Data Integration, Approximate Categorisation and Fuzzy Associations

Trevor Martin[1,2] Yun Shen[1]

1 Artificial Intelligence Group, Department of Engineering Mathematics,
University of Bristol, University Walk, Bristol BS8 1TR, UK
2 BT Intelligent Systems Lab, BT Innovate, Orion Building
Adastral Park, Ipswich IP5 3RE, UK

Email: firstName.lastName@bristol.ac.uk

Abstract—*The use of hierarchical taxonomies to organise information (or sets of objects) is essential to the semantic web and is also fundamental to many aspects of web 2.0. In most cases, the seemingly crisp granulation of a taxonomy disguises the fact that categories are based on loosely defined concepts which are better modelled by allowing graded membership. Fuzzy categories may also arise when integrating information from multiple sources which do not conform to precisely the same taxonomy definitions. Knowledge of relations between categories can be summarised by association rules. In this paper, we outline a new method to calculate fuzzy confidences for association rules between fuzzy categories from different hierarchies. We illustrate with examples drawn from a system that integrates information from web-based sources.*

Keywords— fuzzy association rules, fuzzy data mining, fuzzy hierarchies, soft semantic web, mass assignment.

1 Introduction

The semantic web [1] is a co-ordinated attempt to develop "*common formats for integration and combination of data drawn from diverse sources*"* whilst Web2.0 is an emergent trend involving changes in web use, including user participation, content sharing, collaboration, etc. Both rely heavily on hierarchical taxonomies as a fundamental mechanism for information organisation. In the case of the semantic web, ontologies provide a formal definition of terms, relations, constraints, etc. as well as their correspondence to objects in the real world. For example, an ontology for wine retailers would define terms such as grape type, vintage, region, descriptions of flavour, etc. Web documents using these terms can be processed by machines, knowing that the "meaning" is defined by the ontology. The use of "tags" from the ontology enables automated reasoning as well as assisting human understanding.

In the case of Web 2.0, tagging and categorisation also play a central role. Photo and video sharing sites, wikis, blogs, and large parts of e-commerce rely on hierarchical categorisation of content. This may be implicit rather than explicit - a tag such as *sport* can encompass *athletics*, *swimming*, *football*, etc - but the hierarchical nature is unarguable. Rather than relying on formal, agreed ontologies, the meaning of a tag is essentially defined by the user who adds the tag, and a degree of shared understanding is necessary for the system to work.

The ability to categorise and summarise data is a key feature of human intelligence, enabling us to group multiple entities together into an (approximately) uniform whole and represent / reason about the group as a single concept. Many familiar ways to access information such as books, libraries, computer file structures, the web and other networks provide evidence that hierarchical categorisation is an efficient way to organise and access information.

There is rarely a single unique hierarchy - instead, a person chooses the most appropriate way to split up data according to their expertise / background knowledge and the problem at hand. For example, Fig 1 shows two possible divisions of customers. Here, as in most real-world hierarchies, groups of entities (or conceptual categories) are loosely defined, able to admit elements according to some scale of membership rather than according to an absolute yes/no test - indicating that fuzzy set theory [2] is appropriate.

An association exists when the extensions of two concepts overlap significantly, as indicated in Fig 1 by the directed link. Association rules (in their crisp form) are a well-established technique for knowledge discovery in databases,

