

A new fuzzy reasoning method based on the use of the Choquet integral

J. Sanz¹ C. Lopez-Molina¹ J. Cerrón¹ R. Mesiar² H. Bustince¹

¹Departamento de Automatica y Computacion, Universidad Publica de Navarra
31006 Pamplona, Spain (e-mail: joseantonio.sanz@unavarra.es)

²Faculty of Civil Engineering, Department of Mathematics, Slovak University of Technology in Bratislava
Radlinského 11, 81368 Bratislava, Slovakia

Abstract

In this work we use the Choquet integral as an aggregation function and we apply it in the fuzzy reasoning method of fuzzy rule-based classification systems. We study the behaviour of several fuzzy measures and we propose a genetic learning method of an appropriate fuzzy measure to model the interaction among the set of rules of each class. In the experimental study we show that the new proposal allows the performance of the fuzzy reasoning method of the winning rule to be outperformed when dealing with classification problems in which several fuzzy rules are fired to classify each example.

Keywords: Fuzzy rule-based classification systems, Choquet integral, fuzzy measure, fuzzy reasoning method, evolutionary algorithm.

1. Introduction

Computational intelligent methods have shown to be useful tools to solve classification problems [1]. Among them, Fuzzy Rule-Based Classification Systems (FRBCSs) [2] are widely applied since they allow the inclusion of all the available information in system modelling, that is, the information that comes from expert knowledge and the one coming from empirical measures or mathematical models. Furthermore, FRBCSs provide a good balance between accuracy and interpretability through the use of linguistic labels in the antecedent part of the fuzzy rules.

The Fuzzy Reasoning Method (FRM) [3] is a key component of FRBCSs. It is the mechanism that performs the classification of new examples based on the knowledge which has been learnt beforehand. Basically, to classify an example the FRM computes the compatibility between the example and the fuzzy rules in first place. Then, it aggregates the compatibility degrees of the rules having the same class in their consequents and finally, it assigns the example to the class having the maximum aggregated value. In the aggregation stage, it is possible to take into account only the information given by the best fired rule, by a subset of the fired rules or by all the fired rules. Depending on the way in

which the information is aggregated we find different classical inference procedures like the FRM of the winning rule [4, 5] or the FRM of the additive combination [6, 7].

In this work, we propose a new FRM based on the use of the Choquet integral [8] to perform the aggregation stage. In this manner, all the fired rules are taken into account in the decision process and we provide the FRBCS with the ability to model the interaction among the fuzzy rules by means of the fuzzy measure [8, 9]. We apply several classical aggregation functions written in terms of the Choquet integral with the corresponding fuzzy measures. Furthermore, in order to exploit the use of the Choquet integral, we propose a genetic learning method of the fuzzy measure in which we compute the best fuzzy measure to model the interaction among the rules of each class.

We analyse the performance of the new FRM in forty one data-sets selected from the KEEL repository [10, 11]. We will compare the results provided when using the different fuzzy measures using the accuracy rate as measure of performance and we will support our conclusions by conducting a statistical study as recommended in the specialized literature [12, 13]. In the comparison we will study the suitability of the new FRM depending on the number of elements to be aggregated so as to show the importance of having a proper fuzzy measure.

This work is arranged as follows: in Section 2 we provide the background on both theoretical concepts and FRBCSs. The new FRM making use of the Choquet integral and the genetic learning method of the fuzzy measure is described in detail in Section 3. The experimental results and the corresponding analysis are provided in Section 4. Finally, the concluding remarks are drawn in Section 5.

2. Preliminaries

This section is aimed at introducing the background necessary to understand the new proposal. In first place we recall some theoretical concepts and then we introduce basic concepts about FRBCSs.

2.1. Theoretical concepts

In this work we use fuzzy sets to model the linguistic labels composing the antecedents of the rules.

Definition 1 [14] A fuzzy set F defined on a finite and non-empty universe $U = \{u_1, \dots, u_n\}$ is given by

$$F = \{(u_i, \mu_F(u_i)) | u_i \in U\}$$

where $\mu_F : U \rightarrow [0, 1]$ is the membership function.

