Optimising the Fuzzy Granulation of Attribute Domains

Marcos E. Cintra[1] Heloisa A. Camargo[2] Trevor Martin[3]
1.Mathematics and Computer Science Institute, São Paulo University(USP)
P.O. Box 668, 13561-970, São Carlos-SP, Brazil
2.Computer Science Department, Federal University of São Carlos(UFSCar)
P.O. Box 676, 13565-905 São Carlos-SP
3.Mathematics Engeneering Department, University of Bristol
Queen's Building, University Walk, Bristol - UK - BS8 1TR
Email: cintra@icmc.usp.br, heloisa@dc.ufscar.br, Trevor.Martin@bristol.ac.uk

Abstract— *The definitions of the number of fuzzy sets and their proper distribution on their domains are fundamental issues for fuzzy systems since these basic parameters deeply affect the quality of the systems results, both in terms of performance rates and interpretability. Several methods have been proposed in the literature to define these parameters, although it is common to find works in which the number of fuzzy sets is defined empirically, distributing them equally in the domains. This paper presents a fast and easy method to estimate the number of fuzzy sets for each attribute and compares three methods for the distribution of fuzzy sets. Two of them are non-supervised methods, using same width and same frequency, the third one is an adaptation of the 1-R supervised method to discretize attributes. Experiments with 10 datasets for classification problems, 10-fold cross validation, using the Wang & Mendel method, and the classic and general fuzzy reasoning methods are presented and discussed.*

Keywords— Fuzzy systems, granulation of attributes, Wang & Mendel method, fuzzy data base generation.

1 Introduction

Fuzzy set theory and fuzzy logic [1] proposed by professor Loft A. Zadeh are the base for the fuzzy systems. Fuzzy set theory is used to represent and process information. The main characteristic of fuzzy sets, contrasting with crisp sets, is the progressive transition from one set to another. This natural characteristic of the fuzzy sets provides automatic mechanisms to deal with imprecision and uncertainty, which are inherent to real world knowledge. Moreover, the fuzzy logic theory prevents the creation of unnatural frontiers in the partitioning of attributes domains [2].

A fundamental issue, due to its direct impact on fuzzy systems, is the definition of the fuzzy sets that model the linguistic variables of a given domain, both in terms of shape (triangular, trapezoidal, S-function, etc) and partitioning of the attributes domains (number of fuzzy sets and their distribution). Unfortunately, there are no general rules or guidelines for these tasks that suit every domain [3]. In fact, many methods may have to be tested in order to find the appropriate definitions for a given application.

It is also common to find studies in which the definition of the membership functions is done empirically. In fact, the Gaussian, trapezoidal and triangular shapes are the most used in the literature, probably because they produce comparable results to other shapes and are easily interpretable [4]. Regarding the definition of the number of fuzzy sets and their distribution on the partitions, most of the papers in the literature use from 2 to 10 fuzzy sets for each attribute, equally distributed in their domains, *i.e.*, all fuzzy sets have same width.

Another aspect to be considered when defining fuzzy sets is personal interpretations, which make this task non trivial. For instance, let us consider a linguistic variable *temperature*; the interpretation of the linguist term *low temperature* is closely related to the region where one lives (consider a continental country such as Brazil, for example). Thus, personal interpretations may generate strong variations for the same concept: since each person may have different and personal interpretations on the meaning of a concept, it is natural that different membership functions may be created to define the same concept. This flexibility on the subjective selection of membership functions and, as previously stated, in the distribution of the fuzzy sets in their partitions shows the robustness of the fuzzy logic, which is closely related to the inherent characteristics of fuzzy sets [5].

Although this flexibility exists, the task of defining membership functions is largely ignored. The choice of the method, in fact, depends on the particular application and domain [6]. While the shape of the fuzzy sets may not present expressive differences in the results of fuzzy systems, the number of fuzzy sets and their distribution are relevant parameters that affect the system in terms of performance and in terms of interpretability. Thus, this paper presents a fast and easy method to estimate the number of fuzzy sets, in order to use this estimated number with more costly approaches, such as genetic algorithms. This paper also presents experiments carried out using three distinct methods for the definition of the distribution of fuzzy sets, in an attempt to provide further insight on the task, focusing specifically on classification problems. One of these methods uses same width distribution for the fuzzy sets, the second uses same frequency distribution, and the third takes the classes into consideration in the process.