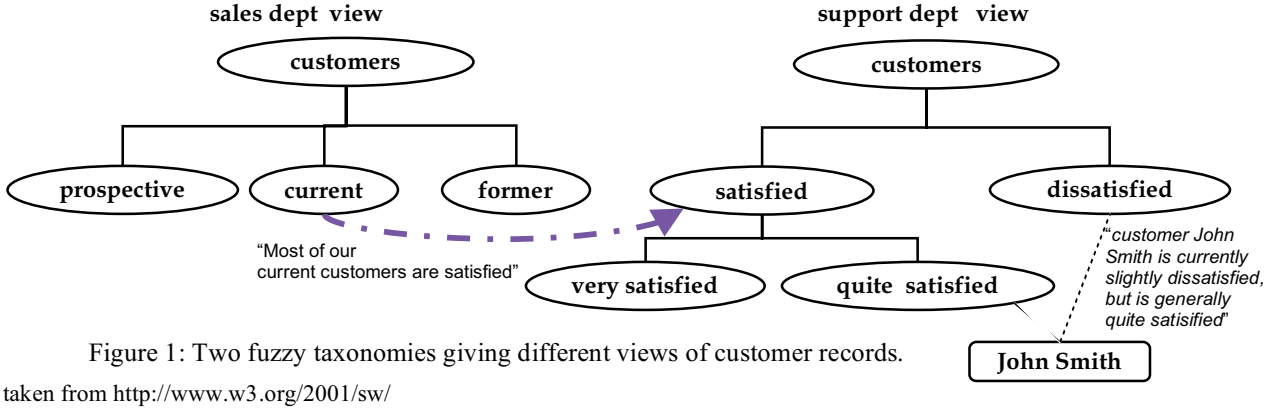

Figure 1: Two fuzzy taxonomies giving different views of customer records.

* taken from http://www.w3.org/2001/sw/

ISBN: 978-989-95079-6-8

enabling "interesting" relations to be discovered. There have been a number of proposals to develop fuzzy association rules, that is to discover the degree of association between fuzzy categories. Some of our recent work has used mass assignment theory [3-5] to find a point valued association strength between fuzzy categories [6], later extended to an interval-valued version [7]. In common with other work on fuzzy association rules, most of our previous work assumes there is a crisp value for the rule confidence.

The main contribution of this paper is a novel mass assignment-based method for calculating a *fuzzy* confidence in associations between fuzzy categories. It relies on a new method of converting fuzzy relations to mass assignments and a membership function for fuzzy confidence related to the movement of mass needed to produce that confidence value. The minimum and maximum values for the confidence can be found quickly, and memberships calculated based on the corresponding mass assignments. A full version, including proofs, will appear subsequently [8].

2 Background

The Smart Queries and Adaptive Data (SQuAD) project is concerned with adding structure (tags) to data and refining approximate knowledge from this data. It aims to assist in the extraction of useful information from diverse sources of semi-structured text data and consists of four main components:

- approximate fuzzy grammars [9] to tag small text fragments.
- SOFT (Structured Object Fusion Toolkit) [10] to help determine that entities from different sources are the same (a process also known as instance-matching, entity resolution)
- iPHI (intelligent Personal Hierarchies of Information) [11, 12] which aims to combine and integrate multiple sources of information and to configure access to the information based on an individual's personal categories.
- TRACK (Time-varying Relations in Approximately Categorised Knowledge). [7] The previous three processes leave us with a collection of objects organised into fuzzy categories, where the taxonomic structure reflects a view of the underlying data. Insight can be gained by considering relations between fuzzy categories, particularly from different category hierarchies. Of particular interest is the change over time in the degree of association between fuzzy categories - for example, in Fig 1, an executive in charge of the company would be interested to know whether the set of current customers (in the sales department's categorisation) is mostly the same as the fuzzy set of satisfied customers (in the support department's categorisation) and how the relation has changed over recent time, over the medium term and over the long term - particularly if there have been significant changes such as company practice, number/nature of competitors, etc during any of these periods.

2.1 Fuzzy Sets - the Conjunctive Interpretation

Many authors (e.g. [13]) have proposed fuzzy sets to model uncertain values in databases and knowledge based applications. The standard (disjunctive) interpretation of a fuzzy set in this context is as a possibility distribution - a single valued attribute which is not known exactly.

The conjunctive interpretation of a fuzzy set occurs when the attribute can have multiple values. These values are not repeated so it is a set, not a multi-set. For example, a person may be able to speak several languages; we could model this as a fuzzy set of languages, where membership would depend on the degree of fluency. This is formally a relation rather than a function on the underlying sets. We distinguish between the conjunctive interpretation - modelled by a monadic fuzzy relation – and the disjunctive interpretation – modelled by a possibility distribution by using the notation

$$F(a) = \{x/\mu(x) \mid x \in U\}$$

to denote a single valued attribute F of some object a (i.e. a possibility distribution over a universe U) and

$$R(a) = [x/\chi(x) \mid x \in U]$$

to denote a multi-valued attribute (relation). Fuzzy categories represent the latter case, since we have multiple values that satisfy the predicate to a greater or lesser degree.