The conjunction among the antecedents of the rules is modelled by means of t-norms.

Definition 2 [15, 16] A triangular norm (t-norm) $T : [0, 1]^2 \rightarrow [0, 1]$, is an associative, commutative, increasing function such that $T(1, x) = x$ for all $x \in [0, 1]$.

We apply aggregation functions to combine several numerical values into a single one.

Definition 3 [15, 16] An aggregation function of dimension n (n -ary aggregation function) is a non decreasing mapping $\mathbb{M} : [0, 1]^n \rightarrow [0, 1]$ such that $\mathbb{M}(0, \dots, 0) = 0$ and $\mathbb{M}(1, \dots, 1) = 1$.

Finally, we recall the concept of fuzzy measure which is used in the aggregation function known as the Choquet integral [8]. We must point out that in the context of aggregation functions, fuzzy measures are used to model the importance of a coalition, that is, the relationship among the elements to be aggregated. In [17], authors studied the interpretation of the Sugeno integral in a fuzzy inference system.

Definition 4 [8, 9] Let $N = \{1, \dots, n\}$. A fuzzy measure is a function $m : 2^N \rightarrow [0, 1]$ which is monotonic (i.e. $m(A) \leq m(B)$ whenever $A \subset B$) and satisfies $m(\emptyset) = 0$ and $m(N) = 1$.

Definition 5 [8] Let m be a fuzzy measure on N . We say that m is:

- Additive if for any disjoint subsets $A, B \subseteq N$, $m(A \cup B) = m(A) + m(B)$;
- Symmetric if for any subsets $A, B \subseteq N$, $|A| = |B|$ implies $m(A) = m(B)$;

Example 1 Let $N = \{1, \dots, n\}$ and $A \subseteq N$, the following equations are examples of fuzzy measures:

(1) Cardinality.

$$m(A) = \frac{|A|}{n}$$

(2) Dirac's measure.

$$m(A) = \begin{cases} 1, & \text{if } i \in A \\ 0, & \text{if } i \notin A \end{cases}$$

where $i \in A$ is selected beforehand. We must point out that the result of the Choquet integral with this fuzzy measure is the i -th smallest value of X , that is, i -th order statistic.

- We take an arbitrary vector of weights $(w_1, \dots, w_n) \in [0, 1]^n$ with $\sum_{i=1}^n w_i = 1$.

(3) Weighted mean. We assign the following values for the fuzzy measure: $m(\{1\}) = w_1, \dots, m(\{n\}) = w_n$. For $|A| > 1$ the fuzzy measure is

$$m(A) = \sum_{i \in A} m(\{i\})$$

(4) Ordered Weighted Averaging (OWA). We assign the following values for the fuzzy measure: $m(\{i\}) = w_j$, with i being the j -th largest component to be aggregated, that is, we construct an OWA operator. For $|A| > 1$ the fuzzy measure is

$$m(A) = \sum_{i \in A} m(\{i\}),$$

(5) Exponential cardinality.

$$m(A) = \left(\frac{|A|}{n}\right)^q, \text{ with } q > 0$$

Definition 6 [8] Let m be a fuzzy measure on N and $x \in [0, \infty]^n$. The discrete Choquet integral of x with respect to m is defined by:

$$C_m(x) = \sum_{i=1}^n (x_{(i)} - x_{(i-1)}) \times m(A_{(i)}),$$

where (\cdot) is a permutation of $\{1, \dots, n\}$ such that $0 \leq x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ with the convention $x_{(0)} = 0$ and $A_{(i)} = \{A_{(i)}, \dots, A_{(n)}\}$.

2.2. Fuzzy rule-based classification systems

A classification problem consists of m training examples $x_p = (x_{p1}, \dots, x_{pn}, y_p)$, $p = 1, 2, \dots, m$ from M classes where x_{pi} is the value of the i th variable ($i = 1, 2, \dots, n$) and y_p is the class label of the p -th training example.

