but for every quadratic function $u(x)$, we can generate the fuzzy number $\mu(u)$ corresponding to the mean $E[u]$.

We will show that in almost all cases (i.e., when the midpoints are all different), based on the resulting fuzzy numbers $\mu(u)$, we can actually reconstruct all the original fuzzy numbers $\mu_i(x_i)$.

In other words, adding fuzzy uncertainty also leads to a loss of privacy.

3.14 Under Fuzzy Uncertainty, Privacy Is Also Not Preserved: Proof

For each α, the α-cuts of the resulting fuzzy number $\mu(u)$ is the range of $E[u]$ when $x_i \in \mathbf{x}_i(\alpha)$.

We already know that from the ranges computed for all possible quadratic functions $u(x)$, we can reconstruct all the intervals. Thus, for every α, we can reconstruct all the α-cuts of all the membership functions $\mu_i(x_i)$.

Since we can thus reconstruct all α-cuts for all α, hence, we can uniquely reconstruct all the membership functions.

The statement is proven.

Acknowledgment

The authors are thankful to the anonymous referees for useful suggestions.

References

[1] D. Denning, *Cryptography and Data Security*, Addison-Wesley, Reading, MA, 1982.

[2] D. Dubois and H. Prade, Operations on fuzzy numbers, *International Journal of Systems Science*, 1978, Vol. 9, pp. 613–626.

[3] P. C. Fishburn, *Nonlinear preference and utility theory*, The John Hopkins Press, Baltimore, MD, 1988.

[4] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, New Jersey, 1995.

[5] V. Kreinovich and L. Longpré, Computational complexity and feasibility of data processing and interval computations, with extension to cases when we have partial information about probabilities, In: V. Brattka, M. Schroeder, K. Weihrauch, and N. Zhong, editors, *Proc. Conf. on Computability and Complexity in Analysis CCA'2003*, Cincinnati, Ohio, USA, Aug. 28–30, 2003, pp. 19–54.

[6] V. Kreinovich, L. Longpré, S. A. Starks, G. Xiang, J. Beck, R. Kandathi, A. Nayak, S. Ferson, and J. Hajagos, "Interval Versions of Statistical Techniques, with Applications to Environmental Analysis, Bioinformatics, and Privacy in Statistical Databases", *Journal of Computational and Applied Mathematics*, 2007, Vol. 199, No. 2, pp. 418–423.

[7] L. Longpré, G. Xiang, V. Kreinovich, and E. Freudenthal, "Interval Approach to Preserving Privacy in Statistical Databases: Related Challenges and Algorithms of Computational Statistics", In: V. Gorodetsky, I. Kotenko, and V. A. Skormin (eds.), *Proceedings of the International Conference "Mathematical Methods, Models and Architectures for Computer Networks Security"*, MMM-ACNS-07, St. Petersburg, Russia, September 13–15, 2007, Springer Lecture Notes in Computer Science, 2007, Vol. CCIS-1, pp. 346–361.

[8] R. E. Moore and W. Lodwick, Interval Analysis and Fuzzy Set Theory, *Fuzzy Sets and Systems*, 2003, Vol. 135, No. 1, pp. 5–9.

[9] H. T. Nguyen and V. Kreinovich, Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications, In: R. B. Kearfott and V. Kreinovich, eds., *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.

[10] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida, 2005.

[11] L. Sweeney, "Weaving technology and policy together to maintain confidentiality", *Journal of Law, Medicine and Ethics*, 1997, Vol. 25, pp. 98–110.

[12] L. Willenborg and T. De Waal, *Statistical disclosure control in practice*, Springer Verlag, New York, 1996.

ISBN: 978-989-95079-6-8

Decentralized Adaptive Fuzzy-Neural Control of an Anaerobic Digestion Bioprocess Plant

Ieroham S. Baruch[1] Rosalba Galvan-Guerra[1]

1 Department of Automatic Control, CINVESTAV-IPN, Mexico D.F., 07360 Mexico
Email: {baruch, rgalvan}@ctrl.cinvestav.mx

***Abstract*—** *The paper proposed to use recurrent Fuzzy-Neural Multi-Model (FNMM) identifier for decentralized identification of a distributed parameter anaerobic wastewater treatment digestion bioprocess, carried out in a fixed bed and a recirculation tank. The distributed parameter analytical model of the digestion bioprocess is used as a plant data generator. It is reduced to a lumped system using the orthogonal collocation method, applied in three collocation points (plus the recirculation tank), which are used as centers of the membership functions of the fuzzyfied plant output variables with respect to the space variable. The local and global weight parameters and states of the proposed FNMM identifier are used to design hierarchical FNMM direct and indirect controllers. The comparative graphical simulation results of the digestion system direct and indirect control, obtained via learning, exhibited a good convergence, and precise reference tracking. The comparative numerical results, giving the final means squared error of control of each output variable showed that the indirect adaptive decentralized fuzzy-neural control outperformed the direct one, and the it outperformed the linearized proportional optimal control too.*

***Keywords*—** Decentralized control, direct adaptive control, indirect adaptive control, distributed parameter digestion bioprocess system, recurrent neural network model, hierarchical fuzzy neural identification and control.

1 Introduction

The Distributed Parameter Systems (DPS) are distinguished by the fact that the states, controls, and outputs may depend on spatial position, [1]. Thus the natural form of the system description is by Partial Differential Equations (PDE), integral equations, or transcendental transfer functions, [1]. On the other side, in the last two decades, a new identification and control tools like Neural Networks (NN), used for biotechnological plants, [2], rose fame. Among several possible network architectures the ones most widely used are the Feedforward NN (FFNN) and the Recurrent NN (RNN), [3]. The main NN property namely the ability to approximate complex non-linear relationships without prior knowledge of the model structure makes them a very attractive alternative to the classical modeling and control techniques. This property has been proved for both types of NNs by the universal approximation theorem [3]. The preference given to NN identification with respect to the classical methods of process identification is clearly demonstrated in the solution of the "bias-variance dilemma" [3]. The FFNN and the RNN have been applied for DPS identification and control too. In [4], a RNN is used for system identification and process prediction of a DPS dynamics - an adsorption column for wastewater treatment of water contaminated with toxic chemicals. In [5, 6], a spectral-approximation-based intelligent modeling approach is proposed for the distributed thermal processing of the snap curing oven DPS that is used in semiconductor packaging industry. In [7], it is presented a new methodology for the identification of DPS, based on NN architectures, motivated by standard numerical discretization techniques used for the solution of PDE. In [8], an attempt is made to use the philosophy of the NN adaptive-critic design to the optimal control of distributed parameter systems. In [9] the concept of proper orthogonal decomposition is used for the model reduction of DPS to form a reduced order lumped parameter problem. In [10], measurement data of an industrial process are generated by solving the PDE numerically using the finite differences method. Both centralized and decentralized NN models are introduced and constructed based on this data. The models are implemented on FFNN using Backpropagation (BP) and Levenberg-Marquardt learning algorithms. In [11, 12, 13, 14], Baruch et al. defined direct and indirect Fuzzy Neural Multi-Model (FNMM) control system, based on Takagi-Sugeno (T-S) fuzzy rules, [15], containing in its consequent parts computational procedure of a Backpropagation Learning (BP) of a Recurrent Trainable NN (RTNN), [14]. In the present paper, the direct and indirect FNMM control system are modified and used for decentralized identification and control of a digestion anaerobic DPS of wastewater treatment, [16]. The anaerobic bioprocess plant model, used as an input/output plant data generator, is described by PDE/ODE, and simplified using the orthogonal collocation technique in three collocation points and a recirculation tank, [16], [17].

2 Analytical Model of the Anaerobic Digestion Bioprocess

The anaerobic digestion systems conformed by a fixed bed reactor and a recirculation tank is depicted on Fig. 1. It contained a fixed bed bioreactor and a recirculation tank. The anaerobic digestion process is modeled using PDE, [16]:

$$\frac{\partial X_1}{\partial t} = (\mu_1 - \varepsilon D) X_1 , \quad \mu_1 = \mu_{1\max} \frac{S_1}{K_{s_1}^{'} X_1 + S_1}, \tag{1}$$

$$\frac{\partial X_2}{\partial t} = (\mu_2 - \varepsilon D) X_2 , \quad \mu_2 = \mu_{2s} \frac{S_2}{K_{s_2}^{'} X_2 + S_2 + \frac{S_2^2}{K_{I_2}}}, \tag{2}$$

Where: X_1 is concentration of acidogenic bacteria; X_2-

ISBN: 978-989-95079-6-8

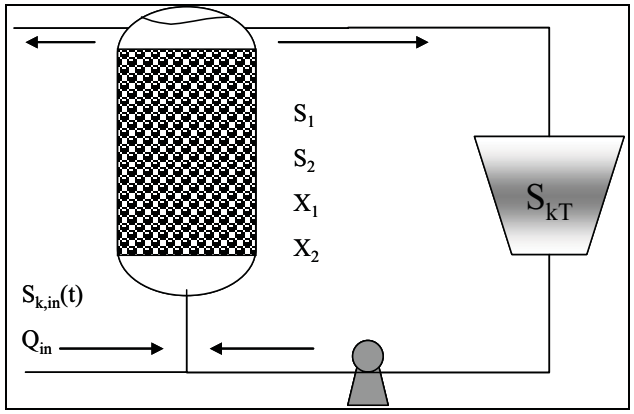


Figure 1: Block-Diagram of Anaerobic Digestion Bioreactor

concentration of methanogenic bacteria; S_1 - chemical oxygen demand; S_2 - volatile fatty acids; μ_1, μ_2- Monod functions, representing the acidogenesis and methanogenesis growth rate; D - dilution rate; ε - bacteria fraction in the liquid phase. The next equations of the bioprocess model are:

$$\frac{\partial S_1}{\partial t} = \frac{E_z}{H^2}\frac{\partial^2 S_1}{\partial z^2} - D\frac{\partial S_1}{\partial t} - k_1\mu_1 X_1\,, \tag{3}$$

$$\frac{\partial S_2}{\partial t} = \frac{E_z}{H^2}\frac{\partial^2 S_2}{\partial z^2} - D\frac{\partial S_2}{\partial t} + k_2\mu_1 X_1\,, \tag{4}$$

$$S_1(0,t) = \frac{S_{1,in}(t) + RS_{1T}}{R+1},\quad S_2(0,t) = \frac{S_{2,in}(t) + RS_{2T}}{R+1},\quad R = \frac{Q_T}{DV_{eff}} \tag{5}$$

$$\frac{\partial S_1}{\partial z}(1,t) = 0\,,\quad \frac{\partial S_2}{\partial z}(1,t) = 0\,. \tag{6}$$

$$\frac{dS_{1T}}{dt} = \frac{Q_T}{V_T}\left(S_1(1,t) - S_{1T}\right),\quad \frac{dS_{2T}}{dt} = \frac{Q_T}{V_T}\left(S_2(1,t) - S_{2T}\right). \tag{7}$$

Where: S_{1T} , S_{2T} are concentrations of the chemical oxygen demand and the volatile fatty acids in the recirculation tank, respectively; H - fixed bed length; Q_T - recycle flow rate; V_T - volume of the recirculation tank. The physical meaning of the other constants and initial values of the variables of the process model are given in [16]. For practical purpose, the full PDE process model, [16], could be reduced to an ODE system using an early lumping technique and the Orthogonal Collocation Method (OCM), [16], [17]. The precision of the OCM approximation of the PDE model depended on the number of measurement (collocation) points, but the approximation is always exact in that points. If the number of points is very high and the point positions are chosen inappropriately, the ODE model could loose identifiability. Furthermore the ODE plant model here is used as a plant data generator for neural identification and control of PDE system and the number of point not need to be too high. So to fulfill this objective we need a reduced order model having only three points, (0.25H, 0.5 H, 0.75H), but generating 14 measured variables. The reference set points generated for all that variables keep the form but differ in amplification due to its position. The plant ODE system model, obtained by OCM is taken from [16] and used as a plant process input/output data generator so to obtain identification and control simulation results.

3 Description of the Direct and Indirect Fuzzy-Neural Multi-Model Control System

3.1 Direct Adaptive FNMM Control System Design

The block-diagrams of the complete control system and its identification and control parts are schematically depicted in Fig. 2, Fig. 3 and Fig. 4.

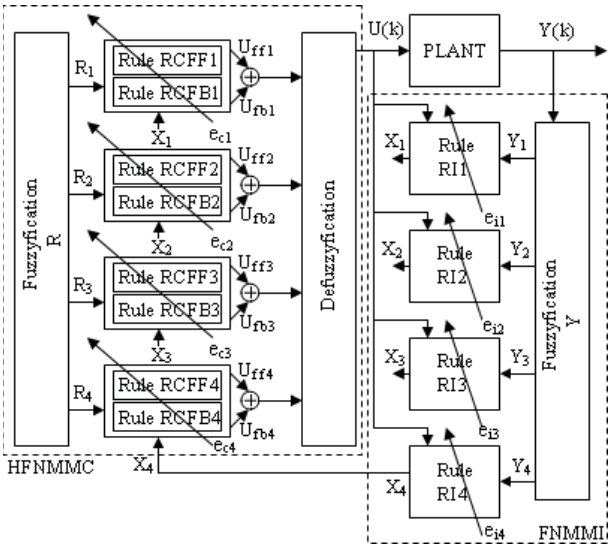


Figure .2: Block-Diagram of the FNMM Control System.