2.2 Extending Association Rules to Fuzzy Categories

In creating association rules within transaction databases (e.g. [14], see also [15] for a clear overview), the standard approach is to consider a table in which columns correspond to items and each row is a transaction. A cell contains 1 if the item was bought, and 0 otherwise. The aim of association rule mining is to find links between disjoint subsets of items – for example, do customers generally buy biscuits and cheese when buying beer and wine? These disjoint subsets represent categories, as described earlier.

Let I denote the set of items, so that any transaction can be represented as $tr \subseteq I$, and consider X, the set of all transactions (strictly speaking, X is a multi-set but can be made into a set by adding a unique identifier to each transaction). We also specify two categories (itemsets) s and t, which are non-empty, non-overlapping subsets of I,

$$t \subset I \qquad s \subset I \quad \text{where} \quad s \cap t = \varnothing$$

and the sets of transactions containing s and t

$$S = \{x | x \in X \wedge s \subseteq x\}$$

$$T = \{x | x \in X \wedge t \subseteq x\}$$

An association rule is of the form $s \Rightarrow t$ and is interpreted as stating that when the items in s appear in a transaction, it is likely that the items in t will also appear i.e. it is not an implication in the formal logical sense. A slight abuse of notation allows us to use $S \Rightarrow T$ or $s \Rightarrow t$ as the rule.

Most authors use two measures to assess the significance of association rules. The support of a rule $s \Rightarrow t$ is the number (or relative number) of transactions in which both s and t appear, and the confidence of the rule is an estimate (based on the samples) of the conditional probability of t being contained in a transaction given that it contains s

$$Support(s,t) = |S \cap T| \qquad (1)$$

$$Conf(s,t) = \frac{|S \in T|}{|S|} \qquad (2)$$

Typically a threshold is chosen for the support, so that only frequently occurring sets of items s and t are considered; a second threshold filters out rules of low confidence.

ISBN: 978-989-95079-6-8

name	sales	salary
a	100	1000
b	80	400
c	50	800
d	20	700

good sales	*high salary*	*confidence*
sales≥80	*high≥400*	1
sales≥50	*high≥500*	0.667
sales>50	*high>500*	0.5
sales≥50	*high>800*	0.333

Figure 2 : A simple database of names (*a, b, c, d*), sales and salaries (top) and (bottom) confidences for the rule *good sales => high salary* arising from different crisp definitions of the terms *good sales* and *high salary*

For example, consider a database of sales employees, salaries and sales figures. A mining task might be to find out whether the good sales figures are achieved by the highly paid employees. Given the database table in Fig 2, we can obtain rule confidences ranging from 1/3 up to 1 by different crisp definitions of "*good sales*" and "*high salary*".. Although this is a contrived example, such sensitivity to the cut-off points adopted for crisp definitions is a good indication that a fuzzy approach is more in line with human understanding of the categories.

Various approaches to fuzzifying association rules have been proposed e.g. [15-17]. The standard extension to the fuzzy case is to treat the sets *S, T* as fuzzy and find the intersection and cardinality using a t-norm and sigma-count respectively.

$$Conf(S,T) = \frac{\in \mu_{S \cap T}(x)}{\in \mu_S(x)} \qquad (3)$$

In the example of Fig 2, a fuzzy approach would categorise employees according to simple membership functions for *good sales* (S) and *high salary* (T), which could lead to

$S = [a/1, b/0.8, c/0.5, d/0.2]$

and

$T = [a/1, b/0.4, c/0.8, d/0.7]$

and confidence 0.72 for the association $S=>T$ using (3).

As pointed out by [15], using min and the sigma count for cardinality can be unsatisfactory because it does not distinguish between several tuples with low memberships and few tuples with high memberships - for example,

$S = [x_1/1]$

$T = [x_2/1]$

leads to *Conf(S, T) = 0* but

$S_1 = [x_1/1, x_2/0.01, x_3/0.01, \ldots, x_{1000}/0.01]$

$T_1 = [x_1/0.01, x_2/1, x_3/0.01, \ldots, x_{1000}/0.01]$

leads to

$$Conf(S_1,T_1) = \frac{1000 \in 0.01}{1 + 999 \in 0.01} \in 0.91$$

which is extremely high for two almost disjoint sets (this example originally appeared in [18]). Using a fuzzy cardinality (i.e. a fuzzy set over the possible cardinality values) is also potentially problematic since the result is a possibility distribution over rational numbers, and the extension principle [19] gives a wider bound than it should, due to neglect of interactions between the numerator and denominator in Eq 2. For example, given

$S = [x1/1, x2/0.8]$

$T = [x1/1, x2/0.4].$

the fuzzy cardinalities are

$|S \cap T| = \{1/1, 2/0.4\}$,

$|S| = \{1/1, 2/0.8\}$

leading (by the extension principle) to a confidence of

$\{0.5/0.8, 1/1, 2/0.4\}$,

clearly incorrect as the confidence cannot be greater than 1.