This paper is organized as follows: Section 2 reviews methods for the automatic definition of fuzzy membership functions; Section 3 describes the three methods used in the experiments for the definition of fuzzy sets. Section 4 describes the heuristic method for the definition of the number of fuzzy sets proposed here. Section 5 presents the experiments and results, followed by the conclusions in Section 6.

ISBN: 978-989-95079-6-8

2 Methods for the Automatic Definition of Fuzzy Sets

There is a wide variety of methods applied for the definition of fuzzy sets, ranging from heuristic methods, genetic algorithms, artificial neural networks, clustering algorithms, the use of indexes, to the adaptation of classic machine learning methods, such as the K-NN algorithm. Next, we present some of these methods. More details on these and on other related methods can be found in [7] and on the provided references.

Several methods explore the use of special parameters for the distribution of fuzzy sets. In [2], for example, the authors use the concept of 2-tuple to define linguistic terms. The 2-tuple strategy is used to define both, the center of a fuzzy set (support), and a value for a symbolic translation, which enables the fuzzy sets to be moved sideways, always keeping the same base width. This symbolic translation of a linguistic term is a number within the interval $[-0.5, 0.5)$. A genetic algorithm is then used to define the best position for the fuzzy sets. The experiments used a fixed number of triangular shaped sets.

In [6] and [8] the authors use the Fuzzy C-Means algorithm for the definition of fuzzy membership functions. In the first method [6], a modified version of the Xie-Beni index [9] is used to define the number of clusters and, therefore, the number of fuzzy sets, for the Fuzzy C-Means algorithm. Once the number of fuzzy sets is defined, the centers of the sets are initialized based on two methods, the equalized universe method, which divides the partition equally by the number of sets, and the subtractive clustering method, which classifies the points present in the training data by their number of neighboring points (thus, a data point will have a high potential value if it has many neighboring points). After the initialization of the centers, the Fuzzy C-Means algorithm is employed to search for the best positions for centers. A hybrid learning algorithm for refining the system parameters based on the ANFIS method [10] is then presented. The second method [8] proposes the use of a Fuzzy C-Means variant for the generation of fuzzy term sets with ½ overlap. This method uses the mean squared error criterion to determine the number of fuzzy sets and the optimal shape of the membership functions associated with each term.

Genetic algorithms are used to tune the parameters of fuzzy sets in [11]. The genetic algorithm is used to find membership functions suitable for mining problems and then uses the final best set of membership functions to mine fuzzy association rules. The fitness of each chromosome is evaluated by the number of large 1-itemsets generated and by the suitability of the membership functions.

The use of artificial neural networks has also been explored for the definition of fuzzy membership functions. The authors in [12] and [13] present an overview of several neuro-fuzzy methods for the definition of fuzzy sets. These methods, although some present good results, have the problem of tunning parameters for the neural networks, and, in some cases, might need further tests.

There are also works focusing on the use of special indexes to generate membership functions. For instance, the concept of entropy is used in [14] and in [15] to determine membership functions. In [16] the authors propose a method for the definition of fuzzy sets based on two indexes, the fuzzy entropy and the fuzziness index, adopting the S shaped fuzzy set. These papers focus on image recognition applications. The Kappa measure was also used. In [17] the authors use the Kappa measure to calculate the *fuzzy Kappa* index in order to define a suitable segmentation of partitions.

Several other approaches can be found in the literature, such as heuristic methods [18], histograms [3], and even a fuzzy version of the classic machine learning K- nearest neighbor algorithm [19].

Although such a variety of methods exist, it is important to emphasize the fact that many studies define the number of fuzzy sets empirically, and employ the equalized universe method [6] to define the distribution of the fuzzy sets, which simply distributes the sets evenly in the partitions. Regarding the number of fuzzy sets, most researchers define this variable empirically using a range usually varying from 2 to 10 fuzzy sets per attribute. The main reasons why researchers decide to define such important variables without the aid of any formal method may include:

- The complexity of the methods available, which may require more time and effort to be implemented than the actual application in focus;
- The flexibility of the fuzzy logic for the definition of fuzzy variables and fuzzy terms, which allows the users to define their own fuzzy data base, since it will be the base for the generation of the fuzzy rules, which in turn will be adjusted to provide a suitable performance;
- The lack of consensus and/or guidelines on which of the available methods is the best for a given application;

The next section presents the methods used in the experiments conducted for this paper.