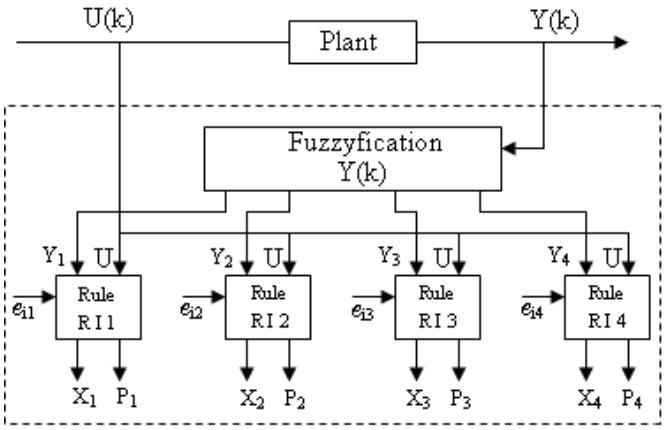


Figure 3: Detailed block-diagram of the FNMM identifier

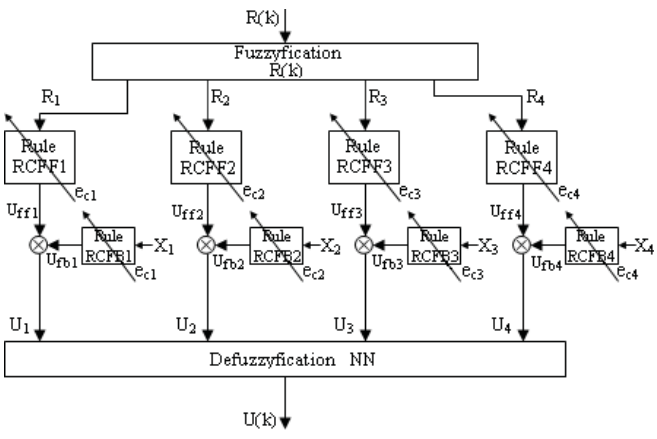


Figure 4: Detailed block-diagram of the HFNMM controller

ISBN: 978-989-95079-6-8

The structure of the entire control system, [12, 14], contained Fuzzyfier, Fuzzy Rule-Based Inference System (FRBIS), containing four identification, four feedback control and four feedforward control T-S rules (RI_i, $RCfb_i$, $RCff_i$) and a Defuzzyfier. The plant output variable and its correspondent reference variable depended on space and time, and they are fuzzyfied on space. The membership functions of the fixed-bed output variables are triangular/trapezoidal ones and that belonging to the output variables of the recirculation tank are singletons. The centers of the membership functions are the respective collocation points of the plant. The main objective of the Fuzzy-Neural Multi-Model Identifier (FNMMI), containing four rules, is to issue states for the direct adaptive Fuzzy-Neural Multi-Model Feedback Controller (FNMMFBC) when the FNMMI outputs follows the outputs of the plant in the four measurement (collocation) points with minimum Means Squared Error (MSE) of approximation. The direct fuzzy neural controller has also a direct adaptive Fuzzy-Neural Multi-Model Feedforward Controller (FNMMFFC). The objective of the direct adaptive FNMM controller, containing four Feedback (FB) and four Feedforward (FF) T-S control rules is to reduce the error of control. The upper hierarchical level of the FNMM control system is one layer perceptron which represented the defuzzyfier, [12, 14]. The hierarchical FNMM controller has two levels – Lower Level of Control (LLC), and Upper Level of Control (ULC). It is composed of three parts: 1) Fuzzyfication, where the normalized reference vector signal contained reference components of four measurement points; 2) Lower Level Inference Engine, which contains twelve T-S fuzzy rules (four rules for identification and eight rules for control- four in the feedback part and four in the feedforward part), operating in the corresponding measurement points; 3) Upper Hierarchical Level of neural defuzzification.

The detailed block-diagram of the FNMMI, given on Fig. 3, contained a space plant output fuzzyfier and four identification T-S fuzzy rules, labeled as RI_i, which consequent parts are RTNN learning procedures, [14]. The identification T-S fuzzy rules have the form:

$$RI_i\text{: } \textbf{\textit{If}}\ \mathrm{x(k)}\ \textbf{\textit{is}}\ \mathrm{A_i}\ \textbf{\textit{and}}\ \mathrm{u(k)}\ \textbf{\textit{is}}\ \mathrm{B_i}\ \textbf{\textit{then}}\ \mathrm{Y_i = \Pi_i\ (L,M,N_i,Y_{di},U,X_i,A_i,B_i,C_i,E_i),\ i=1,2,..,4} \tag{8}$$

The detailed block-diagram of the FNMMC, given on Fig. 4, contained a spaced plant reference fuzzyfier and eight control T-S fuzzy rules (four FB and four FF), which consequent parts are also RTNN learning procedures, [12, 14], using the state information, issued by the corresponding identification rules. The consequent part of each feedforward control rule (the consequent learning procedure) has the M, L, N_i RTNN model dimensions, R_i, Y_{di}, E_{ci} inputs and U_{ffi}, outputs used to form the total control. The T-S fuzzy rule has the form:

$$RCFF_i\text{: } \textbf{\textit{If}}\ \mathrm{R(k)}\ \textbf{\textit{is}}\ B_i\ \textbf{\textit{then}}\ \mathrm{Uff_i = \Pi_i\ (M, L, N_i, R_i, Y_{di}, X_i, J_i, B_i, C_i, E_{ci}),\ i =1,2,..,4} \tag{9}$$

The consequent part of each feedback control rule (the consequent learning procedure) has the M, L, N_i RTNN model dimensions, Y_{di}, X_i, E_{ci} inputs and U_{fbi}, outputs used to form the total control. The T-S fuzzy rule has the form:

$$RCFB_i\text{: } \textbf{\textit{If}}\ \mathrm{Y_{di}}\ \textbf{\textit{is}}\ A_i\ \textbf{\textit{then}}\ \mathrm{Ufb_i = \Pi_i\ (M, L, N_i, Y_{di}, X_i, X_{ci}, J_i, B_i, C_i, E_{ci}),\ i =1,2,..,4} \tag{10}$$

The total control corresponding to each of the four measurement points is a sum of its corresponding feedforward and feedback parts:

$$U_i\,(k) = -U_{ffi}\,(k) + U_{fbi}\,(k) \tag{11}$$

The defuzzyfication learning procedure, which correspond to the single layer perceptron learning, performed a weighted sum of the control variables Ui, [15]. It is described by:

$$\mathrm{U = \Pi\ (M, L, N, Y_d, Uo, X, A, B, C, E)} \tag{12}$$

3.1 Indirect Adaptive FNMM Control System Design

The block-diagram of this control system is schematically depicted in Fig.5. The structure of the entire control system, [14, 15], contained Fuzzyfier, Fuzzy Rule-Based Inference System (FRBIS), containing four identification and four control T-S rules (RI_i, RC_i), and a Defuzzyfier. The plant output variable and its correspondent reference variable depended on space and time, and they are fuzzyfied on space. The membership functions and the Identifier (FNMMI) are the same used in the direct controller. The objective of the indirect adaptive FNMM controller is equivalent to that of the direct controller. The hierarchical FNMM controller has two levels – Lower Level of Control (LLC), and Upper Level of Control (ULC). It is composed of three parts: 1) Fuzzyfication, where the normalized reference vector signal contained reference components of four measurement points; 2) Lower Level Inference Engine, which contains eight T-S fuzzy rules (four rules for identification and four rules for control), operating in the corresponding measurement points; 3) Upper Hierarchical Level of neural defuzzification. The detailed block-diagram of the FNMMC is given on Fig. 6. It contained a spaced plant reference fuzzyfier and four sliding mode control T-S fuzzy rules, which consequent parts are SMC procedures, [13, 14], using the state, and parameter information, issued by the corresponding identification rules. The control T-S fuzzy rules have the form:

$$RC_i\text{: } \textbf{\textit{If}}\ \mathrm{R(k)}\ \textbf{\textit{is}}\ \mathrm{C_i}\ \textbf{\textit{then}}\ \mathrm{U_i = \Pi_i\ (M, L, N_i, R_i, Y_{di}, X_i, A_i, B_i, C_i, E_{ci}),\ i =1, 2,\ldots, 4} \tag{13}$$

The defuzzyfication of the control variable is a learning procedure, which correspond to the single layer perceptron learning. It is described by:

$$\mathrm{U = \Pi\ (M, L, N, Y_d, Uo, X, A, B, C, E)} \tag{14}$$

The T-S rule and the defuzzification of the plant output of the fixed bed with respect to the space variable z ($\lambda i,z$ is the correspondent membership function), [14, 15], are given by:

$$RO_i\text{: } \textbf{\textit{If}}\ \mathrm{Y_{i,t}}\ \textbf{\textit{is}}\ \mathrm{A_i}\ \textbf{\textit{then}}\ \mathrm{Y_{i,t} = a_i^T Y_t + b_i,\ i=1,2,3} \tag{15}$$

$$\mathrm{Y_z=[\Sigma_i\ \gamma_{i,z}\ a_i^T]\ Y_t + \Sigma_i\ \gamma_{i,z}\ b_i\ ;\ \gamma_{i,z} = \lambda_{i,z} / (\Sigma_j\ \lambda_{j,z})} \tag{16}$$

The indirect adaptive neural control algorithm, which is the consequent part of the local fuzzy control rule RCi (13) is viewed as a Sliding Mode Control (SMC), [13, 14], using the

ISBN: 978-989-95079-6-8

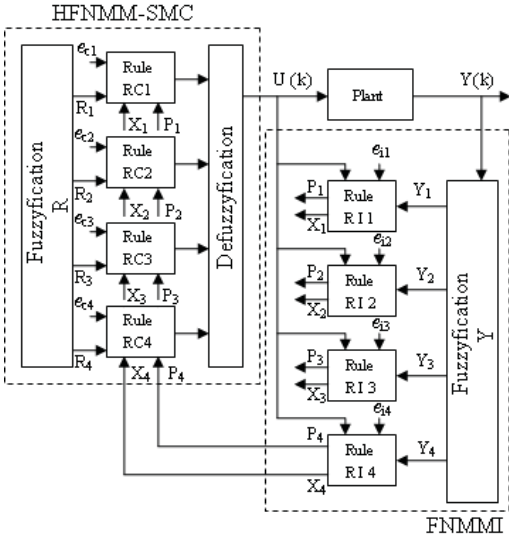


Figure 5: Block-diagram of the FNMM Control system.

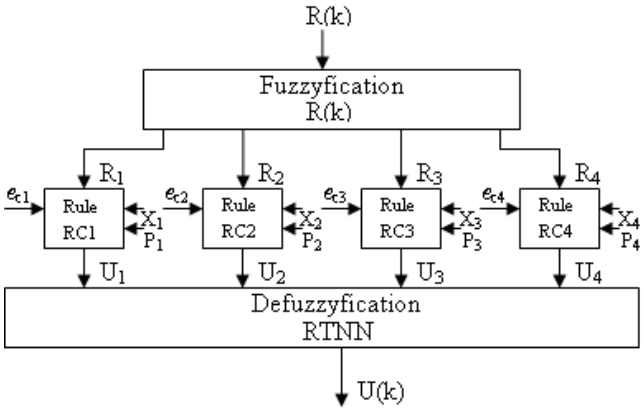


Figure 6: Detailed block-diagram of the HFNMM controller

parameters and states issued by the correspondent identification local fuzzy rule RIi (8). The equivalent control capable to lead the system to the sliding surface in the case when L>M is given by the equations:

$$U_{eq}(k) = (CB)^{+}\left[-CAX(k) + R(k+1) + \sum_{i=1}^{P}\gamma_i E(k-i+1)\right] + Of \qquad (17)$$
$$(CB)^{+} = [(CB)^{T}(CB)]^{-1}(CB)^{T}$$

Here the added offset Of is a learnable M-dimensional constant vector which is learnt using a simple delta rule (see [3]), where the error of the plant input is obtained backpropagating the output error through the adjoint RTNN model (see [18, 14]). The SMC avoiding chattering is taken using a saturation function inside a bounded control level Uo, taking into account plant uncertainties. So the SMC takes the form:

$$U(k) = \begin{cases} U_{eq}(k), & if \ \|U_{eq}(k)\| < U_0 \\ -U_0 U_{eq}(k)/\|U_{eq}(k)\|, & if \ \|U_{eq}(k)\| \geq U_0 \end{cases} \qquad (18)$$

The proposed SMC cope with the characteristics of the wide class of plant model reduction neural control with reference model, and represents an indirect adaptive neural control, given by [11, 13, 14].