2.3 Fuzzy Relations as Mass Assignments

As with previous work [6, 7] we start from the fact that a relation represents a conjunctive set of ordered *n*-tuples i.e. a conjunction of *n* ground clauses. For example, if *U* is the set of dice values then we could define a (crisp) predicate *differBy4or5* on $U \times U$ as the set of pairs

[(1,6), (1,5), (2,6), (5,1), (6,1), (6, 2)]

This is a conjunctive set, in that each pair satisfies the predicate. In a similar way, a fuzzy relation represents a set of *n*-tuples that satisfy a predicate to some degree. Thus *differByLargeAmount* could be represented by

[(1,6)/1, (1,5)/0.6, (2,6)/0.6, (5,1)/0.6, (6,1)/1, (6,2)/0.6]

The interpretation is not that a single pair satisfies this predicate, but that one set of pairs satisfies it (out of several possible sets of pairs).

A mass assignment [3-5] on a universe *U* is a distribution over the power set of *U*. Here, the mass assignment is on possible *relations* :

$R_1 = [(1,6),(6,1)]$

$R_2 = [(1,6),(1,5),(2,6),(5,1),(6,1),(6,2)]$

$m = \{\{R_1\} : 0.4, \{R_1, R_2\} : 0.6\}$

This is equivalent to treating the fuzzy relation as a fuzzy set of crisp relations:

$differByLargeAmount = \{R1/1, R2/0.6\}$

Similarly, a monadic fuzzy predicate *largeValue* defines a set of 1-tuples such as [6/1, 5/0.8, 4/0.3] which is written as a fuzzy set of crisp monadic relations:

$largeValue = \{[6]/1, [6,5]/0.8, [6,5,4]/0.3\}$

and has the mass assignment

$m_{largeValue} = \{\{[6]\} : 0.2, \{[6],[6,5]\} : 0.5, \{[6],[6,5],[6,5,4]\} : 0.3\}$

Our subsequent studies [7] show that this approach can sometimes overestimate the difference between full and nearly-full membership, which can lead to unreasonably large intervals calculated for the confidence of association rules. For example, under this interpretation, the monadic fuzzy relation $S = [a/1\ b/0.98]$ has the mass assignment

$m_S = \{\{[a]\} : 0.02, \{[a],[a,b]\} : 0.98\}$

ISBN: 978-989-95079-6-8

The normal mass assignment interpretation allows us to redistribute the mass on {*[a], [a,b]*} to either of the relations *[a]* or *[a,b]* which leads to the family of distributions:

$S = [a] : 1\text{-}x \quad , \quad [a,b] : x \quad$ where $0 \le x \le 0.98$

This flexibility in re-assigning mass means that for a source

$S = [a/1\ b/0.98]$ and target relation $T = [a/1\ b/0.98\ c/0.02]$

we get an interval [0.51, 1] which is surprisingly wide considering the two relations are so similar. NB this behaviour arises mostly in contrived cases, smaller intervals are calculated in the vast majority of "real" association rules that have arisen in our experimental studies such as [7].

2.4 Alternative Interpretation of Relations as Mass Assignments

The approach discussed above, which we will refer to as an open world approach, treats partial membership of a tuple x in a relation R (i.e. $0 < \cap_R(x) < 1$) as an upper bound for the mass that can be assigned to any set of tuples including x. This leads to a wide range of mass distributions that can be derived from the fuzzy relation R.