3 Evaluated Methods for Fuzzy Sets Distribution

The three methods tested and compared in this paper are the equalized universe method, same frequency method, and an adaptation of the 1-R method for attribute discretization. These methods are described next.

3.1 Equalized Universe Method

This method uses the same width for each fuzzy set [6]. Figure 1 presents an attribute described by five triangular fuzzy sets using the equalized universe method. Notice that, usually, the fuzzy sets at the extremes have half the width of the other fuzzy sets in the middle of the domain.

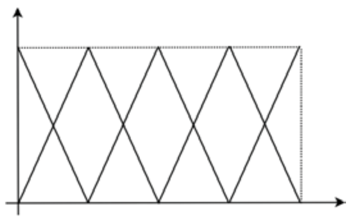

Figure 1: Attribute defined by 5 fuzzy sets using the equalized universe method.

This method is the most widely found in the literature. Some of the reasons for that may include the fact that it is

ISBN: 978-989-95079-6-8

a simple and easy method to use, and that it generates isosceles triangles, with ½ overlap, so no area of the domain will have a membership degree inferior to 0.5 [20]. Besides, they are easy to interpret.

3.2 Same Frequency Method

This method divides the domains of the attributes according to the frequency distribution of the examples. For our implementation, percentiles [21] were used to find the correct distribution for the sets. This way, if an attribute is described by 5 triangular fuzzy sets, the maximum membership degree for the first fuzzy set will be on the 16.7th percentile, on the 33.3th percentile for the second fuzzy set, on the 50.0th percentile for the third one, on the 66.7th percentile for the fourth one, and on the 83.3th percentile for the last fuzzy set. The triangles will not have the same base width, on the contrary, depending on the distribution of the examples, they are likely to be different. Notice that in the case of many examples with the same value, two or more fuzzy sets might be defined with exactly the same middle point. In order to avoid this problem, if the same example is chosen by two or more different percentile values, the neighboring examples of the left and right are chosen to replace the original value.

Figure 2 shows an attribute represented by 5 fuzzy sets distributed using the same frequency method, having 20 examples in the interval $[0, 10]$ (0, 0, 0, 1, 1, 2, 2, 2, 3, 5, 5, 6, 7, 7, 7, 9, 9, 9, 9, 9).

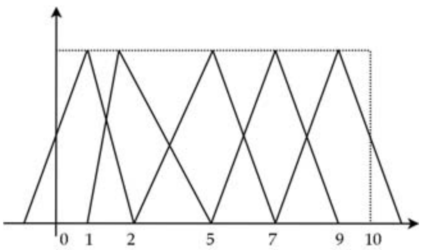

Figure 2: Attribute defined by 5 fuzzy sets using the same frequency method.

Notice that the fuzzy sets of the laterals are defined beyond the actual domain values. This strategy was used to avoid leaving any spaces of the domain with maximum membership degrees close to zero. The space between the actual beginning and end of the domain and the vertices of the triangles is calculated as the same space between the middle vertex of the triangle and the adjacent extremity of the domain.

3.3 1-R Supervised Method

The 1-R method is used for discretization of attributes. It basically divides the domain of an attribute in n intervals, in relation to the frequency of the examples, calculates the optimum class for each interval, and tries to find the best split point using the class. The original algorithm can be found in [22]. Algorithm 1 presents the adapted version of the 1-R algorithm used in the experiments.

Input: n = number of examples;
m = number of attributes;
i = number of fuzzy sets describing the attribute;
for a = *1* **to** m **do**
Sort the examples by attribute a;
Divide the examples in i intervals;
for b = *1* **to** $i-1$ **do**
Calculate the optimal(majority) class for the interval;
for c = *1* **to** *interval* $i-2$ **do**
for d = *index of first example of interval* $c+1$ **to** *index of last example of interval* $c+1$ **do**
if *first example of interval* $c+1$ *belongs to optimal class of interval* c **then** ;
Adjust intervals so that the first example of interval $c+1$ is transfered to interval c;
else d = index of last example of interval $c+1$;
if d = *index of last example of interval* $c+1$
then Split the examples of interval c between intervals c and $c+1$;
end
end
end
For each interval, create a fuzzy set with core in the middle of the interval and laterals in the cores of the neighboring intervals;
end

Algorithm 1: 1-R adapted algoritghm.

The same strategy used for the same frequency method, to avoid leaving any spaces of the domain with maximum membership degrees close to zero, was used similarly for the 1-R adapted method.