3.1 Description of the RTNN and its Learning Algorithm

The RTNN topology, including thresholds in both layers, and its learning algorithm, which appeared in the consequent part of the identification fuzzy rule as a learning procedure, are described in vector-matrix form as:

$$X(k+1) = AX(k) + BU(k);\ B = [B_1 ; B_0];\ U^T = [U_1 ; U_2] \quad (19)$$
$$Z_1(k) = G[X(k)] \quad (20)$$
$$V(k) = CZ(k);\ C = [C_1 ; C_0];\ Z^T = [Z_1 ; Z_2] \quad (21)$$
$$Y(k) = F[V(k)];\ A = \text{block-diag}\ (Ai),\ |Ai| < 1 \quad (22)$$
$$W(k+1) = W(k) + \eta\,\Delta W(k) + \alpha\,\Delta W_{ij}(k-1) \quad (23)$$
$$E(k) = T(k) - Y(k);\ E_1(k) = F'[Y(k)]\,E(k) \quad (24)$$
$$\Delta C(k) = E_1(k)\,Z^T(k);\ F'[Y(k)] = [1 - Y^2(k)] \quad (25)$$
$$E_3(k) = G'[Z(k)]\,E_2(k);\ E_2(k) = C^T(k)\,E_1(k) \quad (26)$$
$$\Delta B(k) = E_3(k)\,U^T(k);\ G'[Z(k)] = [1 - Z^2(k)] \quad (27)$$
$$\Delta A(k) = E_3(k)\,X^T(k);\ Vec(\Delta A(k)) = E_3(k) \square X(k) \quad (28)$$

Where: X, Y, U are vectors of state, output, and augmented input with dimensions N, L, (M+1), respectively, Z is an (L+1) –dimensional input of the feedforward output layer, where Z_1 and U_1 are the (Nx1) output and (Mx1) input of the hidden layer; the constant scalar threshold entries are $Z_2 = -1$, $U_2 = -1$, respectively; V is a (Lx1) pre-synaptic activity of the output layer; T is the (Lx1) plant output vector, considered as a RNN reference; A is (NxN) block-diagonal weight matrix; B and C are [Nx(M+1)] and [Lx(N+1)]-augmented weight matrices; B_0 and C_0 are (Nx1) and (Lx1) threshold weights of the hidden and output layers; F[.], G[.] are vector-valued tanh(.)-activation functions with corresponding dimensions; F'[.], G'[.] are the derivatives of these tanh(.) functions; W is a general weight, denoting each weight matrix (C, A, B) in the RTNN model, to be updated; ΔW (ΔC, ΔA, ΔB), is the weight correction of W; η, α are learning rate parameters; ΔC is an weight correction of the learned matrix C; ΔB, ΔA are weight corrections of the learned matrices B, A; the diagonal of the matrix A is denoted by Vec (.) where (28) represents its learning as an element-by-element vector product; E, E_1, E_2, E_3, are error vectors with appropriate dimensions. The stability of the RTNN model is assured by the activation functions [-1, 1] bounds and by the local stability weight bound condition, given by (22). The learning procedure having forward and backward steps (19)-(28), could be denoted by Y(k) = Π (L, M, N, Y_d, U, X, A, B, C, E). The defuzzyfication learning procedure, corresponding to the single layer perceptron, denoted by (14) is a simple delta rule (see [3]), where the error of the plant controlled input is obtained backpropagating the output error through the adjoint RTNN model, [18].

4 Simulation Results

The decentralized FNMM identifier used a set of four T-S fuzzy rules containing in its consequent part RTNN learning procedures (see Fig. 3). The topology of the first three RTNNs is (2-6-4) (2 inputs, 6 neurons in the hidden layer, 4 outputs) and the last one has topology (2-6-2) corresponding to the fixed bed plant behavior in each collocation point and the recirculation tank. The RTNNs identified the following fixed bed variables: X_1 (acidogenic bacteria), X_2

ISBN: 978-989-95079-6-8

(methanogenic bacteria), S_1 (chemical oxygen demand) and S_2 (volatile fatty acids), in the following collocation points, z=0.25H, z=0.5H, z=0.75H, and the following variables in the recirculation tank: S_{1T} (chemical oxygen demand) and S_{2T} (volatile fatty acids). The graphical simulation results of RTNNs learning are obtained on-line during 100 days with a step of 0.1 day (To=0.1 sec.; Nt=1000 iterations). The learning rate parameters of RTNN have small values which are different for the different measurement point variables. The Fig. 7, Fig.8, Fig. 9, Fig. 10 showed similar three dimensional and two dimensional graphical simulation results of the direct and indirect decentralized FNMM control of X_1, X_2, where the outputs of the plant are compared with the reference signals at the collocation points. The reference signals for all variables are proportional train of pulses with uniform duration and random amplitude. The Means Squared Error (MSE%) of identification, direct, indirect and optimal control for each output signal and each measurement point are given on Table 1, 2, 3, 4, respectively. The comparison showed a slight preference of the indirect FNMM-SMC control over the direct FNMM control, due to the better adaptation of the first low.

Table 1: MSE% of the decentralized FNMM approximation of the bioprocess output variables

Coll. point	X_1	X_2	S_1/S_{1T}	S_2/S_{2T}
z=0.25H	1.2524e-8	6.5791e-8	2.9615e-5	4.3302e-4
z=0.5H	5.0180e-9	2.9067e-8	1.1840e-8	2.7851e-6
z=0.75H	1.0487e-9	2.7977e-9	9.3562e-5	2.8941e-4
Recir Tank	-	-	8.6967e-7	2.0205e-6

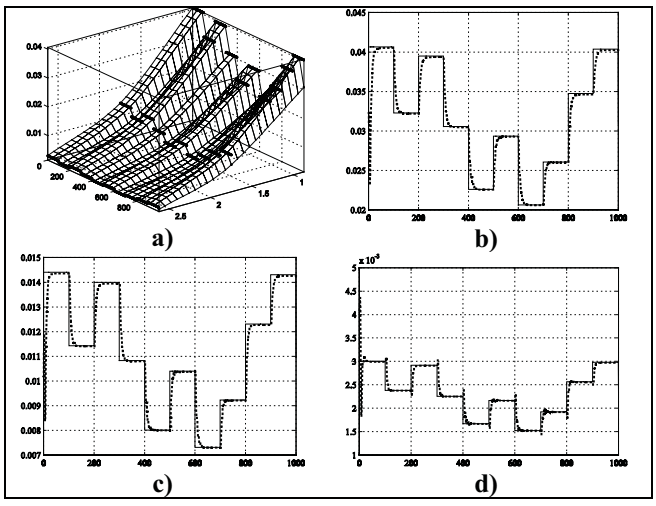


Figure 7: Results of the direct decentralized FNMM control of X_1 (acidogenic bacteria in the fixed bed) (dotted line-plant output, continuous-reference); a) 3d view of X_1 ; b) Ref vs X_1 in z=0.25H; c) Ref vs X_1 in z=0.5H; d) Ref vs X_1 in z=0.75H.

Table 2: MSE% of the direct decentralized FNMM control of the bioprocess output variables

Coll. point	X_1	X_2	S_1/S_{1T}	S_2/S_{2T}
z=0.25H	1.2524e-8	6.5791e-8	2.9615e-5	4.3302e-4
z=0.5H	5.0180e-9	2.9067e-8	1.1840e-8	2.7851e-6
z=0.75H	1.0487e-9	2.7977e-9	9.3562e-5	2.8941e-4
Recir Tank	-	-	8.6967e-5	2.0205e-4

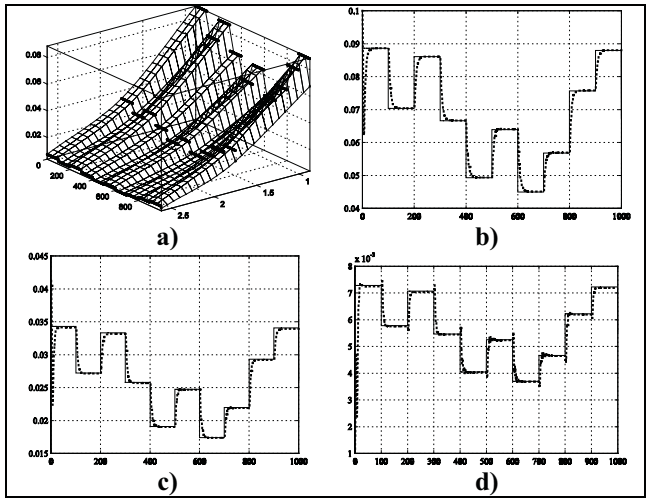


Figure 8: Results of the direct decentralized FNMM control of X_2 (methanogenic bacteria in the fixed bed) (dotted line-plant output, continuous-reference); a) 3d view of X_2 ; b) Ref vs X_2 in z=0.25H; c) Ref vs X_2 in z=0.5H; d) Ref vs X_2 in z=0.75H.

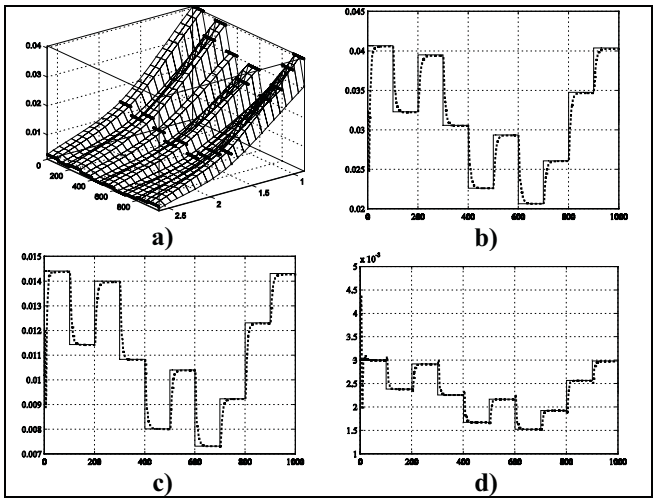


Figure 9: Results of the decentralized FNMM-SMC of X_1 (acidogenic bacteria in the fixed bed) (dotted line-plant output, continuous-reference); a) 3d view of X_1; b) SMC of X_1 in z=0.25H; c) SMC of X_1 in z=0.5H; d) SMC of X_1 in z=0.75H.

Table 3: MSE% of the decentralized FNMM–SMC of the bioprocess output variables

Coll. point	X_1	X_2	S_1/S_{1T}	S_2/S_{2T}
z=0.25H	1.7494E-8	3.0157E-9	3.9538E-6	7.9391E-9
z=0.5H	2.2131E-9	1.9669E-8	4.8951E-7	2.1116E-6
z=0.75H	1.0415E-10	1.3238E-9	2.3548E-8	1.3095E-7
Recir Tank			8.4352E-8	5.8734E-9

Table 4: MSE% of the proportional optimal control of the bioprocess output variables

Coll. point	X_1	X_2	S_1/S_{1T}	S_2/S_{2T}
z=0.25H	5.3057E-8	1.7632E-7	1.1978E-5	2.1078E-5
z=0.5H	6.6925E-9	4.2626E-8	1.4922E-6	4.4276E-6
z=0.75H	3.0440E-10	2.0501E-9	6.8737E-8	2.0178E-7
Recir Tank			2.7323E-7	6.0146E-7

The comparison also showed a slight preference of the FNMM-SMC control over the linearized optimal control, due to the adaptation of the FNMM-SMC control low.

ISBN: 978-989-95079-6-8

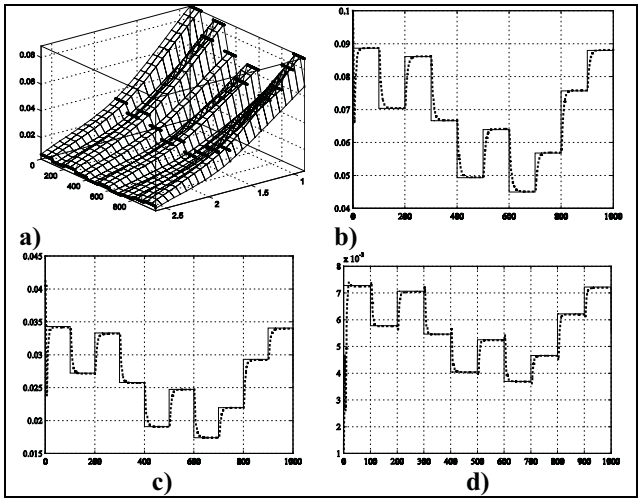


Figure 10: Results of the decentralized FNMM-SMC of X_2 (methanogenic bacteria in the fixed bed) (dotted line-plant output, continuous-reference); a) 3d view of X_2; b) SMC of X_2 in z=0.25H; c) SMC of X_2 in z=0.5H; d) SMC of X_2 in z=0.75H.