FRBCSs are widely used in data mining to deal with classification problems, since they allow the inclusion of all the available information in system modelling, i.e. expert knowledge, empirical measures or mathematical models. They have the advantage of generating an interpretable model and therefore, allowing the knowledge representation to be understandable for the users of the system. The two main components of FRBCSs are:

- Knowledge Base: it is composed of both the Rule Base (RB) and the Data Base, where the rules and the membership functions are stored respectively. Specifically, we use fuzzy rules of the following form:

Rule R_j : If x_1 is A_{j1} and \dots and x_n is A_{jn} then Class = C_j with RW_j (1)

where R_j is the label of the j th rule, $x = (x_1, \dots, x_n)$ is an n -dimensional example vector, A_{ji} is an antecedent fuzzy set representing a linguistic term, C_j is a class label, and $RW_j \in [0, 1]$ is the rule weight [18].

- Fuzzy Reasoning Method: it is the mechanism used to classify examples using the information stored in the knowledge base.

In the remainder of this subsection, the FRM applied to determine the classes of new examples and the fuzzy rule learning algorithm used to generate the RB are described in detail.

2.2.1. Fuzzy reasoning method

Let $x_p = (x_{p1}, \dots, x_{pn})$ be a new example to be classified, L the number of rules in the RB and M the number of classes of the problem, the steps of the FRM [3] are the following ones:

1. *Matching degree*, that is, *the strength of activation of the if-part for all rules in the RB with the example x_p* . To compute it we use a t-norm.

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \dots, \mu_{A_{jn}}(x_{pn})),$$

$$j = 1, \dots, L. \quad (2)$$

2. *Association degree*. The *association degree of the example x_p with the class of each rule in the RB*.

$$b_j^k = \mu_{A_j}(x_p) \times RW_j \quad k = \text{Class}(R_j),$$

$$j = 1, \dots, L. \quad (3)$$

3. *Example classification soundness degree for all classes*. We use an aggregation function that combines the positive association degrees calculated in the previous step.

$$Y_k = \mathbb{M}(b_j^k | j = 1, \dots, L \text{ and } b_j^k > 0),$$

$$k = 1, \dots, M. \quad (4)$$

4. *Classification*. We apply a decision function F over the example classification soundness degree for all classes. This function determines the class corresponding to the maximum value.

$$F(Y_1, \dots, Y_M) = \arg \max_{k=1, \dots, M} (Y_k) \quad (5)$$

2.2.2. Chi et al. rule generation algorithm

Chi et al. fuzzy rule learning method [19] is the extension of the Wang and Mendel algorithm [20] to solve classification problems. This method is one of the most used learning algorithms in the specialized literature due to the simplicity of the fuzzy rule generation method.

To generate the fuzzy RB this FRBCSs design method determines the relationship between the

variables of the problem and establishes an association between the space of the features and the space of the classes by means of the following steps:

1. *Establishment of the linguistic partitions*. Once the domain of variation of each feature A_i is determined, the fuzzy partitions are computed.
2. *Generation of a fuzzy rule for each example $x_p = (x_{p1}, \dots, x_{pn}, C_p)$* applying the following process:
 - 2.1 To compute the matching degree $\mu(x_p)$ of the example with all the fuzzy regions using a conjunction operator (usually modelled with the minimum or product t-norm).
 - 2.2 To assign the example x_p to the fuzzy region with the greatest matching degree.
 - 2.3 To generate a rule for the example, whose antecedent is determined by the selected fuzzy region and whose consequent is the class label of the example.
 - 2.4 To compute the rule weight. In this paper, we use the Penalized Certainty Factor (PCF) defined in [21] as:

$$RW_j = PCF_j = \frac{\sum_{x_p \in \text{Class } C_j} \mu_{A_j}(x_p) - \sum_{x_p \notin \text{Class } C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^m \mu_{A_j}(x_p)} \quad (6)$$

where $\mu_{A_j}(x_p)$ is the matching degree of the example x_p with the antecedent of the rule that is being generated.