Figure 3 shows an attribute represented by 5 fuzzy sets distributed by the same frequency method, having the following 20 examples (attribute value, class) in the interval $[0, 10]$: 0, 3; 0, 3; 0, 1; 1, 3; 1, 3; 2, 3; 2, 2; 2, 2; 3, 2; 5, 2; 5, 1; 6, 1; 7, 1; 7, 3; 7, 3; 9, 2; 9, 3; 9, 1; 9, 2; 9, 3.

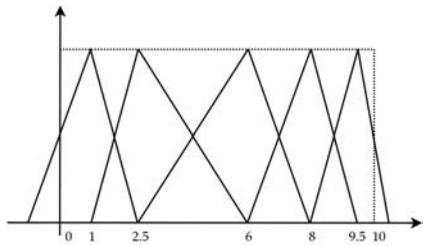

Figure 3: Attribute defined by 5 fuzzy sets using the 1-R adapted method.

4 Heuristic Method to Define the Number of Fuzzy Sets

The approaches used by the recently proposed methods to generate fuzzy rules usually demand a great computational effort, such as the genetic algorithms, that can take from minutes to days to converge to good solutions. Since the definition of the number of fuzzy sets is an open area of research, it is common to see researchers using a few different numbers of fuzzy sets for each attribute in order to validate their methods, usually ranging from 2 to 10, thus, repeating the complete round of experiments several times, according to the range of number of fuzzy sets used. If it were possible to establish an approximate optimal number of fuzzy sets, the total number of experiments would be greatly reduced. Thus, we propose a heuris-

ISBN: 978-989-95079-6-8

tic strategy to define this number using the Wang & Mendel method [23], which is able to generate fuzzy rule bases from data examples.

The Wang & Mendel method has been widely used due to its low complexity ($O(m \times n)$, m attributes, n examples) and the fact that it produces relatively small rule bases with good classification rates and no conflicting or redundant rules. However, nowadays it is possible to generate more precise fuzzy rule bases with even lower number of rules using other approaches, such as the genetic fuzzy systems or neural networks. The heuristic of the proposed method is based on the assumption that the Wang & Mendel method can be used as an indicator of the quality of the results that can be achieved by other methods, when using the same data base, *i.e.*, the same number, distribution, and shape of fuzzy sets.

Experiments shown in Section 5 were carried out using from 2 to 10 triangular shaped fuzzy sets defining each attribute of the domains. The idea is to run the experiments with the Wang & Mendel method before running the experiments with more time consuming approaches, and select the number of fuzzy sets that produced the best accuracy rates, or, in case of ties, the smaller number, due to interpretability reasons. Notice that a tolerance interval may be defined to allow the user to select the smaller number of fuzzy sets when a small difference between the accuracy rates is not significant, but the difference in the number of fuzzy sets is.

5 Experiments and Results

The experiments were conducted using 10-fold cross validation on 10 datasets. Table 1 shows a summary of the characteristics of the datasets used in the experiments, presenting the number of examples in each dataset (# Examples), number of attributes (# Attribs), number of classes (# Classes), and the majority class error (MCE).

Table 1: Characteristics of the datasets.

Dataset	# Examples	# Attribs	# Classes	MCE
Breast	684	10	3	34.99
Bupa	345	4	2	42.00
Credit	653	15	2	45.33
Diabetes	769	8	2	34.98
Glass	220	9	7	65.45
Heart	270	13	2	44.40
Iris	150	4	3	66.60
Segment	210	19	7	85.72
Vehicle	846	18	4	74.20
Wine	178	13	3	59.70

Two fuzzy reasoning methods, frequently employed in the fuzzy classification domain, were used, the classic and the general fuzzy reasoning methods. The classic fuzzy reasoning method uses the class of the rule with highest compatibility with an input pattern to classify it. The general fuzzy reasoning method, on the other hand, calculates the sum of compatibility degrees for each class and uses the class with highest sum to classify an input pattern. The accuracy was measured in terms of correct classification rates.

Table 2 presents the results for the experiments that were carried out using the classic fuzzy reasoning method. **Sets** describes the number of fuzzy sets used for each attribute in the experiments; **EU** represents the equalized universe method, which uses the same width for each fuzzy set, **F** represents the method that uses same frequency, and **1-R** represents the 1-R adapted method, which takes the classes into consideration when distributing the fuzzy sets.

Table 2: Classification rates - classic fuzzy reasoning method.