5 Conclusions

The paper performed decentralized recurrent fuzzy-neural identification, direct and indirect control of an anaerobic digestion wastewater treatment bioprocess, composed by a fixed bed and a recirculation tank represented a DPS. The simplification of the PDE process model by ODE is realized using the orthogonal collocation method in three collocation points (plus the recirculation tank) represented centers of membership functions of the space fuzzyfied output variables. The obtained from the FNMMI state and parameter information is used by a HFNMM direct and indirect control. All graphical control results exhibited good convergence and precise reference tracking. All obtained comparative numerical identification and control results (final MSE%) exhibited a high precision and showed that the indirect decentralized FNMM control is the better one outperforming the linearized optimal control and the direct decentralized fuzzy-neural control.

Acknowledgment

The Ph.D. student Rosalba Galvan-Guerra is thankful to CONACYT, Mexico for the scholarship received during her studies at the Department of Automatic Control, CINVESTAV-IPN, Mexico City, Mexico.

References

[1] W.H. Ray, *Advanced Process Control*, Butterworths, Boston, London, Singapore, Sydney, Toronto, Wellington, 133-242, 1989.

[2] J.D. Boskovic, and K. S. Narendra, Comparison of Linear, Nonlinear and Neural-Network-Based Adaptive Controllers for a Class of Fed-Batch Fermentation Processes, *Automatica,* 31: 817-840, 1995.

[3] S. Haykin, *Neural Networks, a Comprehensive Foundation*, Second Edition, Prentice-Hall, Upper Saddle River, New Jersey 07458, Section 2.13: 84-89; Section 4.13: 208-213, 1999.

[4] A. Bulsari, and S. Palosari, Application of Neural Networks for System Identification of an Adsorption Column, *Neural Computing and Applications*, 1: 160-165, 1993.

[5] H. Deng, and H.X. Li, Hybrid Intelligence Based Modeling for Nonlinear Distributed Parameter Process with Applications to the Curing Process, *IEEE Transactions on SMC*, 4: 3506-3511, 2003.

[6] H. Deng, and H.X. Li, Spectral-Approximation-Based Intelligent Modeling for Distributed Thermal Processes, *IEEE Transactions on Control Systems Technology*, 13: 686-700, 2005.

[7] R. Gonzalez-Garcia, R. Rico-Martinez, and I. Kevrekidis, Identification of Distributed Parameter Systems: A Neural Net Based Approach, *Computers and Chemical Engineering*, 22 (4-supl. 1): 965-968, 1998.

[8] R. Padhi, S. Balakrishnan, and T. Randolph, Adaptive Critic based Optimal Neuro Control Synthesis for Distributed Parameter Systems, *Automatica*, 37: 1223-1234, 2001.

[9] R. Padhi, and S. Balakrishnan, Proper Orthogonal Decomposition Based Optimal Neurocontrol Synthesis of a Chemical Reactor Process Using Approximate Dynamic Programming, *Neural Networks*, 16: 719-728, 2003.

[10] S. Pietil, and H.N. Koivo, Centralized and Decentralized Neural Network Models for Distributed Parameter Systems, *In: Proc. of the Symposium on Control, Optimization and Supervision, In: Proc. of CESA'96, IMACS Multiconference on Computational Engineering in Systems Applications*, Lille, France, 1043-1048, 1996.

[11] I. Baruch, J.M. Flores, and R. Garrido, A Fuzzy-Neural Recurrent Multi-Model for Systems Identification and Control, *In: Proc. of the ECC'01*, Porto, Portugal, 3540-3545, Sept. 4-7, 2001.

[12] I. Baruch, J.L. Olivares-Guzman, C.R. Mariaca-Gaspar, and R. Galvan-Guerra, A Fuzzy-Neural Hierarchical Multi-Model for Systems Identification and Direct Adaptive Control, *In: P. Melin, O.Castillo, E.G. Ramirez,, J. Kacprzyk, W. Pedrycz (eds.): Analysis and Design of Intelligent Systems Using Soft Computing Techniques,* ASC Vol. 41, Springer-Verlag Berlin Heidelberg, 163-172, 2007.

[13] I. Baruch, J.L. Olivares-Guzman, C.R. Mariaca-Gaspar, and R. Galvan-Guerra, A Sliding Mode Control Using Fuzzy-Neural Hierarchical Multi-Model Identifier, *In: O.Castillo, P. Melin, O.M. Ross, R.S. Cruz, W. Pedrycz, J. Kacprzyk (eds.): Theoretical Advances and Applications of Fuzzy Logic and Soft Computing,* ASC Vol. 42, Springer-Verlag Berlin Heidelberg, 762-771, 2007.

[14] I. Baruch, R. Beltran-Lopez, J.L. Olivares-Guzman, and J.M. Flores, A Fuzzy-Neural Multi-Model for Nonlinear Systems Identification and Control, *Fuzzy-Sets and Systems*, 159: 2650-2667, 2008.

[15] T.Takagi, and M. Sugeno, Fuzzy Identification of Systems and its Applications to Modeling and Control, *IEEE Trans. on SMC*, 15: 116-132, 1985.

[16] F. Aguilar-Garnica, V. Alcaraz-Gonzalez, and V. Gonzalez-Alvarez, Interval Observer Design for an Anaerobic Digestion Process Described by a Distributed Parameter Model, *In: Proc. of the 2-nd Intern. Meeting on Environm. Biotech. and Engng. (2IMEBE),* Mexico City, paper 117: 1-16, 2006.

[17] B. Bialecki, and G. Fairweather, Orthogonal Spline Collocation Methods for Partial Differential Equations, *Journal of Computational and Applied Mathematics*, 128: 55-82, 2001.

[18] E. Wan, and F. Beaufays, Diagrammatic Method for Deriving and Relating Temporal Neural Networks Algorithms, *Neural Computations*, 8: 182-201, 1996.

ISBN: 978-989-95079-6-8

Enforcing Local Properties in Online Learning First Order TS-fuzzy Systems by Incremental Regularization

Nils Rosemann Werner Brockmann Bastian Neumann
University of Osnabrück
Albrechtstraße 28
D–49069 Osnabrück

Abstract— *Embedded systems deseminate more and more. Because their complexity increases and their design time has to be reduced, they have to be increasingly equipped with self-tuning properties. One form is self-adaption of the system behavior, which can potentially lead the system into safety critical states. In order to avoid this and to speed up the self-tuning process, we apply a specific form of regularization, incremental regularization. The SILKE approach has been developed as an incremental regularization scheme for a special class of online learning Takagi-Sugeno fuzzy systems. Its aim is to control the process of self-tuning by guiding the online learning process towards local meta-level characteristics such as a smooth system behavior without outliers. This ability has been investigated experimentally and formally for zero order systems before. This paper now analyzes the regularization ability of the SILKE approach to enforce local smoothness in first order TS-fuzzy systems in order to enlarge the methodological basis for more complex applications.*

Keywords— First order Takagi Sugeno fuzzy systems, incremental learning, self-optimization, incremental regularization, fuzzy control

1 Introduction

1.1 Background

Due to the growing dissemination of embedded technical systems and their higher integration into critical environments, the requirements for the next generation of those technical systems will change. The systems will have to adapt intelligently to different environments, different aims and changes within the environment in order to open up new application areas. This greatly increases the complexity of such new systems. At the same time, it will become necessary to decrease the design time. We tackle these problems within the context of Organic Computing, There one tries to face these challenges by introducing self-x properties into the systems, e.g., self-organization and self-optimization [1]. This means that technical systems are equipped with a much higher flexibility. They have to adapt their behavior based on changing demands and prior experience.

But although such a technical system adapts its behavior at run time, safety and trustworthiness are a must. This is especially problematic as the higher integration into natural, i.e., not specially prepared, environments comes with many sources of uncertainty. Thus, any form of adaption based on uncertain interaction with the environment will be uncertain itself. Accordingly, a self-tuning system has to be both flexible enough to react dynamically to new conditions, and stable enough to protect already learned knowledge from transient disturbances, noise and other uncertainties (*stability-plasticity-dilemma*). And, what is more, at the same time it has to be guaranteed that the system does not enter critical states by learning. This means that the self-tuning process has to be controlled, but without limiting its flexibility.

Within this paper, we address issues of self-tuning the control behavior of an embedded system, i.e., a form of self-adaptive control. This online changing of the behavior is done by *incremental online learning*. It has a strong influence on the stability of the system: In addition to the feedback loop between the control system and the environment into which it is embedded, online learning creates another, indirect feedback loop [2]. Whatever is learned in a certain situation changes the system's (future) behavior, and thus also what is learned in the future. Because of this, small deviations of (initial) conditions or the self-optimization process as such might have arbitrarily high impact and can potentially lead to a chaotic system behavior.

Different candidate function approximators have been used in adaptive control by online learning, each of them fullfilling the different requirements to a varying degree. An overview over the different families of function approximators is given in [3]. Accordingly, one has to choose an appropriate approximator to represent the control behavior in a way which suits the application specific needs, e.g., concerning expressiveness, complexity, interpretability and, what is more in our context, controllability of the learning process itself.

For adaptive nonlinear modeling and control, Takagi-Sugeno systems are a favoured trade-off [4, 5]. We hence build upon a specific variant. Then, in order to employ incremental online learning for adaptive control in uncertain environments, one has to address the stability-flexibility-dilemma. In the next section it is argued that this can be done by an appropriate form of regularization.

1.2 Related Work

In data-driven function approximation, regularization techniques have been successfully applied in many different areas, especially in nonlinear system identification [6] and nonlinear adaptive control [7]. Typical for these applications is that the distribution of data points is dense in one region of the input space and sparse in another. For practical reasons it might be impossible to obtain enough data in a sparse region, e.g., because it is related to critical states of a controlled process. Hence regularization techniques have found widespread use in data-driven fields. The reason is that in many cases data-driven problems are ill-posed because there is no unique solution [8]. This is usually due to a misbalance between the degrees of freedom of the function approximator and the amount of data available.

The idea of regularization is to incorporate additional in-

ISBN: 978-989-95079-6-8

formation about the solution into the problem statement, i.e., to reformulate it [9]. Usually this is some kind of structural or meta-information. A well known example is to assign a penalty to non-zero parameter values [10]. Another example is to assign a penalty to solutions with a high overall curvature [11]. They are called ridge regression or Tikhonov regularization. Despite missing data, sparse regions and large input spaces, regularization is able to improve the generalization performance of the identified models.

Basically, these approaches require all data points to be accessible at once in order to find a globally optimal solution, i.e., they are offline approaches. But for many online applications, there is a change of requirements while new data points are generated (continuously), for example because the system which is to be modeled changes or is modelled from scratch.

Recent approaches for these kind of applications have so far neglected the problem of requiring regularization [4]. In order to tackle the above challenges online, these approaches learn incrementally, i.e., by looking at every data point only once. But without access to the whole dataset, it is not *a priori* clear how to apply the conventional regularization techniques. Other approaches use regularization techniques for ill-posed matrix inversions [12] which is suitable for certain modeling applications.

Because of our background of adaptive nonlinear control by direct adaptation [2], we are developing schemes for *incremental regularization*, i.e., which perform special forms of regularization online, for appropriate function approximators. The control tasks which we address typically have only a low number of input dimensions (less than 10). In the case of more input dimensions, it is usually possible and advisable to decompose the overall problem into subproblems. Thus, the first one of the schemes has been developed for special Takagi-Sugeno fuzzy systems which are especially suited for adaptive nonlinear modeling and control [13, 5]. They employ an all-coverage approach for generating the rule base which allows for a high interpretability albeit limiting the maximal possible number of input dimensions. These fuzzy systems and the incremental regularization, called SILKE approach, will be reviewed in section 2. Up to now, we have developed incremental regularization schemes for zero order Takagi-Sugeno fuzzy systems [14]. They have been investigated experimentally and formally [15, 16]. The purpose of this paper is to extend the SILKE approach to first order Takagi-Sugeno fuzzy systems. In section 3 it is then shown that each incremental application of our approach to a first order Takagi-Sugeno system increases its smoothness. Because of this, the SILKE approach can be seen as a kind of incremental Tikhonov regularization which is also applicable to first order TS-fuzzy systems.