We must remark that rules with the same antecedent can be generated during the learning process. If they have the same class in the consequent we just remove one of the duplicated rules, but if they have a different class only the rule with the highest weight is kept in the RB.

3. A fuzzy reasoning method based on the use of the Choquet integral

In this section we present our new FRM making use of the Choquet integral to aggregate the local information given by the rules in the RB. Furthermore, we propose a learning proposal of the fuzzy measure for each class of the problem in order to exploit the use of the Choquet integral in classification problems.

In first place we present the modification of the classical FRM in which the Choquet integral is applied in the third step, that is, to compute the classification soundness degree for all the classes of the problem.

Let $x_p = (x_{p1}, \dots, x_{pn})$ be a new example to be classified, L the number of rules in the RB and M the number of classes of the problem, the new FRM

applies the following steps to classify the example x_p :

1. *Matching degree*: we compute the conjunction among the antecedents of the rules by applying a t-norm.

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \dots, \mu_{A_{jn}}(x_{pn})), \quad j = 1, \dots, L. \quad (7)$$

We must point out that in this work we apply t-norms to compute this step but functions satisfying different properties could be used, like overlap functions that are non-associative [22, 23, 24].

2. *Association degree*. For each rule in the RB we weight the matching degree by the corresponding rule weight.

$$b_j^k = \mu_{A_j}(x_p) \times RW_j \quad k = \text{Class}(R_j), \quad j = 1, \dots, L. \quad (8)$$

3. *Example classification soundness degree for all classes*. We aggregate by classes the positive association degrees using the Choquet integral.

$$Y_k = C_{m_k}(b_j^k | j = 1, \dots, L \text{ and } b_j^k > 0), \quad k = 1, \dots, M, \quad (9)$$

where m_k is the fuzzy measure considered for the k -th class of the problem.

4. *Classification*. We classify the example x_p in the class having the maximum example classification soundness degree.

$$F(Y_1, \dots, Y_M) = \arg \max_{k=1, \dots, M} (Y_k) \quad (10)$$

As can be observed, the Choquet integral is applied in Equation (9). At this point, we propose to use the five fuzzy measures presented in Example 1. We must stress that these fuzzy measures are additive (the exponential cardinality is additive only when $q = 1$) and the cardinality and exponential cardinality are also symmetric.

The key point for the success of this algorithm is to make a proper definition of the fuzzy measure. In the new FRM, the first idea is to select one of the five fuzzy measures presented in Section 2 and to apply the Choquet integral with this fuzzy measure for each class of the problem. However, the set of rules of each class can interact in a different way depending on the class, which could be modelled by using a different fuzzy measure for each class of the problem.

For this reason, we propose to use the exponential cardinality and to apply an evolutionary algorithm to learn the most suitable value of the parameter q for each class. In this manner, a specific fuzzy measure would be constructed for the different

classes of the problem, that is, $m_k(A) = \left(\frac{|A|}{n}\right)^{q_k}$, with $k = 1, \dots, M$.

In order to accomplish this learning process, we consider the use of the CHC evolutionary model [25], which is a genetic algorithm that presents a good trade-off between exploration and exploitation and therefore, it is a good choice in problems with complex search spaces. The CHC evolutionary model considers a population-based selection approach in order to perform a suitable global search. It makes use of a ‘‘Population-based Selection’’ approach, where N parents and their corresponding offspring are combined to select the best N individuals to form the next population. The CHC approach uses an incest prevention mechanism and a restarting process to provoke diversity in the population, instead of the well known mutation operator.

The incest prevention mechanism is only considered in order to apply the crossover operator. In our case, two parents are only crossed if half their Hamming distance is above a predetermined threshold, Th . Since we consider a real coding scheme, we have to transform each gene considering a Gray Code (binary code) with a fixed number of bits per gene ($BITSGENE$), which is determined by the system expert. In this way, the threshold value is initialized as:

$$Th = (\#Genes \cdot BITSGENE)/4.0$$

where $\#Genes$ stands for the total length of the chromosome. Following the original CHC scheme, Th is decremented by one ($BITSGENE$ in our case) when there are no new individuals in the next generation. The scheme of this model is depicted in Fig. 1.