In the open world approach, for any tuple x where $\cap_R(x)<1$, the total mass that can be assigned to relations containing x is given by

$$0 \le \sum_{\substack{t=[x1,\ldots,xn] \\ x \in t}} m(t) \le \chi_R(x)$$

In the *largeValue* example above, consider the element $x=5$, which has $\cap_R(x) = 0.8$; the relations containing 5 are [5, 6] and [4, 5, 6] and we have

$$0 \le m_{largeValue}([5,6]) + m_{largeValue}([4,5,6]) \le 0.8$$

This gives a considerable degree of flexibility in assigning mass. The alternative interpretation used here - the closed world approach - regards partial membership of a tuple x in a relation R (i.e. $0 < \cap_R(x) < 1$) as *strictly equal* to the total mass assigned to the sets of tuples which include x, i.e.

$$\sum_{\substack{t=[x1,\ldots,xn] \\ x \in t}} m(t) = \cap_R(x)$$

This means there is no flexibility in the range of mass distributions that can be derived from the fuzzy relation R. However, there is flexibility in the mass assignments when R is combined with an assignment corresponding to another relation, for example in calculating association confidences as described later.

Under this interpretation, the monadic fuzzy relation *largeValue* discussed above has

$$m_{largeValue} = \{[6]:0.2,\ [6,5]:0.5,\ [6,5,4]:0.3\}$$

and clearly

$$m_{largeValue}([5,6]) + m_{largeValue}([4,5,6]) = 0.8$$

2.5 Closed world Mass-based Association Rules

For a source category

$$S = \left[x_1/\chi_S(x_1),\ x_2/\chi_S(x_2),\ldots,\ x_{|S|}/\chi_S(x_{|S|})\right]$$

and a target category

$$T = \left[x_1/\chi_T(x_1),\ x_2/\chi_T(x_2),\ldots,\ x_{|T|}/\chi_T(x_{|T|})\right]$$

we can define the corresponding mass assignments as follows. Let the set of distinct memberships in S be

$$\Lambda_S = \left\{\chi_S^{(1)}, \chi_S^{(2)}, \ldots, \chi_S^{(n_S)}\right\}$$

where

$$\chi_S^{(1)} > \chi_S^{(2)} > \ldots > \chi_S^{(n_S)}$$

and $n_S \le |S|$

Let

$$S_i = \left\{\left[x \,\middle|\, \chi_S(x) \ge \chi_S^{(i)}\right]\right\}$$

Then the mass assignment corresponding to S is

$$\left\{S_i : m_S(S_i)\right\}, \quad 1 \le i \le n_S$$

where

$$m_S(S_k) = \chi_S^{(k)} - \chi_S^{(k+1)} \qquad (7)$$

and we define

$$\chi_S^{(i)} = 0 \quad if \quad i > n_S$$

For example, the fuzzy category

$S = [a/1,\ b/0.8,\ c/0.5,\ d/0.2]$

has the corresponding mass assignment

$$M_S = \{[a]:0.2,\quad [a,b]:0.3,\quad [a,b,c]:0.3,\quad [a,b,c,d]:0.2\}$$

We can now calculate the confidence in the association between the categories S and T using mass assignment theory. In general, this is not a unique value as we are free to move mass (consistently) between the cells corresponding to S_i and T_j for each i, j.

For two mass assignments

$$M_S = \left\{S_i : m_S(S_i)\right\}, \quad 1 \le i \le n_S$$

$$M_T = \left\{T_j : m_T(T_j)\right\}, \quad 1 \le j \le n_T$$

the composite mass assignment is

$$M = M_S \oplus M_T = \left\{X : m(X)\right\}$$

where m is specified by the composite mass allocation function, subject to

$$\sum_{j=1}^{n_T} m_{ij} = m_S(S_i) \qquad \sum_{i=1}^{n_S} m_{ij} = m_T(T_j)$$

This can be visualised using a mass tableau (see [3]) as shown in Fig 3. Each row (column) represents a relation of the source (target) mass assignment. We label the rows $S_1, S_2, \ldots S_{nS}$ and columns $T_1, T_2, \ldots T_{nT}$, and assign mass m_{ij} to cell *(i, j)* subject to row and column constraints. The confidence in the association rule is given by

$$conf(M) = \frac{\sum_{i,j}\left(m_{ij} \times |S_i \cap T_j|\right)}{\sum_{i=1}^{n_S}\left(\sum_{j}^{n_T} m_{ij} \times |S_i|\right)} = \frac{n}{d}$$

$$where \quad n = \sum_{i,j}\left(m_{ij} \times |S_i \cap T_j|\right)$$

$$d = \sum_{i=1}^{n_S}\left(\sum_{j}^{n_T} m_{ij} \times |S_i|\right) = \sum_{i=1}^{n_S} |S_i| \times m_S^{(i)} \qquad (7)$$

ISBN: 978-989-95079-6-8

Clearly $n \geq 0$, $d > 0$ and d is a constant for a given source relation S, irrespective of T and M.
For example consider the fuzzy categories

$S = [a/1, b/0.8, c/0.5, d/0.2]$ and

$T = [a/1, b/0.4, c/0.8, d/0.7]$

One notable assignment is the least prejudiced distribution, obtained by taking the product of source and target masses for each cell as shown in Fig 3. This corresponds to the minimum entropy combination of the two mass assignments.