2 Controlled Self-optimization by the SILKE Approach

2.1 Basic Concept

The basis of the SILKE approach is a special class of Takagi-Sugeno fuzzy systems (hereinafter called sTS fuzzy systems) with polynomial functions as rule conclusions. They use linear B-spline input membership functions, such that each input dimension is split by a partition of unity. This allows for an intuitive handling by the system designer and an implementation which requires a very low computational effort. The

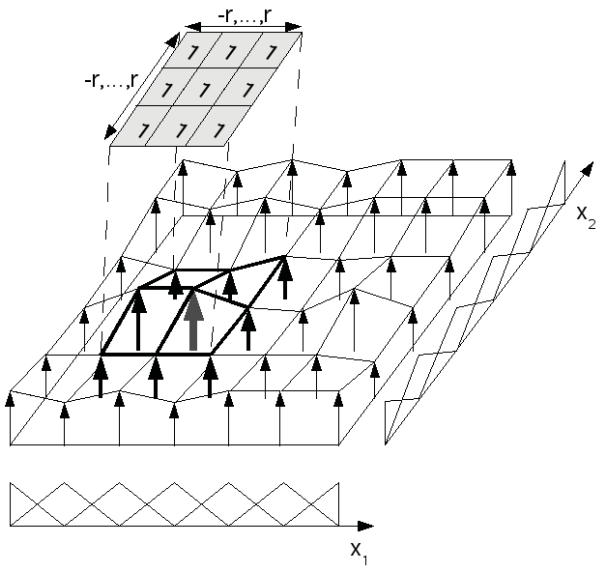


Figure 1: Schematic correspondence of a SILKE-template (shaded in grey) to the conclusions of a part of an sTS-fuzzy rule base in a two-dimensional system. The membership functions determine the lattice of rule conclusions. The thick arrow indicates the active rule under consideration. The numbers within the template are the coefficients of the averaging template mask [15].

rule inference is done by the *sum-prod* technique. These sTS-systems are ideally suited for application to online learning within embedded systems [3].

The idea of the SILKE approach is then to enhance learning in sTS-systems by a correction step which introduces an additional feedback loop at a higher, functional level in order to rate the learning process itself online [17]. This is done by implementing so called *SILKE templates* which determine whether the rule base of the sTS-system complies with predefined meta-level properties. If this *monitoring* detects too strong deviations, then the rule base should be corrected, which closes the feedback loop at a higher level. It should be noted that such a template acts only on those rules which apply to the current system state and which form a local neighborhood (*active rules*), i.e., a template acts only locally and incrementally on each new data point.

Many different meta-level properties can be expressed, e.g., steadiness, monotonicity, local linearity or gradient ratios of the underlying function. One of the most important examples of meta-properties in the field of adaptive control is *smoothness* of the approximated function, i.e., the learned knowledge, because smoothness means that small changes of the inputs do not cause strong changes at the output. So given a process which is smooth, then one wants a rule base which is also smooth. If online learning changes the rule base in a way that it is smooth, then the process of learning harmonizes with its environment. Hence, robustness of learning is increased when faced with learning stimuli which would otherwise violate the smoothness property and smoothness is enforced, for example the effect of outliers is reduced.

In the next section, it is formally shown that this smoothness property can be enforced for first order sTS-systems by

ISBN: 978-989-95079-6-8

using the SILKE approach. For this purpose, the formalism of the SILKE approach is presented in its general form. Afterwards, a measure for local smoothness in zero and first order sTS-systems is developed. This measure is then used to investigate the effect of the SILKE approach in a simplified, one-dimensional scenario for clarity reasons.

2.2 Formalization

Let n be the number of inputs of a first order sTS-fuzzy system. By compilation of an sTS rule base, it becomes normalized, so that one can imagine it as a lattice [17]. This is illustrated by the schematic in Fig. 1. It shows a two-dimensional lattice of a rule base with two input variables, x_1 and x_2. Both inputs are partitioned by triangular membership functions. For every combination of two membership functions, one from each dimension, there is a single rule. Due to normalization, the rules thus form a lattice given by these membership functions. The membership functions can be numbered sequentially in each input dimension. As there is a single rule for each combination of membership functions, one can assign a vector of the individual membership function indixes to each rule. For example, in a two-dimensional rule base, the index vector $(2, 1)$ is assigned to the rule which has the second membership function of the first input dimension and the first membership function of the second input dimension as its antecedents.

The rule conclusions (which are affine functions) can be viewed as being placed upon the nodes of this lattice. It is useful here to identify the affine conclusions by vectors of their parameters, i.e., a function $g : \mathbb{R}^n \to \mathbb{R}$ with

$$g(\mathbf{x}) = c_0 + c_1 x_1 + \ldots + c_n x_n, \quad \mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}, \tag{1}$$

can be identified by the vector of the parameters

$$\mathbf{c} = (c_0, \ldots, c_n)^T \in \mathbb{R}^{n+1}. \tag{2}$$

With this, the rule base of a first-order sTS fuzzy system can be seen as a map $o : \mathbb{N}^n \to \mathbb{R}^{n+1}$ from the space $\mathbb{N}^n$ of rule index vectors to the space $\mathbb{R}^{n+1}$ of parameter vectors. A single node $p \in \mathbb{N}^n$ of the lattice corresponds to a rule, and $o(p)$ to the parameters of an affine function, i.e., the rule conclusion.

A so called SILKE template T is than a map

$$T : \mathbb{N}^n \to \mathbb{R}^{n+1}. \tag{3}$$

The value $\|T(p)\|$ is called *violation degree* of the rule p with respect to T. It indicates how much a rule p violates the meta-level property which is given by T with respect to the neighboring rules. $\|T(p)\| = 0$ means a complete compliance with this property. The higher $\|T(p)\|$, the further rule p is violating the property.

The core of the SILKE approach is then to correct the given rule p according to

$$o(p) \leftarrow (1 - \alpha) \cdot o(p) + \alpha \cdot (o(p) - T(p)). \tag{4}$$

Obviously, if $\|T(p)\| = 0$, the rule is not changed. The *adjustment rate* $\alpha \in [0; 1]$ determines how much the meta-level property is enforced, i.e., $\alpha = 0$ results in no correction of the rule base, whereas $\alpha = 1$ makes it fully compliant with the meta-level property in one step. As said before, such a correction step is done directly after each learning step. Every rule which has been changed by learning is subject to the SILKE template(s). This means that $\alpha = 1$ ignores all changes of the rules by learning. Instead, the meta-level property is fully enforced. Thus, high values of α limit the influence of learning severely but increase the stability.

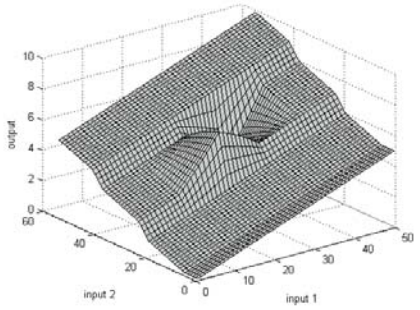


Figure 2: Original graph of a sample two-dimensional sTS-system which is smoothed by the SILKE approach, see Fig. 3 and Fig. 4.

In the following, only SILKE templates are considered which can be represented by a convolution of the rule base:

$$T(p) = o(p) - \frac{1}{(2r+1)^n} \sum_{u \in U} o(p-u)m(u). \tag{5}$$

Here, $r \in \mathbb{N}^+$ is a radius. All neighbor rules with a L_1-distance to the given rule p of less than or equal to r are considered by the template (For illustration, in Fig. 1 an exemplary template is shown with radius $r = 1$). So the sum indices u are in

$$U = [-r, r]^n. \tag{6}$$

The map $m : U \to \mathbb{R}$ is called a *mask* and determines the effect of the SILKE template, i. e., the represented meta-level property, by specific coefficients. Because of this, the mask is the most important part of the SILKE approach concerning design. The designer has to use the right coefficients in the mask in order to represent the desired meta-level properties. The mask thus gives the SILKE approach the necessary flexibility to express a wide range of local meta-level properties. The example in Fig. 1 shows the mask for an averaging template which is given by $m(u) = 1 \forall u \in U$. This mask can be used to guide zero order sTS systems towards a higher smoothness [15]. Another example is the one-dimensional averaging template which only use one of the input dimensions. It is given by $m(u) = 1 \forall u \in \{0\}^{a-1} \times [-r, r] \times \{0\}^{n-a}$ with a denoting the dimension to be used.

3 Smoothness in First Order sTS Fuzzy Systems

In order to extend the formal investigations of the SILKE approach to first order sTS fuzzy systems, they are investigated concerning their smoothness properties. In this section, based on the definition of a measure of smoothness for the individual rules, it is proven that the application of an averaging SILKE

ISBN: 978-989-95079-6-8

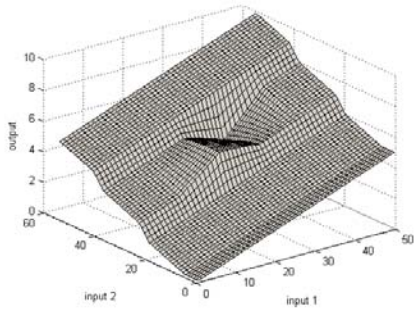


Figure 3: Graph of the sample two-dimensional sTS-system from Fig. 2 after the first application of an averaging template ($\alpha = 0.5$).

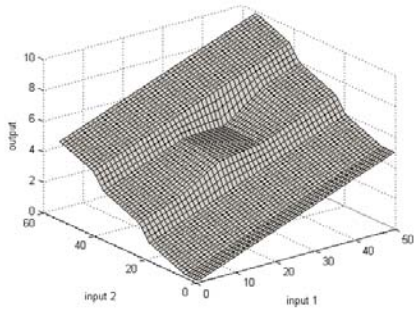


Figure 4: Graph of the sample two-dimensional sTS-system from Fig. 2 after applying an averaging template for two times ($\alpha = 0.5$).

template to a rule increases its smoothness. This can be seen as a form of incremental regularization.

For clarity reasons, we first consider a one dimensional first order sTS-fuzzy system f with triangular membership functions which sum up to 1. Then the input space is covered completely by a set of neighboring membership functions. Due to the normalization of the rule base, there is a single rule for each membership function. Thus, the rules can be ordered according to the ordinality of the membership functions so that they can be addressed by an index $i \in \mathbb{N}$, such that the indices i and $i+1$ address neighboring rules. The rule conclusion of rule number i is given by

$$f_i(x) = c_{0,i} + c_{1,i} \cdot x \tag{7}$$

with x as the input to the fuzzy system.

In the following, let $p_i \in \mathbb{R}$ denote the core of the membership function of rule i. As the membership functions are triangular, it is convenient here to distinguish between the left and the right part, $\mu_{i,l}(x)$ and $\mu_{i,r}(x)$, respectively, of each of them:

$$\begin{aligned} \mu_{i,l}(x) &= \frac{x - p_{i-1}}{p_i - p_{i-1}}, \quad p_{i-1} \le x < p_i \\ \mu_{i,r}(x) &= \frac{p_{i+1} - x}{p_{i+1} - p_i}, \quad p_i \le x < p_{i+1}. \end{aligned} \tag{8}$$

We are now considering the smoothness of the sTS fuzzy system in the environment of p_i. The functional behavior in this environment is determined by the neighboring rules, i.e., the rules $i-1$, i and $i+1$. To analyze the smoothness, we first calculate the first derivative of the sTS fuzzy system on the left and on the right of p_i. Between p_{i-1} and p_i, the output of the fuzzy system is then given by

$$f(x) = \mu_{i-1,r}(x) f_{i-1}(x) + \mu_{i,l}(x) f_i(x) \tag{9}$$

because of the linear interpolation due to sum-prod inference. Similarly, between p_i and p_{i+1}, the output is

$$f(x) = \mu_{i,r}(x) f_i(x) + \mu_{i+1,l}(x) f_{i+1}(x). \tag{10}$$

With these preparations, one can calculate the first derivative f' of f between the cores. For $p_{i-1} < x < p_i$ one gets

$$\begin{aligned} f'(x) &= \mu'_{i-1,r}(x) f_{i-1}(x) + \mu_{i-1,r}(x) f'_{i-1}(x) \\ &\quad + \mu'_{i,l}(x) f_i(x) + \mu_{i,l}(x) f'_i(x) \end{aligned} \tag{11}$$

and for $p_i < x < p_{i+1}$

$$\begin{aligned} f'(x) &= \mu'_{i,r}(x) f_i(x) + \mu_{i,r}(x) f'_i(x) + \\ &\quad \mu'_{i+1,l}(x) f_{i+1}(x) + \mu_{i+1,l}(x) f'_{i+1}(x). \end{aligned} \tag{12}$$

We can now determine the difference between the right and the left limit of $f'(x)$ with $x \to p_i$ and simplify it as the membership functions are either 0 or 1 at p_i:

$$\begin{aligned} \Delta_i &= \lim_{x \searrow p_i} f'(x) - \lim_{x \nearrow p_i} f'(x) \\ &= (\mu'_{i,r}(p_i) - \mu'_{i,l}(p_i)) f_i(p_i) \\ &\quad + \mu_{i+1,l}(p_i) f_{i+1}(p_i) - \mu'_{i-1,r}(p_i) f_{i-1}(p_i) \\ &= \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) (c_{0,i} + c_{1,i} p_i) \\ &\quad + \frac{1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad + \frac{1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i). \end{aligned} \tag{13}$$

ISBN: 978-989-95079-6-8

This expression means that in general one obtains a different result for the first derivative when approaching p_i from left than from right. To put it otherwise, Δ_i is a measure of the violation of the smoothness property at rule i.