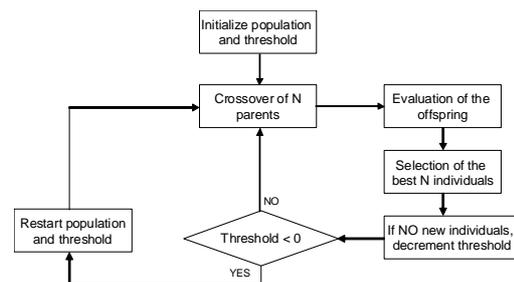


Figure 1: CHC scheme.

The specific components of the CHC model regarding our new proposal are the following ones:

- **Coding scheme**. We have a set of real parameters to be optimized (q_k , with $k = 1, \dots, M$), where the range in which we suggest to vary each one is $[0.01, 100]$. However, we do not directly encode them in a chromosome but we adapt them using chromosomes in the form:

$$C_{CHOQUET} = \{G_1, \dots, G_M\},$$

where $G_k \in [0.01, 1.99]$ with $k = 1, \dots, M$. In order to compute their real values (in the range $[0.01, 100]$) we apply Eq.(11).

$$q_k = \begin{cases} G_k, & \text{if } 0 < G_k \leq 1 \\ \frac{1}{2-G_k}, & \text{if } 1 < G_k < 2 \end{cases} \quad (11)$$

- **Initial Gene Pool.** We include an individual having all genes with value 1. In this manner, we consider the well-know cardinality fuzzy measure as a smart initialization in order to obtain at least its results.
- **Chromosome Evaluation.** We use the most common metric for classification, i.e. the accuracy rate.
- **Crossover Operator.** The crossover operator is based on the concept of environments (the offspring are generated around their parents). These kinds of operators present a good cooperation when they are introduced within evolutionary models forcing the convergence by pressure on the offspring (as the case of CHC). Figure 2 depicts the behaviour of these kinds of operators, which allow the offspring genes to be around the genes of one parent, Parent Centric BLX (PCBLX), or around a wide zone determined by both parent genes BLX- α [26]. Specifically, we consider the PCBLX operator which is described as follows:

Let $X = (x_1 \dots x_n)$ and $Y = (y_1 \dots y_n)$, ($x_i, y_i \in [a_i, b_i] \subset \mathbb{R}, i = 1 \dots n$) be the two real-coded chromosomes to be crossed. The PCBLX operator generates the two following offspring:

- $O_1 = (o_{11} \dots o_{1n})$, where o_{1i} is a randomly (uniformly) chosen number from the interval $[l_i^1, u_i^1]$, with $l_i^1 = \max\{a_i, x_i - I_i\}$, $u_i^1 = \min\{b_i, x_i + I_i\}$, and $I_i = |x_i - y_i|$.
- $O_2 = (o_{21} \dots o_{2n})$, where o_{2i} is a randomly (uniformly) chosen number from the interval $[l_i^2, u_i^2]$, with $l_i^2 = \max\{a_i, y_i - I_i\}$ and $u_i^2 = \min\{b_i, y_i + I_i\}$.

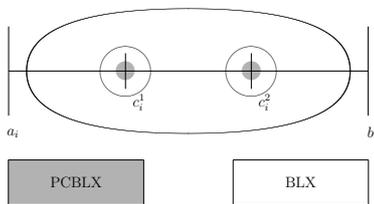


Figure 2: Scheme of the behaviour of the BLX and PCBLX operators.

- **Restarting Approach.** To get away from local optima, this algorithm uses a restarting approach since it does not apply mutation during the recombination phase. Therefore, when the threshold value is lower than zero, all the

chromosomes are regenerated randomly to introduce new diversity to the search. Furthermore, the best global solution found is included in the population to increase the convergence of the algorithm as in the elitist scheme.