Sets	**2**			**3**			**4**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.96	0.53	0.64	0.96	0.59	0.89	0.95	0.71	0.84
Bupa	0.49	0.46	0.43	0.52	0.51	0.55	0.55	0.53	0.57
Credit	0.54	0.58	0.60	0.54	0.58	0.73	0.53	0.55	0.63
Dia.	0.99	1.00	0.98	0.99	0.99	1.00	1.00	0.98	0.97
Glass	0.41	0.58	0.48	0.49	0.55	0.67	0.55	0.60	0.65
Heart	0.98	0.98	0.93	1.00	1.00	0.97	1.00	0.99	0.98
Iris	0.67	0.43	0.37	0.65	0.91	0.65	0.92	0.93	0.91
Seg.	0.47	0.14	0.26	0.74	0.15	0.54	0.71	0.15	0.50
Vehi.	0.99	0.95	1.00	1.00	0.98	1.00	1.00	0.97	1.00
Wine	0.94	0.99	0.99	0.99	0.98	0.97	1.00	1.00	0.98

Sets	**5**			**6**			**7**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.92	0.69	0.76	0.89	0.66	0.71	0.91	0.65	0.70
Bupa	0.56	0.52	0.54	0.55	0.48	0.52	0.55	0.46	0.48
Credit	0.53	0.55	0.58	0.53	0.55	0.56	0.54	0.55	0.56
Dia.	1.00	0.94	0.98	1.00	1.00	0.97	0.99	1.00	0.97
Glass	0.61	0.56	0.61	0.61	0.53	0.57	0.63	0.52	0.55
Heart	1.00	0.98	1.00	1.00	0.97	1.00	0.93	0.95	0.99
Iris	0.91	0.90	0.91	0.87	0.83	0.87	0.93	0.80	0.87
Seg.	0.67	0.15	0.39	0.66	0.14	0.30	0.62	0.15	0.23
Vehi.	0.98	0.95	1.00	0.98	1.00	0.98	0.97	1.00	0.99
Wine	1.00	0.99	0.99	0.99	0.98	0.99	0.99	0.93	1.00

Sets	**8**			**9**			**10**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.90	0.63	0.69	0.74	0.63	0.69	0.80	0.61	0.66
Bupa	0.61	0.45	0.42	0.58	0.43	0.42	0.55	0.43	0.45
Credit	0.54	0.55	0.55	0.60	0.55	0.55	0.57	0.55	0.55
Dia.	0.97	0.99	0.95	0.93	0.93	0.89	0.88	0.91	0.89
Glass	0.58	0.51	0.54	0.60	0.50	0.51	0.61	0.50	0.50
Heart	0.92	0.98	0.99	0.91	0.89	0.98	0.91	0.87	0.97
Iris	0.89	0.71	0.76	0.89	0.67	0.72	0.89	0.59	0.68
Seg.	0.54	0.14	0.21	0.53	0.15	0.18	0.45	0.14	0.16
Vehi.	0.95	0.95	0.99	0.94	0.97	0.97	0.93	0.95	0.98
Wine	0.93	0.88	0.95	0.81	0.85	0.93	0.73	0.81	0.91

Table 3 presents the results for the experiments that were carried out using the general fuzzy reasoning method. **Sets** describes the number of fuzzy sets used for each attribute; **EU** represents the equalized universe method, **F** represents the method that uses same frequency, and **1-R** represents the 1-R adapted method.

Table 4 shows the best estimated number of fuzzy sets, according to the proposed method, for the classic and general fuzzy reasoning methods, using the equalized universe method (**EU**), frequency method (**F**) and 1-R adapted method (**1-R**). The suffixes **C** and **G** were used to describe the classic and general fuzzy reasoning methods, respectively. A tolerance interval of 0.02 was used. Notice that the best classification rates used to define these values are light-gray shaded in Tables 2 and 3. The proposed method was able to estimate a small number of fuzzy sets for most of the datasets.

Figures 4 and 5 present the classification rates (vertical axis) for the classic and general fuzzy reasoning methods, respectively, for each of the 10 datasets, using from 2 to 10 fuzzy sets (horizontal axis) to define each of the attributes.

To test whether there is a significant difference among the methods, the Friedman test [24] was used with the null-hypothesis that the performance of the three methods, assessed in terms of the error rate, are comparable. As the null-hypothesis was rejected with a 95% confidence level, the Bonferroni-Dunn post-hoc test (to detect whether the differences among the methods are significant) was used [24]. Results showed that the equalized universe method is signifi-

ISBN: 978-989-95079-6-8

Table 3: Classification rates - general fuzzy reasoning method.