Now suppose that our sTS fuzzy system contains a rule i with $\Delta_i \neq 0$. We will now prove that we can use a special SILKE template to reduce the violation of the smoothness property. This special SILKE template is given by the radius $r = 1$ and the mask

$$\begin{aligned} m(-1) &= \frac{3(p_{i+1} - p_i)}{p_{i+1} - p_{i-1}} \\ m(0) &= 0 \\ m(1) &= \frac{3(p_i - p_{i-1})}{p_{i+1} - p_{i-1}}. \end{aligned} \tag{14}$$

The application of this mask according to (4) and (5) changes $c_{0,i}$ and $c_{1,i}$ to $\hat{c}_{0,i}$ and $\hat{c}_{1,i}$, respectively:

$$\begin{aligned} \hat{c}_{a,i} &= c_{a,i} + \alpha \left(\frac{p_{i+1} - p_i}{p_{i+1} - p_{i-1}} c_{a,i-1} \right. \\ &\quad \left. + \frac{p_i - p_{i-1}}{p_{i+1} - p_{i-1}} c_{a,i+1} - c_{a,i} \right), \; a \in \{1, 2\}. \end{aligned} \tag{15}$$

This yields an updated value $\hat{\Delta}_i$ for the violation of the smoothness property. By comparing the updated value $\hat{\Delta}_i$ with the old value Δ_i, we can prove that the SILKE template (14) has the desired effect:

$$\begin{aligned} \hat{\Delta}_i &= \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) (\hat{c}_{0,i} + \hat{c}_{1,i} p_i) \\ &\quad + \frac{1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad + \frac{1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \\ &= \Delta_i + \alpha \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) \\ &\quad \cdot \left(\frac{p_{i+1} - p_i}{p_{i+1} - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \right. \\ &\quad + \frac{p_i - p_{i-1}}{p_{i+1} - p_{i-1}} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad \left. + \frac{-1}{1} (c_{0,i} + c_{1,i} p_i) \right) \\ &= \Delta_i + \alpha \left(\left(\frac{1}{p_{i+1} - p_i} + \frac{1}{p_i - p_{i-1}} \right) (c_{0,1} + c_{1,i} p_i) \right. \\ &\quad + \frac{-1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad \left. + \frac{-1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \right) \\ &= \Delta_i - \alpha \Delta_i \\ &= (1 - \alpha) \Delta_i. \end{aligned} \tag{16}$$

From this, it can be concluded that

$$|\hat{\Delta}_i| \leq |\Delta_i|. \tag{17}$$

The effect of this parameter update on a sample two-dimensional sTS-system is illustrated in Fig. 2 to 4. The figures demonstrate that each application of (4) on the central rule (in this case with $\alpha = 0.5$) increases the smoothness.

4 Discussion

The above investigations have shown that first order sTS-fuzzy systems with triangular membership functions have discontinuities of their first derivative precisely at the cores of the input membership functions, i.e., at the nodes of the normalized rule base. These discontinuities correspond to violating the *local smoothness* of the system behavior. It was shown for the one dimensional case that the SILKE approach reduces the discontinuity locally by a special averaging template, i.e., it increases the local smoothness. The effect depends on the adjustment rate α. For $\alpha = 0$, the rule base is not changed at all. But for higher values of α, the discontinuity of the first derivative is more and more reduced. Thus, the learning process is guided towards the meta-property of global smoothness expressed by the template function. Up to now, the formal proof is for the one dimensional case, although the SILKE algorithm is of course applicable to arbitrary dimensionality. We have shown by examplary investigations that the smoothing property also holds for higher dimensions. The extension of the formal proof is currently in progress.

In contrast to prior work on regularization, the SILKE approach has the advantage of working incrementally. It is thus well suited for incremental online learning which is required for self-adapting systems or building models online. During online operation, it is applied periodically and affects all currently active rules of the fuzzy system. This way, SILKE templates get a global influence although they work locally, which results in a low computational effort and good real-time capabilities. But what is more, sufficient plasticity of learning is still given due to locality, although the learning process is guided towards global meta-level properties.

Incremental online learning in a closed-loop setting is potentially chaotic. But with the SILKE approach, the influence of learning on the behavior of an embedded system in its environment can be guided by correcting the learned knowledge locally towards higher smoothness or towards compliance with other local meta-level properties when other templates are used. Depending on the adjustment rate, this yields a system which is very flexible in the face of learning stimuli that comply the meta-properties, but which is also very robust against learning stimuli which violate these meta-properties. So for this case, self-adaptive systems can fight outliers and overcome the stability-plasticity-dilemma to a certain extent. Based on this it is clear that the focus of this work is not to achieve the highest possible modeling quality, but to stabilize adaptive control systems.

At system level, the presented approach is an effective and efficient method for controlling incremental online learning without limiting the required flexibility. Such a system can cope with changes online and in a reliable way. It is based on the use of rules. On the one hand this aids engineerability at design time. On the other hand, rules which have been modified by self-tuning can also be analyzed directly. By that good engineerability and high trustworthiness are achieved.

In future work, the interplay between closed-loop incremental online learning and the SILKE approach will be investigated especially for the application of other template functions to zero and first order sTS-fuzzy systems in order to achieve other meta-properties such as monotony and curvature. As before, this will be done in experiments as well as

ISBN: 978-989-95079-6-8

in formal analysis. In addition, the extension of the results to more than one dimension has to be addressed.

Acknowledgement

This project is funded by the German Research Foundation DFG in the framework of the *Organic Computing* priority research program under No. BR 1980/3-1.

References

[1] C. Müller-Schloer, C. von der Malsburg, and R. P. Würtz. Organic Computing. *Informatik-Spektrum*, 27(4):332–336, 2004.

[2] W. Brockmann. Online Machine Learning For Adaptive Control. *IEEE Int. Work. Emerging Technologies and Factory Automation, 1992*, pages 190–195, 1992.

[3] J. Farrell and M. Polycarpou. *Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. Wiley Interscience, Hoboken, New Jersey, 2006.

[4] P. P. Angelov and D. P. Filev. An Approach to Online Identification of Takagi-Sugeno Fuzzy Models. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):484–498, 2004.

[5] T. A. Johansen and R. Murray-Smith. The operating regime approach to nonlinear modelling and control. *Multiple Model Approaches to Modelling and Control*, pages 3–72, 1997.

[6] T. A. Johansen. Robust identification of Takagi-Sugeno-Kang fuzzy models using regularization. In *Fifth IEEE Int. Conf. Fuzzy Systems*, volume 1, pages 180–193, 1996.

[7] J. Farrell and M. Polycarpou. *Adaptive Approximation Based Control*. Wiley–Interscience, 2006.

[8] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.

[9] F. Bauer, S. Pereverzev, and L. Rosasco. On regularization algorithms in learning theory. *Journal of Complexity*, 23(1):52–72, 2007.

[10] A.E. Hoerl and R.W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 42(1):80–86, 2000.

[11] A. Tikhonov. On the stability of inverse problems. *CR (Dokl.) Acad. Sci. URSS, n. Ser.*, 39:176–179, 1943.

[12] E. Lughofer and S. Kindermann. Improving the Robustness of Data-driven Fuzzy Systems with Regularization. In *Proc. Int. Conf. Fuzzy Systems*, pages 703–709. IEEE, 2008.

[13] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

[14] W. Brockmann and A. Horst. Stabilizing the Convergence of Online-Learning in Neuro-Fuzzy Systems by an Immune System-inspired Approach. In *Proc. IEEE Int. Conf. Fuzzy Systems, FUZZ-IEEE*, pages 351–356. IEEE, 2007.

[15] N. Rosemann and W. Brockmann. Concept for Controlled Self-optimization in Online Learning Neuro-fuzzy Systems. In J. Hertzberg, M. Beetz, and R. Englert, editors, *KI 2007: Advances in Artificial Intelligence, LNAI*, volume 4667, pages 498–501. Springer, Berlin, 2007.

[16] N. Rosemann, B. Neumann, and W. Brockmann. Formale Eigenschaften des SILKE-Ansatzes zur Kontrolle selbstoptimierender Systeme. In H.-G. Hegering and A. et al. Lehmann, editors, *Proc. Informatik 2008 Beherrschbare Systeme – dank Informatik, LNI*, volume 134, pages 755–762. Gesellschaft für Informatik, Bonn, 2008.

[17] W. Brockmann and N. Rosemann. Instantaneous Anomaly Detection in Online Learning Fuzzy-Systems. In F. Hoffmann and O. et al. Cordon, editors, *3rd Int. Workshop on Genetic and Evolving Fuzzy Systems*, pages 23–28. IEEE Press, Piscataway, 2008.

ISBN: 978-989-95079-6-8

Dynamic vs. static decision strategies in adversarial reasoning

David A. Pelta[1] Ronald R. Yager[2]
1. Models of Decision and Optimization Research Group
Department of Computer Science and A.I., University of Granada,
C/Periodista Daniel Saucedo s/n, 18071 Granada, Spain
2. Machine Intelligence Institute, Iona College
New Rochelle, NY 10801, USA
Email: dpelta@decsai.ugr.es, yager@panix.com

Abstract— Adversarial decision making is aimed at determining optimal decision strategies to deal with an adversarial and adaptive opponent. One defense against this adversary is to make decisions that are intended to confuse him, although our rewards can be diminished. It is assumed that making decisions in an uncertain environment is a hard task. However, this situation is of upmost interest in the case of adversarial reasoning as what we want is to *force the presence of uncertainty* in order to confuse the adversary in situations of repeated conflicting encounters. Using simulations, the use of dynamic vs. static decision strategies is analyzed. The main conclusions are: a) the use of the proposed dynamic strategies has sense, b) the presence of an adversary may produce a decrease of, at least, 45% with respect to the theoretical best payoff and c) the relation between this reduction and the way the uncertainty is forced should be further investigated.

Keywords— Adversarial reasoning, uncertain environment, decision strategies, simulation

1 Introduction

Adversarial decision is largely about understanding the minds and actions of one's opponent. It is relevant to a broad range of problems where the actors are actively and consciously contesting at least some of each others' objectives and actions [1]. The field is also known as decision making in the presence of adversaries or adversarial reasoning.

In its most basic form, adversarial decision making involves two participants, white and black, each of which chooses an action to respond to a given event without knowing the choice of the other. As a result of these choices, a payoff is assigned to the participants. When this scenario is repeated many times, i.e. situations of repeated conflicting encounters arise, then the situation becomes complex as the participants have the possibility to learn the others strategy. Examples of this type can be found in the military field, but also in problems of real-time strategy games, government vs government conflicts, economic adversarial domains, team sports (e.g., RoboCup), competitions (e.g., Poker), etc. [1]

Adversarial decision making is aimed at determining optimal strategies (for white) against an adversarial and adaptive opponent (black). One defense against this adversary is to make decisions that are intended to confuse him, although white's rewards can be diminished.

It is assumed that making decisions in an uncertain environment is a hard task. However, this situation is of upmost interest in the case of adversarial reasoning as what white wants is to make its behaviour as uncertain or unpredictable as possible. In other words, white wants to *force the presence of uncertainty* in order to confuse the adversary while its payoff is as less affected as possible.

In previous work [2], we proposed a model to study the balance between the level of confusion induced and the payoff obtained and we concluded that one way to produce uncertainty is through decision strategies for white that contain certain amount of randomness. Here we focus on learning strategies that white can use as a means of optimizing his payoffs in situations of repeated conflicting encounters. Essentially we are studying how white can defend against an opponent who is trying to learn their decision rules.

In this paper, we want to analyze the case where white's decision strategy is not constant along the time, but modified following certain rules. We explore two alternatives, one is to vary the number of candidates alternatives in terms of their associated payoffs and second is based on the basic concept of α-cuts, where the value of α is varied.

The contribution is organized as follows: some basic concepts on adversarial reasoning are outlined in Section 2. Then, Section 3 describes the main characteristics and components of the model used. Section 4 introduces static decision strategies for both agents and then, shows how they can be transformed into dynamic ones. In Section 5 we describe the computational experiments performed and the results obtained and finally, Section 6 is devoted to discussions and further work.

2 Adversarial Reasoning

As stated before, adversarial decision making is largely about understanding the minds and actions of one's opponent. A typical example is the threat of terrorism and other applications in Defense, but it is possible to envisage less dramatic applications in computer games where the user is the adversary and the computer characters are provided with adversarial reasoning features in order to enhance the quality, hardness and adaptivity of the game. The development of intelligent training systems is also an interesting field.

The threat of terrorism, and in particular the 9/11 event, fueled the investments and interest in the development of computational tools and techniques for adversarial reasoning. However, the field has earlier developments. For example, almost twenty years ago, P. Thagard [3] states

> In *adversarial problem solving*, one must anticipate, understand and counteract the actions of an opponent. Military strategy, business, and game playing all require an agent to construct a model of an

ISBN: 978-989-95079-6-8

opponent that includes the opponent's model of the agent.