4. Experimental analysis

The aim of the experimental study is to show the usefulness of the new FRM to tackle classification problems. Specifically, we compare the performance of the FRBCS when using the Choquet integral with the different fuzzy measures considered in this work. In the remainder of this section we first describe the experimental set-up and then, we analyse the obtained results.

Id.	Data-set	#Ex.	#Atts.	#Class.
app	Appendicitis	106	7	2
aus	Australian	690	14	2
bal	Balance	625	4	3
ban	Banana	5,300	2	2
bre	Breast	277	9	2
bup	Bupa	345	6	2
car	Car	1,728	6	4
cle	Cleveland	297	13	5
crx	Crx	653	15	2
eco	Ecoli	336	7	8
fla	Flare	1,066	11	6
ger	German	1,000	20	2
gla	Glass	214	9	7
hab	Haberman	306	3	2
hay	Hayes-Roth	160	4	3
hea	Heart	270	13	2
hou	Housevotes	232	16	2
ion	Ionosphere	351	33	10
iri	Iris	150	4	3
led	Led7digit	500	7	10
mag	Magic	1,902	10	2
mam	Mammographic	830	5	2
mon	Monk-2	432	6	2
new	New-Thyroid	215	5	3
pag	Page-blocks	548	10	5
pen	Penbased	1,992	16	10
pho	Phoneme	5,404	5	2
pim	Pima	768	8	2
rin	Ring	7,400	20	2
sat	Satimage	6,435	36	7
seg	Segment	2,310	19	7
shu	Shuttle	5,800	9	7
spe	Spectfheart	267	44	2
tae	Tae	151	5	3
tic	Tictactoe	958	9	2
tit	Titanic	2,201	3	2
two	Twonorm	740	20	2
veh	Vehicle	846	18	4
wine	Wine	178	13	3
wis	Wisconsin	683	9	2
zoo	Zoo	101	16	7

Table 1: Summary description for the considered data-sets.

4.1. Experimental framework

We have tested the system's performance in forty one data-sets selected from the KEEL repository [10, 11] applying a *5-folder cross-validation model*. Table 1 summarizes the following characteristics of each data-set: number of examples (#Ex.), number of attributes (#Atts.) and number of classes (#Class.). We must point out that the *Magic*, *Page-blocks*, *Penbased*, *Shuttle* and *Twonorm* data-sets are stratified-sampled to the 10% to improve

the learning process efficiency and we have removed the missing values of *Breast*, *Cleveland*, *Crx*, *House-votes*, *Mammographic* and *Wisconsin* after the partitionate.

For the Chi et al. rule learning algorithm we have used three linguistic labels per variable and the product t-norm as conjunction operator. For the genetic learning proposal presented in Section 3 we have considered populations composed by 50 individuals, 30 bits per gen in order to perform the *gray* codification and the number of evaluations is 5,000 times the number of variables.

Finally, for the use of the Dirac’s measure (Delta_Dirac) we have selected the value n for the parameter i and consequently, the results provided by this fuzzy measure are the same ones than those of the well-known FRM of the winning rule.

4.2. Analysis of the usefulness of the new fuzzy reasoning method

Table 2 contains the results in testing obtained by the new proposal when using the different fuzzy measures considered in this work. For each dataset the best results is highlighted in **bold-face**. The last column (#FiredRules) shows the average number of fired rules in each data-set, that is, we compute the average number of fired rules per class within the five folders and then we average those classes which have fired some rule. The datasets in this table are order by the average number of fired rules in order to analyse the importance of the new proposal depending on the number of elements to be aggregated.

From the results presented in Table 2 it can be observed that when the number of elements to be aggregated is less than three the behaviour of the different methods is very similar whereas when the number of elements is greater than three the performance of the proposal in which the fuzzy measure is learned for each class of the problem is clearly the best one.

In order to support these findings we have applied the Friedman aligned ranks test [27] in order to check whether there are statistical differences among the five proposals and we have depicted graphically the obtained ranks in order to easily show which is the best ranking method. Finally, we have also applied the Holm’s post-hoc test [28] to study if the best ranking method statistically outperforms the remainder ones.