Sets	2			3			4		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.93	0.51	0.83	0.96	0.55	0.89	0.95	0.69	0.82
Bupa	0.57	0.55	0.44	0.53	0.59	0.57	0.55	0.61	0.57
Credit	0.49	0.48	0.53	0.49	0.48	0.65	0.48	0.47	0.51
Dia.	0.99	0.15	0.99	0.99	0.32	0.95	0.97	0.24	0.81
Glass	0.44	0.33	0.51	0.49	0.36	0.61	0.57	0.37	0.59
Heart	0.09	0.19	0.13	0.09	0.15	0.34	0.08	0.08	0.11
Iris	0.69	0.50	0.38	0.70	0.93	0.63	0.97	0.92	0.93
Seg.	0.41	0.14	0.20	0.70	0.15	0.56	0.71	0.16	0.51
Vehi.	0.98	0.44	1.00	0.98	0.44	0.95	0.97	0.28	0.65
Wine	0.93	0.52	0.99	0.89	0.22	0.81	0.73	0.05	0.36
Sets	**5**			**6**			**7**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.92	0.67	0.73	0.89	0.66	0.71	0.91	0.65	0.69
Bupa	0.59	0.63	0.61	0.57	0.58	0.62	0.57	0.57	0.63
Credit	0.48	0.46	0.47	0.48	0.46	0.46	0.49	0.45	0.46
Dia.	0.93	0.11	0.51	0.88	0.06	0.25	0.80	0.02	0.11
Glass	0.54	0.27	0.50	0.56	0.25	0.39	0.55	0.17	0.29
Heart	0.06	0.02	0.04	0.05	0.01	0.04	0.04	0.01	0.02
Iris	0.94	0.90	0.91	0.89	0.80	0.84	0.94	0.77	0.83
Seg.	0.70	0.16	0.42	0.70	0.16	0.32	0.63	0.15	0.24
Vehi.	0.93	0.11	0.30	0.79	0.05	0.17	0.69	0.02	0.06
Wine	0.39	0.03	0.12	0.18	0.00	0.02	0.07	0.02	0.00
Sets	**8**			**9**			**10**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.90	0.63	0.69	0.74	0.63	0.69	0.80	0.61	0.66
Bupa	0.60	0.60	0.61	0.58	0.59	0.61	0.61	0.60	0.59
Credit	0.48	0.45	0.45	0.57	0.45	0.45	0.50	0.45	0.45
Dia.	0.69	0.01	0.06	0.58	0.00	0.02	0.47	0.01	0.01
Glass	0.51	0.15	0.21	0.53	0.10	0.14	0.52	0.07	0.11
Heart	0.03	0.00	0.01	0.03	0.01	0.00	0.02	0.00	0.00
Iris	0.91	0.73	0.77	0.94	0.69	0.73	0.91	0.57	0.61
Seg.	0.60	0.14	0.21	0.59	0.15	0.18	0.51	0.14	0.17
Vehi.	0.41	0.00	0.02	0.31	0.01	0.01	0.17	0.01	0.01
Wine	0.02	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.01

Table 4: Best estimated number of fuzzy sets for each dataset.

Method	EU - C	EU - G	F - C	F - G	1-R - C	1-R - G
Breast	2	3	4	4	3	3
Bupa	8	5	3	4	3	5
Credit	9	9	2	2	3	3
Diabetes	2	2	2	3	2	2
Glass	5	4	2	3	3	3
Heart	2	2	2	2	4	3
Iris	4	4	3	3	4	4
Seg.	3	3	2	2	3	3
Vehicle	2	2	3	2	2	2
Wine	3	2	2	2	2	2

cantly better than the 1-R method and the frequency method, with a 95% confidence level. The results of the statistical tests also showed that the 1-R method performed significantly better than the frequency method with a 95% confidence level.

Table 5 shows the average ranks for each of the methods tested for the classic and general fuzzy reasoning methods.

Table 5: Average ranks.