Game theory is perceived as a natural good choice to deal with adversarial reasoning problems. For example, a brief survey of techniques where the combination of game theory with other approaches is highlighted, jointly with probabilistic risk analysis and stochastic games is presented in [4]

However, nowadays it is assumed that the field transcends the boundaries of game theory [1]. As stated in [5]: "we argue that practical adversarial reasoning calls for a broader range of disciplines: artificial intelligence planning, cognitive modeling, control theory, and machine learning in addition to game theory. An effective approach to problems of adversarial reasoning must combine contributions from disciplines that unfortunately rarely come together".

3 The model

The framework used to conduct our study is a slight modification of the previously proposed in [2].

It is based on two agents white W and black B (the adversary), a set of possible inputs or events $E = \{e_1, e_2, \ldots, e_n\}$ issued by a third agent R, and a fuzzy set of potential responses or actions $A_i = \{a_1, a_2, \ldots, a_m\}$ associated with every event. These fuzzy sets are organized as rows in a matrix P as :

$$P(n \times m) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ p_{31} & p_{32} & \cdots & p_{3m} \\ \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{pmatrix}$$

where $p_{ij} \in [0, 1]$ is the level of suitability of action j to respond to the event i. More precisely, p_{ij} is the degree of membership of action j to the fuzzy set of *suitable actions* associated with event e_i. We do not require these fuzzy sets to be normalized.

Agent W has a strategy to decide which action to take given a particular event e_k and perfect knowledge of matrix P. The aim for W is to maximize the sum of the profits or payoffs given a set of inputs. These inputs or events are issued one at a time by R and, in principle, they are independent.

The payoff of a given action is proportional to its suitability with respect to the given event. For the sake of simplicity, in this contribution we assume that the payoff is the value of suitability.

Agent B wishes to learn the actions that W is going to take given a particular input e_k so as to reduce agent W payoff. Agent B does not know matrix P. It has access to the decisions made previously by W and, if the guess matched the decision taken by W, then B obtains some reward. We may think the situation as an "imitation game", where the aim for W is to avoid being imitated.

A graphical view of the model is shown in Figure 1 while the whole procedure is described in Algorithm 1.

The payoff's calculation for W at stage j is defined as:

$$p' = p_{jk} \times F(a_g, a_k) \tag{1}$$

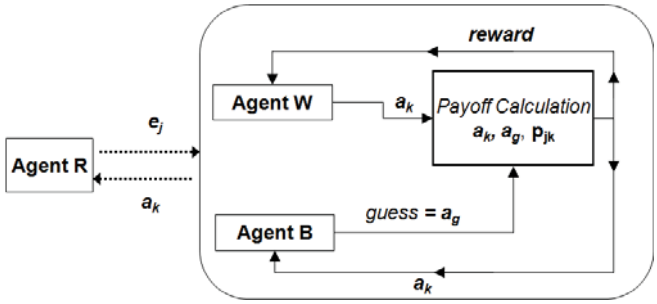


Figure 1: Graphical representation of the model. Events e_j are issued by agent R while response or actions a_k are taken by agent W.

Algorithm 1 Sequence of steps in the model.

for $j = 1$ to N **do**
  A new event e_j arises.
  Agent B "guesses" an action a_g
  Agent W determines an action a_k
  Calculate payoff for W
  Agent B records the pair e_j, a_k
end for

where F is:

$$F(a, b) = \begin{cases} 0 & \text{if a = b} \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

As stated before, the aim for W is to maximize the sum of the payoffs. In other words, a strategy for W should be aimed at defining a sequence of actions that gives the higher payoff while avoiding being correctly guessed.

The terms *event* and *action* used here should be understood in a broad sense. For example, an event may represent a particular simulation scenario while an action may represent a full plan. Also, the actions may represent a Dempster-Schaffer belief structure to reflect the ideas posed in [6].

4 Modeling the Behavior of the Agents

In this section, we provide alternatives for modeling the behavior of both agents. For simplicity, we assume that the inputs issued by agent R are equiprobable and that the number of inputs equals the number of actions (i.e, the payoff matrix is square).

4.1 Strategies for Agent B

Agent B applies a very simple frequency-based decision strategy. We define a matrix of observations O with $M \times M$ dimensions, where each O_{ij} stores the number of times that action i was observed (from W) when the event was e_j. Given an event e_j, the following decision strategy for B is used:

Proportional to the Frequency (PF): the probability of selecting an action i is proportional to O_{ij} (the observed frequency from agent W) [2].

4.2 Strategies for Agent W

Agent W knows he is being observed, so the idea is to change its behavior in order to confuse the adversary. The agent needs to take sub-optimal decisions in order to get benefits in the long term.

ISBN: 978-989-95079-6-8

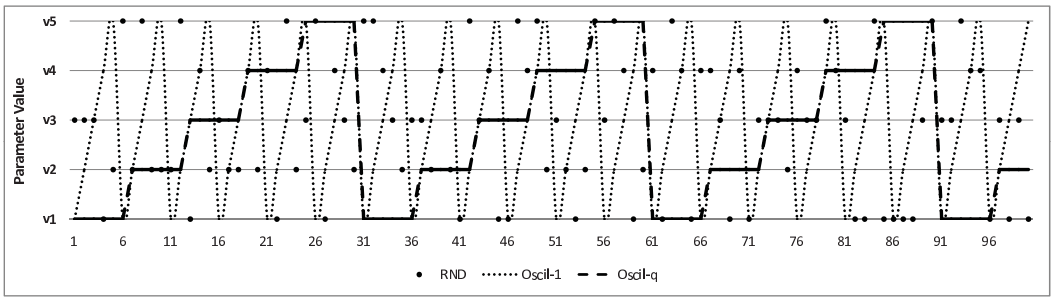


Figure 2: Adaptation schemes for the parameters in the dynamic strategies. The example shows the case where a parameter takes values from a set of possible discrete alternatives $\{v_1, v_2, v_3, v_4, v_5\}$

Given an observed input e_i and the suitability matrix P, we propose two basic strategies:

1. *Random-Among $< k >$ Best Actions (R-k-B)*: Select randomly an alternative among the k most suitable ones [2].
2. *Random-From α-cut (R-α)*: select randomly an alternative among those with a minimum level of suitability α.

We remark here that these strategies are not *mixed strategies* in the sense of game theory as we do not define a probability distribution over the set of alternatives. Once a subset of alternatives is selected, then any of them can be selected with equal probability.

It is clear that both parameters, k and α, can be used to control the amount of uncertainty in the action selection. Some specific values for those parameters lead to interesting behaviours:

- $k = 1$: always select the most suitable action.
- $k = M$: any action is considered equally suitable.
- $0 < k < M$ lead to a behaviour where suboptimal actions may have the chance of being selected.

Particular cases for α:

- $\alpha = 1$: always select the action whose level of suitability is 1. This is not a good strategy as it is easily learnable.
- $\alpha = 0$: any action is considered equally suitable.
- $0 < \alpha < 1$ lead to a behaviour where suboptimal actions may have the chance of being selected.

Although at first sight both strategies look similar, there is an important difference. The former strategy (*R-k-B*) is independent of the suitability scale: it just selects the k most suitable alternatives and then takes a decision. The later strategy is based on α-cuts, so if the value of α is high, then it may happen that the subset of alternatives of the corresponding α-cut becomes empty. In this case, an alternative would be randomly selected.

4.2.1 Dynamic Strategies

When α and k are assigned specific values, then a "static strategy" is obtained. Here, we propose a set of "dynamic strategies" where the values of the control parameter are varied along the time following the patterns shown in Fig. 2.

In order to adapt the parameter k, we propose three schemes:

1. *RND*: at each stage, k is a value from a uniform distribution in $[1, M]$.
2. *Oscil-1*: after each event, $k = k + 1$. When $k = M + 1$ then $k = 1$.
3. *Oscil-10*: the same as before but the value of k is changed after ten events.

For the adaptation of α we propose similar schemes:

1. *RND*: at each stage, α is selected randomly from the set $V = \{0.15, 0.3, 0.45, 0.6, 0.75\}$
2. *Oscil-1*: after each event, $i = i + 1$, $\alpha = V[i]$. When $i = M + 1$ then $i = 1$
3. *Oscil-10*: the same as before but the value of α is changed after ten events.

5 Experiments and Results

The aim of the experiment is to evaluate how W's payoff is affected when dynamic vs. static decision strategies are used.

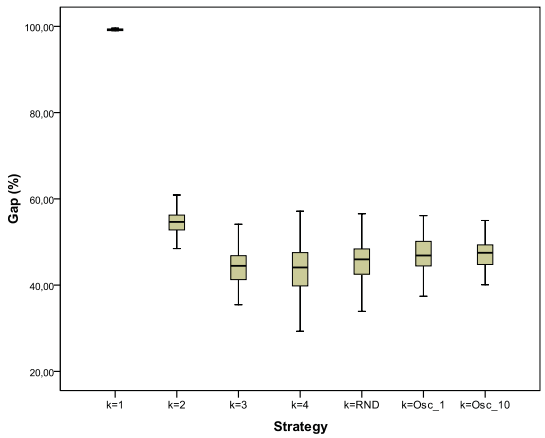
In order to do this, the following considerations are taken. We fix the number of events and alternatives to $M = 5$. Then, for every W's strategy we made 100 repetitions of the scheme shown in Algorithm 1, where $N = 500$. The matrix P is randomly generated each repetition. In this way we avoid potential biases due to particular configuration of values in P. At the end of each repetition, we record:

- Gap: the gap (as percentage) between the payoff obtained by W and the optimum (calculated as the sum of the payoffs associated with the best action for every event),
- $Guess$: the number of times that B correctly guessed the action of W (also as percentage with respect to 500 actions).

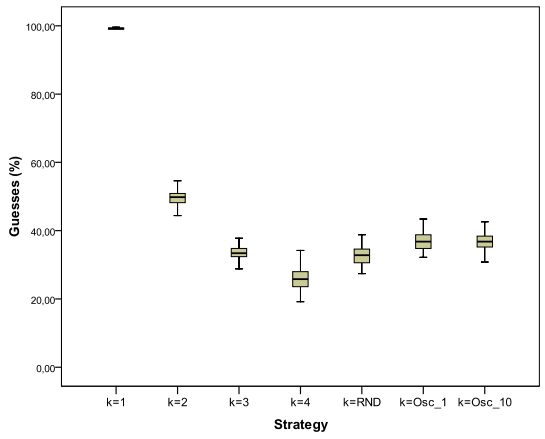
For both cases, the lower the values are, the better the performance of the strategy is.

We evaluated six dynamic strategies (three for α and three for k), four static strategies *R-k-B* using $k = 1, 2, 3, 4$ and five static strategies *R-α* using $\alpha = \{0.15, 0.3, 0.45, 0.6, 0.75\}$.

ISBN: 978-989-95079-6-8



(a)



(b)

Figure 3: Strategy *R-k-B*: average Gap (a) and $Guess$ (b) for each static and dynamic configuration

5.1 Results

In first place we will analyze the results for the *R-k-B*. Figure 3 shows boxplots for the average Gap and $Guess$ (the average number of correct guesses) when the parameter k is fixed or adapted during the repetition.

As it also occurred in [2], the value $k = 1$ is the worst alternative in terms of both measures. In this case, the best action is always selected, thus after a few events the frequency of those actions in the observations matrix kept by B are the only ones different than zero and B always choose them. As k increases, the performance is better on average but the standard deviation is increased. The proposed adaptive schemes for k do not show big differences among them. What is clear, is the reduction in the standard deviation from the $k = RND$ strategy to the one that force oscillations for k every ten events. In terms of the $Guess$ measurement, it is interesting to note that a variation of this single parameter can decrease the number of correct predictions from 50% when $k = 2$ to a lowest value of 25% when $k = 4$.

In order to assess if the differences among the strategies in the average gap have statistical significance, we performed an ANOVA test followed by a post-hoc analysis using Tamhane's test with $p < 0.05$. The results are shown in Table 1; the signs indicate that the average gap between strategies (i, j) (row,col) is different with statistical significance. A '+ sign indicates that strategy i is better, while a '- denotes that j is better. Absence of symbol indicates that the difference had no statistical significance. In this case, the best results are obtained by static strategies with parameters $k = 3$ and $k = 4$. The strategy RND shows a similar performance, however it is not so good to outperform the other dynamic strategies.

R-k-B	1	2	3	4	RND	Osc-1	Osc-10
1		-	-	-	-	-	-
2	+		-	-	-	-	-
3	+	+				+	+
4	+	+				+	+
RND	+	+					
Osc-1	+	+	-	-			
Osc-10	+	+	-	-			

Table 1: Summary of the statistical testing for *R-k-B* strategies. See text for details.