When applying the Friedman aligned ranks test using the datasets in which the average number of fired rules is less than three we obtain a p-value of 0.004 and the ranks shown in Figure 3. It can be observed that the ranks associated with the different proposals are similar being the best ranking method the proposal in which we learn a different fuzzy measure per class. However, from results in Table 3 it can be concluded that under this conditions there are not statistical differences among the

i	Algorithm	Hypothesis	APV
1	Delta_Dirac	Not Rejected	0.58
2	WMean	Not Rejected	0.67
3	OWA	Not Rejected	0.67
4	Card.	Not Rejected	0.67

Table 3: Holm test to compare our approaches dealing with datasets with an average number of fired rules per class lower than 3.

different proposals.

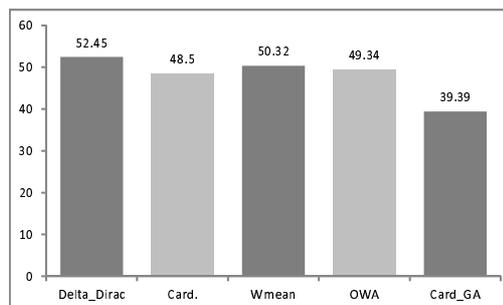


Figure 3: Rankings of the approaches tackling datasets with an average number of fired rules per class lower than 3.

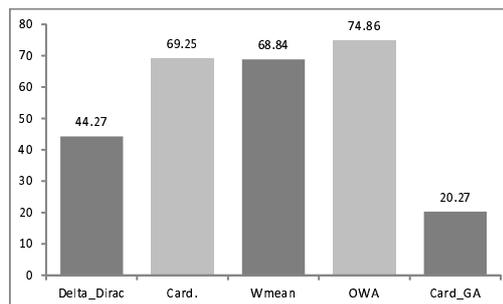


Figure 4: Rankings of the approaches using datasets with an average number of fired rules per class greater than 3.

We have conducted the same statistical study when using data-sets in which more than three rules have been fired per class. In this case, the obtained p-value with the Friedman aligned ranks test is 0.001 being the best ranking method the proposal in which the fuzzy measure has been genetically learnt (see Figure 4. Unlike the previous situation, when applying the Holm’s post-hoc test, whose results are shown in Table 4, we can conclude that the performance of the Card_GA proposal allows the results of the remainder methods to be statistically outperformed.

5. Conclusions

In this work we have presented a new FRM in which the aggregation step is performed by using the Choquet integral. We have considered five fuzzy measures in order to model the interaction among the