Methods	EU	Frequency	1-R
Classic	1.694	2.411	1.894
General	1.350	2.711	1.939

One of the reasons for the good performance of the equalized universe method may be the fact that it generates fuzzy sets with the same distance amongst them, promoting the creation of rules that will cover equal areas of the attributes partitions, rather than promoting the creation of fuzzy sets in areas with a concentration of examples but with less differences regarding the classes, which could have a deeper effect on the classification rates. The 1-R method, since it uses the classes to adjust the fuzzy sets defining each attribute, should perform

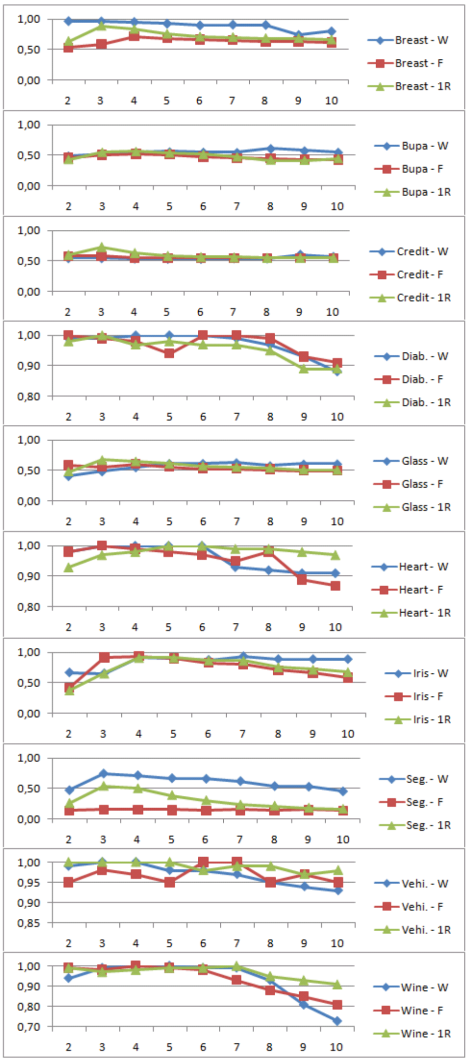

Figure 4: Results for the classic fuzzy reasoning method.

better than the frequency method.

6 Conclusions

In this paper we have tested and compared three different methods for the definition of fuzzy sets, the equalized universe method, the frequency method, and an adaptation of the 1-R supervised method. We have also proposed a heuristic method to estimate the number of fuzzy sets for each attribute using the Wang & Mendel method, to be used by other more costly approaches, able of finding smaller sets of rules with better accuracy. The experiments were carried out with 10 datasets for classification problems using 10-fold cross validation and the classic and general fuzzy reasoning methods. Each attribute of each dataset was defined using from 2 to 10 fuzzy triangular shaped sets. Statistical methods were carried out in order to find significant differences among the methods. The results show that the equalized universe method performed significantly better than the frequency and 1-R methods. The results also showed that the 1-R method performed better than the frequency method. Regarding the estimation of the number of fuzzy sets for the attributes, the proposed method provides support for the researcher to select small intervals of numbers of fuzzy sets that would be worthwhile using in more costly

ISBN: 978-989-95079-6-8

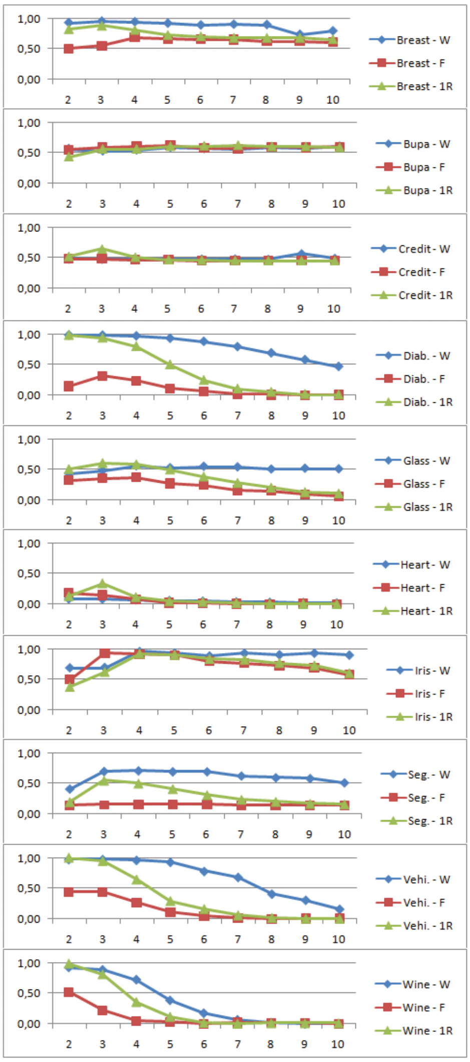

Figure 5: Results for the general fuzzy reasoning method.

approaches, based on the performance of the generated rule bases.