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0.04	[a] 0.02	[a] 0.06	[a] 0.08
0.3	[ab]	[a] 0.06	[a] 0.03	[a] 0.09	[ab] 0.12
0.3	[abc]	[a] 0.06	[ac] 0.03	[ac] 0.09	[abc] 0.12
0.2	[abcd]	[a] 0.04	[ac] 0.02	[acd] 0.06	[abcd] 0.08

Figure 3: mass tableau, showing intersections $S_i \in T_j$ and the least prejudiced mass distribution. The corresponding point valued rule confidence is 1.86 / 2.5 = 0.744

It is possible to show [8] that maximum rule confidence is obtained by moving mass towards the diagonal (top left to bottom right) and that minimum confidence is obtained by moving mass towards the bottom left - top right diagonal, as illustrated in figs 4 and 5.

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0.2	[a] 0	[a] 0	[a] 0
0.3	[ab]	[a] 0	[a] 0.1	[a] 0.2	[ab] 0
0.3	[abc]	[a] 0	[ac] 0	[ac] 0.1	[abc] 0.2
0.2	[abcd]	[a] 0	[ac] 0	[acd] 0	[abcd] 0.2

Figure 4 : mass tableau, showing intersections $S_i \in T_j$ and the mass distribution leading to minimum rule confidence 2.1 / 2.5 = 0.84

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0	[a] 0	[a] 0	[a] 0.2
0.3	[ab]	[a] 0	[a] 0.1	[a] 0.1	[ab] 0.2
0.3	[abc]	[a] 0	[ac] 0.1	[ac] 0.2	[abc] 0
0.2	[abcd]	[a] 0.2	[ac] 0	[acd] 0	[abcd] 0

Figure 5 : mass tableau, showing intersections $S_i \in T_j$ and the mass distribution leading to minimum rule confidence 1.5 / 2.5 = 0.6

2.6 Membership Function for Fuzzy Confidence

We define the membership function in terms of the mass which must be moved (relative to the least prejudiced distribution, where confidence has membership 1). Any other assignment of mass requires one or more elementary mass transfers relative to the LPD, and we are particularly interested in mass assignments corresponding to minimum and maximum confidence, M^{MIN} and M^{MAX}. We define a fuzzy interval C representing the confidence such that

$$\mu_C(conf(M)) = 1 - \frac{pos(M^{LPD} - M)}{N}$$

where $$N = \max\left(pos(M^{\max} - M^{LPD}), pos(M^{\min} - M^{LPD})\right)$$

Because the membership varies linearly with the amount of mass moved, it is triangular and can be calculated quickly by considering the end points. We note that it is possible for the membership function to be discontinuous at one end (i.e. to drop abruptly to zero). Fig 6 shows the membership function for the fuzzy confidence in the *good sales - high salary* example, with the calculations shown in figs 3, 4, and 5.

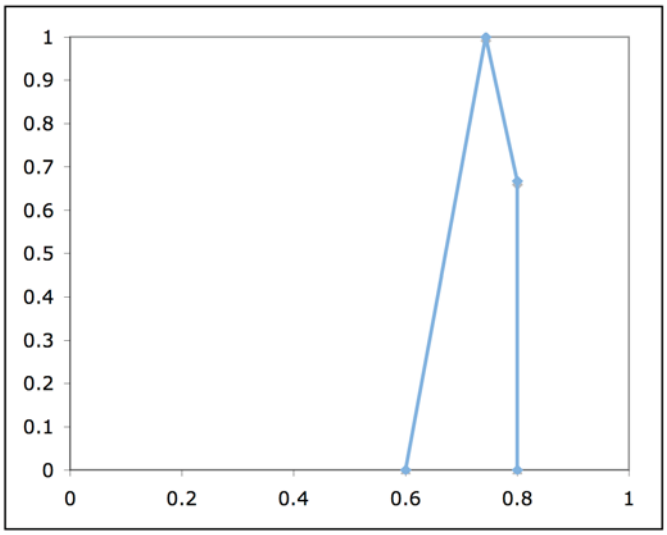

Figure 6 : Membership function for fuzzy confidence in the good sales - high salary example. Note the discontinuity at 0.8