R-α	0.15	0.30	0.45	0.60	0.75	RND	Osc-1	Osc-10
0.15			+	+	+			
0.30			+	+	+			-
0.45	-	-		+	+	-	-	-
0.60	-	-	-		+	-	-	-
0.75	-	-	-	-		-	-	-
RND			+	+	+			
Osc-1			+	+	+			
Osc-10		+	+	+	+			

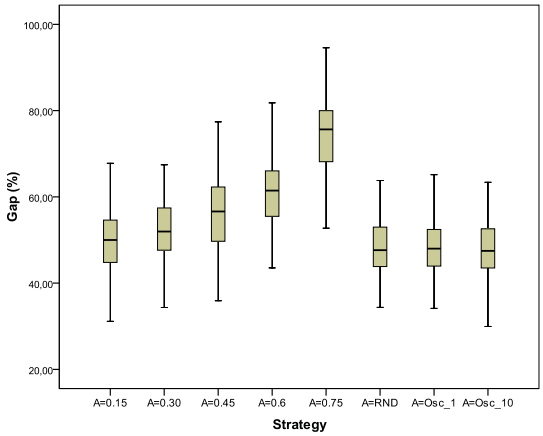
Table 2: Summary of the statistical testing for *R-*α strategies. See text for details.

A similar analysis can be done for the *R-*α strategy. Figure 4 shows boxplots for the average Gap and $Guess$ when the parameter α is fixed or adapted during the repetition. The label A in the plots stands for α.

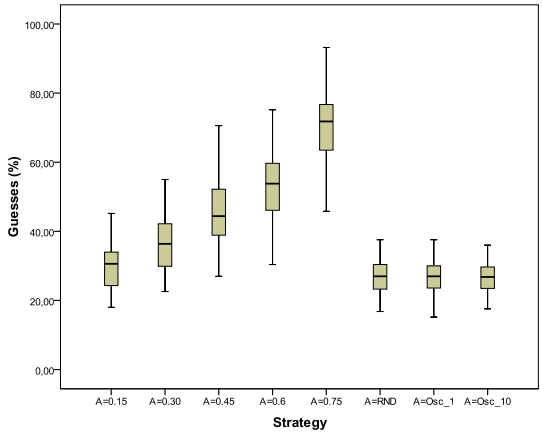
As α increases, the performance becomes worst. The reason is related with the way the matrix P is generated. It may happen that no action has a level of suitability with a degree of membership higher than α. In other words, the corresponding α-cut leads to an empty set of alternatives. In this case, a pure random decision strategy is taken. The counterpart is that now, the proposed adaptive schemes for α are clearly better than the static alternatives. Between the adaptive schemes, no clear differences appeared.

As before, in order to assess if the differences among the strategies in the average Gap have statistical significance, we performed statistical testing. The results are shown in Table 2. It can be confirmed that when using a static strategy, lower values of α like 0.15 or 0.3 led to better results than those obtained with higher ones. However, any of the dynamic strategies proposed, even RND, obtained the same average Gap. Moreover, the strategy *Osc-10* (that changed the α value every ten events) provided better performance than all the static strategies, excepting $\alpha = 0.15$.

ISBN: 978-989-95079-6-8



(a)



(b)

Figure 4: Strategy *R*-α: average Gap (a) and $Guess$ (b) for each static (A stands for α) and dynamic configuration

It is reasonable to assume that a relation exist between the number of correct guesses and the payoff obtained. If the former is high, it is clear that the payoff should be low (this is the case when $k = 1$). But what happen when the number of correct guesses is low? What are the payoff values that can be reached?. Figure 5 shows the average of each measure for every strategy (the case when $k = 1$ is omitted for visualization purposes). Within each kind of strategy (*R-k-B*, R-α), the alternatives are ordered increasingly in terms of Gap.

It is clear that just reducing the number of guesses is not enough to improve the gap. For example, *A=Osc-10* achieved a lower value of $Guess$ than $k = 3$ but a higher value of Gap. This was also already notice in [2], where a purely random strategy led to the lowest value of $Guess$. Interestingly, all the cases where the curve for $Guess$ stands below the one for Gap correspond to strategies based on α-cuts. The reason is not clear and further investigations are needed.

To conclude the analysis, Figure 6 shows several scatter plots displaying the relation between both measures. Every point correspond to a particular repetition, thus 100 point per plot are displayed. Plots on the left correspond to *R*-α strategy while those on the right to *R-k-B*. Three static values are shown per each parameter, while the bottom plot corresponds to the best dynamic strategy. The X axis is the Gap while Y axis shows the measure $Guess$.

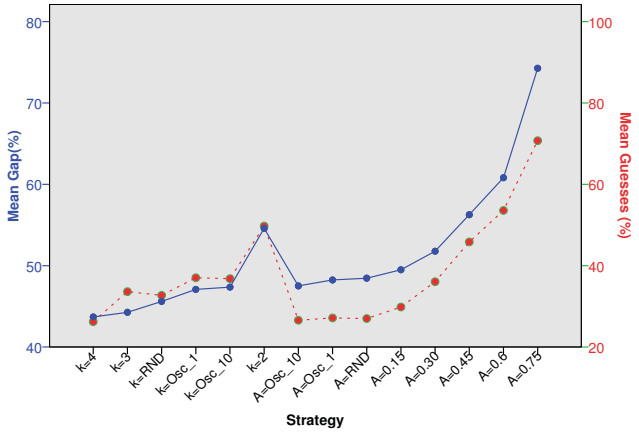


Figure 5: Average Gap (left Y axis) and $Guess$ (right Y axis) for every strategy. The dotted line is used for $Guess$.

Several elements can be observed in the Figure. Let's start with the *R*-α strategy (left). As α increases, we can observe a progressive concentration of the points towards the upper right corner of the plot. This means that the results are getting worse: more guesses, higher gap (agent W is getting much lower payoff than the potential optimum). It is noticeable the high variation the results are showing. For instance, when $\alpha = 0.45$ (second plot, in the left), one can observe simulation with guesses around 30% and gap around $30\% - 45\%$ and others with guesses higher than 60% and gap higher than 60%. When a simple dynamic strategy is used ($\alpha = RND$), then the number of correct guesses is almost always below 30%. However, the range of values that can be obtained for gap value is quite wide, going from 30% to 60%.

The plots are a bit different for the *R-k-B* strategy. As k increases, the percentage of guesses decrease. This is reasonable taking into account how the strategy works: take the k best alternatives and then, choose random. In the experiments performed, the total number of actions is 5, so, when $k = 4$, then just the worst strategy is eliminated. In other words, as k increases, agent B may conclude that W is behaving randomly. In turn, the gap value is more variable, ranging from 35% to 55% approx.

6 Discussion and Future Work

In this work we focused in the context of adversarial reasoning where a player W wants to *force the presence of uncertainty* in order to confuse the adversary B while its payoff is as less affected as possible in situations of repeated conflicting encounters. We extended a previous work where the decision strategies were fixed along the time, to consider dynamic strategies: the way the decisions are taken varies with the time.

When the decision strategy is based on α-cuts, then a dynamic variation of α led to better results than the static counterpart. Moreover, none of the adaptation schemes proposed led to worst results than those obtained by a specific configuration of the control parameter. However, in the $R - k - B$ case, the results are not so clear but still good. The best results are obtained with $k = 3, k = 4$. Using any of these values, the strategy obtained performed better than four of the other

ISBN: 978-989-95079-6-8

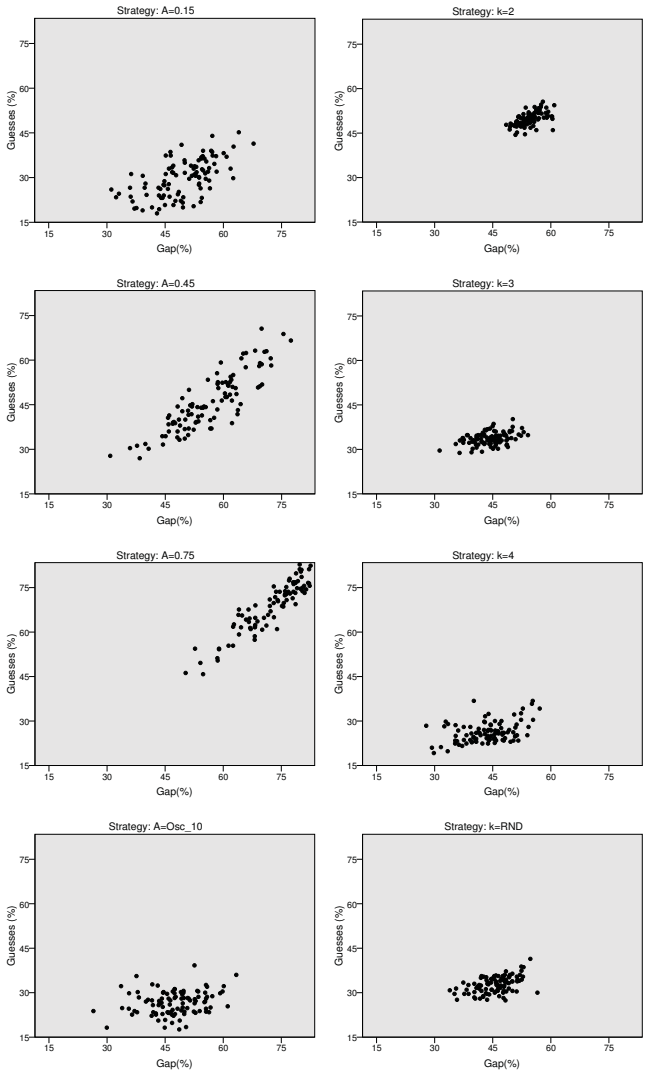


Figure 6: Every point represents the result $(Gap, Guess)$ of a repetition.

strategies available. The best dynamic alternative is RND which assigns a random value to k before applying the strategy. This strategy achieved a similar performance (on average) with respect to $k = 3, k = 4$.

The use of dynamic strategies, besides its potential performance, has another benefit which is to avoid determining a specific value for the control parameter. This is not a trivial matter as there is no simple way to infer which setting will provide the best value for a different simulation scenario.

The study performed also revealed an interesting point: the lowest average Gap value for a dynamic strategy was 45% when using $k = RND$ and this implies that W's payoff is 45% lower than the theoretical optimum one. In [2], we showed that using a purely random decision strategy, the gap is higher than 50%. The question now is: would it be possible to design a strategy with guaranteed (in a formal sense) performance?.

The way the dynamism is included in this proposal is quite simple, but several alternatives are available for producing improvements. Now, agent W is not analyzing the payoff obtained against the one expected (which is stored in matrix P). Agent W can use the fact that its payoff is being affected in order to produce an intelligent adaptation of its strategy. In this context, the use of fuzzy rules may play an important role as it would help to model adaptation strategies as: *if the reduction of payoff is high, then increase uncertainty*, or *if this event occurred a high number of times then increase uncertainty*. Other lines of research are related with the model used. Here, the way the payoff matrix is generated may affect the results obtained by the strategy based on α-cuts as the differences between the suitability of alternatives may be low. Other ways to define such matrix may be necessary. Also, the sequence of events is completely random now, but other options are available.

The relation between the number of correct guesses and the gap value is clear in some situations but in others (as shown in Fig. 6), it is possible to obtain a broad range of gap values for the same number of guesses. In this context, it seems not trivial to use the number of guesses as a predictor for the gap value. A different alternative may take into account that the presence of uncertainty is reflected in the matrix of observations O that agent B construct, so it would be interesting to look for correlations between the gap value and some measures about this matrix O. The ideas posed here are left as future work.

Finally, we would like to mention that one of the reviewers claimed that the problem posed here can be solved with the standard tools of game theory arguing that the problem can be seen as a sequence of n independent non cooperative games. However, we do not think that this is the case as, for example, the matrix of observations changes at each step, we are not dealing with mixed strategies and so on. Moreover, we claim that this kind of analysis can be applied when the events are correlated in some way, and also, when more sophisticated adaptation and learning mechanisms are considered in both agents.

Acknowledgments

This work is supported in part by projects TIN2008-01948 from the Spanish Ministry of Science and Innovation, and P07-TIC-02970 from the Andalusian Government.

References

[1] A. Kott and W. M. McEneany. *Adversarial Reasoning: Computational Approaches to Reading the Opponents Mind*. Chapman and Hall/ CRC Boca Raton, 2007.

[2] D. Pelta and R. Yager. On the conflict between inducing confusion and attaining payoff in adversarial decision making. *Information Science*, 179:33–40, 2009.

[3] P. Thagard. Adversarial problem solving: Modeling an opponent using explanatory coherence. *Cognitive Science*, 16(1):123 – 149, 1992.

[4] E. Kardes and R. Hall. Survey of literature on strategic decision making in the presence of adversaries. Report 05-006, National Center for Risk and Economic Analysis of Terrorism Events, 2005.

[5] A. Kott and M. Ownby. Tools for real-time anticipation of enemy actions in tactical ground operations. In *Proceedings of the 10th International Command and Control Research and Technology Symposium*, 2005.

[6] R. R. Yager. A knowledge-based approach to adversarial decision making. *International Journal of Intelligent Systems*, 23:1–21, 2008.

ISBN: 978-989-95079-6-8