Acknowledgments

This research was supported by CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, and by the São Paulo State Research Foundation - FAPESP (Proc. No. 2007/05390-0).

References

[1] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] Rafael Alcalá, Jesús Alcalá-Fdez, M.J. Gacto, and Francisco Herrera. Genetic learning of membership functions for mining fuzzy association rules. *IEEE International Conference on Fuzzy Systems*, 1:1538–1543, 2007.

[3] Swarup Medasani, Jaeseok Kim, and Raghu Krishnapuram. An overview of membership function generation techniques for pattern recognition. *International Journal of Approxiamate Reasoning*, 19:391–417, 1998.

[4] M. Kaya and R. Alhajj. A clustering algorithm with genetically optimized membership functions for fuzzy association rules mining. *12th IEEE International Conference on Fuzzy Systems*, 2:881–886, 2003.

[5] Ling Zhang and Bo Zhang. The structure analysis of fuzzy sets. *Int. Journal of Approx. Reasoning*, 40:92–108, 2005.

[6] M. S. Chen and S. W. Wang. Fuzzy clustering analysis for optimizing membership functions. *Fuzzy Sets and Systems*, 103:239–254, 1999.

[7] M. E. Cintra, H. A. Camargo, and M. C. Monard. A study on techniques for the automatic generation of membership functions for pattern recognition. *C3N - Congresso da Academia Trinacional de Ciências*, 1:1–10, 2008.

[8] T. Warren Liao, Aivars K. Celmins, and Robert J. Hammell. A fuzzy c-means variant for the generation of fuzzy term sets. *Fuzzy Sets and Systems*, 135:241–257, 2003.

[9] Xuanli Lisa Xie and Gerardo Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):841–847, 1991.

[10] J.S. Jang. Adaptive-network-based fuzzy inference systems. *IEEE Trans. on SMC*, 23:665–685, 1993.

[11] Tzung-Pei Hong, Chun-Hao Chen, Yu-Lung Wu, and Yeong-Chyi Lee. A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions. *Soft Computing*, 10(11):1091–1101, 2006.

[12] Ajith Abraham. Neuro fuzzy systems: State-of-the-art modeling techniques. *Lec. Notes in CS*, 2084:269–276, 2001.

[13] Detlef Nauck, Frank Klawonn, and Rudolf Kruse. *Foundations of Neuro-Fuzzy Systems*. John Wiley & Sons, Inc., 1997.

[14] H. D. Cheng and Jim-Rong Chen. Automatically determine the membership function based on the maximum entropy principle. *Information Sciences*, 96(3-4):163–182, 1997.

[15] Hisao Ishibuchi, Takashi Yamamoto, and Tomoharu Nakashima. Fuzzy data mining: Effect of fuzzy discretization. In *ICDM '01: Proc. of the 2001 IEEE Int. Conf. on Data Mining*, pages 241–248, USA, 2001. IEEE Computer Society.

[16] Grzegorz Nieradka and Bohdan Butkiewicz. A method for automatic membership function estimation based on fuzzy measures. *International Fuzzy Systems Association World Congress - IFSA 2007*, 4529:451–460, 2007.

[17] Weibei Dou, Yuan Ren, Qian Wu, Su Ruan, Yanping Chen, Daniel Bloyet, and Jean-Marc Constans. Fuzzy kappa for the agreement measure of fuzzy classifications. *Neurocomputing*, 70(4-6):726–734, 2007.

[18] Hisao Ishibuchi, Ken Nozaki, and Hideo Tanaka. Efficient fuzzy partition of pattern space for classification problems. *Fuzzy Sets and Systems*, 59(3):295–304, 1993.

[19] J.M. Keller, M.R. Gray, and J.A. Given Jr. A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(4):580–585, 1985.

[20] Enrique H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

[21] Anthony J. Jakeman Jun and Michael McAleer Bai. *Effects of Misspecification in Estimating the Percentiles of Some Two- & Three-Parameter Distributions*. ANU, Australia, 1989.

[22] Robert C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11(1):63–90, 1993.

[23] L. Wang. The WM method completed: a flexible fuzzy system approach to data mining. *IEEE International Conference on Fuzzy Systems*, 11:768–782, 2003.

[24] Janez Demšar. Statistical comparison of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(1):1–30, 2006.