3 Demonstrator Application

We have applied the algorithm to calculation of associations in an integrated taxonomic database of terrorist incidents, as described in [7]. Our primary source is the Worldwide Incidents Tracking System (WITS) [20] augmented by additional information from the MIPT Terrorism Knowledge Base (TKB) database[†]. WITS consists of incidents in which *"subnational or clandestine groups or individuals deliberately or recklessly attacked civilians or noncombatants (including military personnel and assets outside war zones and war-like settings)"*. We are not concerned with the correctness or otherwise of this definition, and treat the data simply as a testbed for SQuAD .

Once data has been integrated and categorised into the iPHI hierarchies, the methods outlined in this paper and [20]

[†] TKB ceased operation on 31 March 2008 and is now part of Global Terrorism Database (GTD) at www.start.umd.edu/data/gtd/

ISBN: 978-989-95079-6-8

have been applied to find significant associations between categories in different hierarchies (e.g. *geographic region*, *weapon type, perpetrator, casualty level* etc). Since the data is not static, we cannot assume that significant associations remain significant – indeed, valuable insight arises from changes in association levels relative to other associations, and trends in the strength of an association. 15,900 incidents between January 2005 and January 2006 were analysed.

Various associations can be extracted by consideration of fuzzy categories in different taxonomies. Although the vast majority of results from [7] led to reasonable intervals, there were a few cases in which intervals were quite large. As can be seen from Fig 7, much smaller intervals were obtained under the closed world method. The interval for closed world rules indicates the extremes of the fuzzy association confidence, with a symmetric triangular membership function (possibly truncated on one side). The interval for open world calculations has membership 1 throughout.

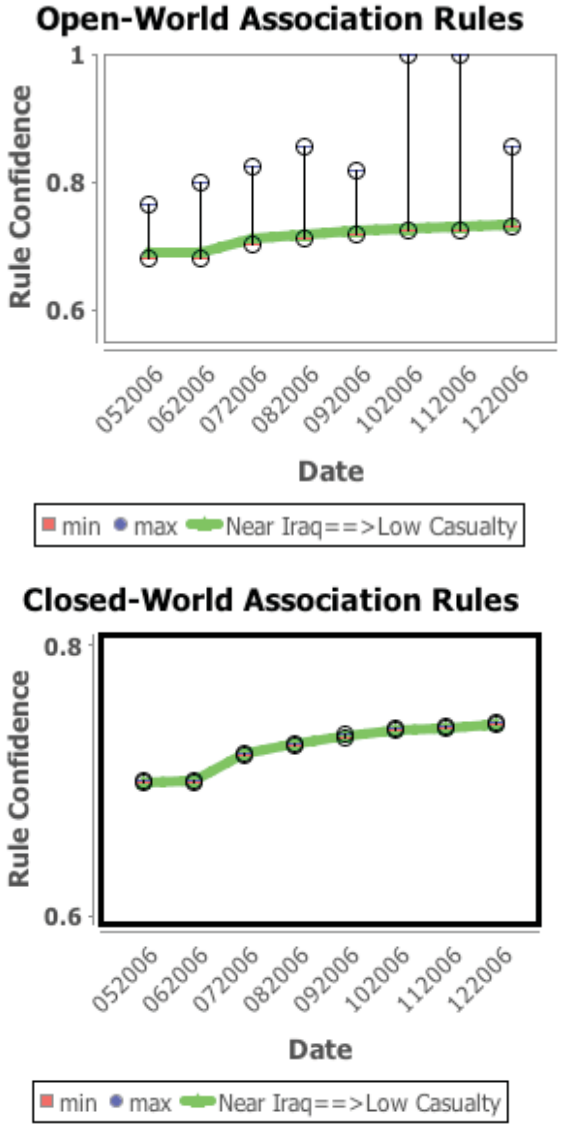

Figure 7 : Confidence intervals and point values for associations between the fuzzy categories *in/near Iraq* and *Medium casualty levels*. The fuzzy confidence (closed world) gives a much tighter interval than the open world calculation

4 Summary and Future Work

We have described a new method for calculating the confidence in an association rule between two (or more) fuzzy categories, based on mass assignment theory and yielding a fuzzy interval within which the confidence must lie. Combined with the point value confidence described in previous work, this enables us to order and plot association strengths with an indication of the degree of uncertainty.

Acknowledgment

This work was partly funded by BT and the Defence Technology Centre for Data and Information Fusion.

References

[1] Berners-Lee, T., J. Hendler, and O. Lassila, "The Semantic Web," in *Scientific American*, vol. 284, 2001, pp. 28-37.

[2] Zadeh, L. A., "Fuzzy Sets," *Inf. and Control*, 8, pp. 338-353, 1965.

[3] Baldwin, J. F., "Management of Fuzzy and Probabilistic Uncertainties for KB Systems," in *Encyclopedia of AI*, S. A. Shapiro, Ed., 2nd ed: John Wiley, 1992, pp. 528-537.

[4] Baldwin, J. F., "Mass Assignments and Fuzzy Sets for Fuzzy Databases," in *Adv. in the Shafer Dempster Theory of Evidence*, M. Fedrizzi, J. Kacprzyk, and R. R. Yager, Eds.: Wiley, 1994.

[5] Baldwin, J. F., T. P. Martin, and B. W. Pilsworth, *FRIL - Fuzzy and Evidential Reasoning in AI*. Wiley, 1995.

[6] Martin, T. P., B. Azvine, and Y. Shen, "A Mass Assignment Approach to Granular Association Rules for Multiple Taxonomies," presented at Uncertain Reasoning in the Semantic Web / Semantic Web 2007, 2007.

[7] Martin, T. P. and Y. Shen, "TRACK - Time-varying Relations in Approximately Categorised Knowledge," *Int. J. Comp Intelligence Research*, (to appear), 2009.

[8] Martin, T. P. and Y. Shen, "Fuzzy Association Rules to Summarise Multiple Taxonomies," in *Fuzziness and Scalability*, A. Laurent and Lesot. M-J, Eds.: 2009.

[9] Martin, T. P., Y. Shen, and B. Azvine, "Incremental Evolution of Fuzzy Grammar Fragments to Enhance Instance Matching and Text Mining," *IEEE Transactions on Fuzzy Systems*, 2008.

[10] Martin, T. P. and B. Azvine, "Soft Integration of Information with Semantic Gaps," in *Fuzzy Logic and the Semantic Web*, E. Sanchez, Ed.: Elsevier, 2005.

[11] Martin, T. P. and B. Azvine, "Acquisition of Soft Taxonomies for Intelligent Personal Hierarchies and the Soft Semantic Web," *BT Tech Jl*, 21, pp. 113-122, 2003.

[12] Martin, T. P., B. Azvine, and Y. Shen, "Intelligent Hierarchy Mapping: A Soft Computing Approach," *Information Tech and Intelligent Computing*, 2008.

[13] Bosc, P. and B. Bouchon-Meunier, "Databases and Fuzziness - Introduction," *Int J Intel Systems*, 9, 419, 1994.

[14] Agrawal, R. and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," VLDB, 1994.

[15] Dubois, D., E. Hullermeier, and H. Prade, "A systematic approach to the assessment of fuzzy association rules," *Data Mining and Knowl Discovery*, 13, 167-192, 2006.

[16] Bosc, P. and O. Pivert, "On Some Fuzzy Extensions of Association Rules," proc IFSA , 2001.

[17] Kacprzyk, J. and S.Zadrozny, "Linguistic Summarization of Data Sets Using Association Rules," proc Fuzz-IEEE, 2003.

[18] Martin-Bautista, M.J., M.A.Vila, H.L. Larsen, D. Sanchez, "Measuring Effectiveness in Fuzzy Information Retrieval," Flexible Query Answering Systems (FQAS), 2000.

[19] Zadeh, L. A., "The Concept of a Linguistic Variable and its Application to Approximate Reasoning (Part 1)," *Information Sciences*, vol. 8, pp. 199-249, 1975.

[20] WITS, "WITS - Worldwide Incidents Tracking System," National Counterterrorism Center Office of the Director of National Intelligence 2007.

ISBN: 978-989-95079-6-8