Dataset	Delta_Dirac	Card.	WMean	OWA	Card_GA	#FiredRules
breast	0.71	0.71	0.71	0.71	0.71	1.00
car	0.70	0.70	0.70	0.70	0.70	1.00
flare	0.63	0.63	0.63	0.63	0.63	1.00
housevotes	0.64	0.64	0.64	0.64	0.64	1.00
led7digit	0.64	0.64	0.64	0.64	0.64	1.00
tictactoe	0.65	0.65	0.65	0.65	0.65	1.00
zoo	0.72	0.72	0.72	0.72	0.72	1.00
german	0.70	0.70	0.70	0.70	0.70	1.05
hayes-roth	0.66	0.67	0.67	0.66	0.67	1.26
monk-2	0.75	0.75	0.75	0.75	0.75	1.29
crx	0.69	0.69	0.69	0.69	0.69	1.42
titanic	0.78	0.78	0.78	0.78	0.78	1.46
glass	0.58	0.59	0.57	0.60	0.60	1.74
cleveland	0.55	0.54	0.53	0.54	0.53	1.78
tae	0.55	0.56	0.56	0.56	0.56	1.84
banana	0.60	0.60	0.60	0.60	0.62	1.94
balance	0.91	0.87	0.87	0.87	0.90	2.52
pageblocks	0.91	0.92	0.92	0.92	0.92	2.76
shuttle	0.80	0.80	0.80	0.80	0.80	2.76
Mean (FR<3)	0.70	0.69	0.69	0.69	0.70	
haberman	0.72	0.72	0.73	0.72	0.73	3.03
heart	0.71	0.71	0.71	0.71	0.71	3.07
iris	0.93	0.87	0.87	0.87	0.95	3.32
newthyroid	0.85	0.86	0.87	0.87	0.93	3.44
ecoli	0.72	0.70	0.71	0.71	0.75	3.62
australian	0.83	0.82	0.82	0.82	0.83	4.10
mammographic	0.81	0.80	0.81	0.80	0.80	5.32
wisconsin	0.96	0.96	0.96	0.96	0.95	5.42
appendicitis	0.87	0.85	0.87	0.85	0.87	6.54
ion	0.92	0.92	0.91	0.91	0.92	7.26
satimage	0.47	0.50	0.50	0.51	0.58	7.42
ring	0.53	0.51	0.51	0.51	0.78	7.81
phoneme	0.72	0.71	0.72	0.72	0.75	8.67
segment	0.85	0.84	0.83	0.83	0.85	8.84
penbased	0.94	0.93	0.93	0.93	0.94	10.64
bupa	0.58	0.59	0.59	0.59	0.63	11.37
wine	0.93	0.91	0.91	0.90	0.92	11.68
vehicle	0.61	0.61	0.60	0.60	0.61	12.01
pima	0.73	0.74	0.75	0.74	0.75	16.70
magic	0.75	0.73	0.73	0.73	0.78	21.39
twonorm	0.84	0.70	0.69	0.71	0.88	26.68
spectfheart	0.63	0.62	0.61	0.62	0.63	46.92
FR>3	0.77	0.76	0.76	0.75	0.80	

Table 2: Results in testing achieved by the different approaches.

i	Algorithm	Hypothesis	APV
1	OWA	Rejected for Card_GA	5.51E-8
1	Card.	Rejected for Card_GA	1.06E-6
1	WMean	Rejected for Card_GA	1.06E-6
1	Delta_Dirac	Rejected for Card_GA	0.01

Table 4: Holm test to compare our approaches dealing with datasets with an average number of fired rules per class greater than 3.

set of rules having the same class in the consequent. The definition of the fuzzy measure is the key point for the success of the new proposal and therefore, we have proposed a learning method of the fuzzy measure associated with each class of the problem. Specifically, we have applied an evolutionary algorithm to compute the best value of the exponent when using the exponential cardinality as fuzzy measure.

Throughout the experimental study it has been shown that when the number of elements, which in our case is the number of fired rules, to be aggregated is low the results provided by the use of the different fuzzy measures are very similar. However, when the number of average fired rules is greater than three it is recommendable the use of the pro-

posal in which a proper fuzzy measure is learnt to model the interaction of the set of rules related to each class. All in all, we can conclude that: 1) the FRM of the winning rule is advisable when having few elements to be aggregated due to its low computational cost and 2) the use of the Choquet integral with the genetic learning of the fuzzy measure allows to clearly enhance the FRBCS's performance to tackle classification problems with several fired rules.

Acknowledgements

This work was partially supported by the Spanish Ministry of Science and Technology under project TIN2010-15055 and by the grant VEGA 1/0171/12.

References

- [1] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley, 2nd edition, 2001.
- [2] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer-Verlag, Berlin, 2004.

- [3] Oscar Cordon, María José del Jesus, and Francisco Herrera. A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning*, 20(1):21–45, 1999.
- [4] Ludmila I. Kuncheva. On the equivalence between fuzzy and statistical classifiers. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 4(3):245–253, 1996.
- [5] D.P. Mandal, C.A. Murthy, and S.K. Pal. Formulation of a multivalued recognition system. *Systems, Man and Cybernetics, IEEE Transactions on*, 22(4):607–620, 1992.
- [6] I.B. Turksen and Y. Tian. Combination of rules or their consequences in fuzzy expert systems. *Fuzzy Sets and Systems*, 58(1):3–40, 1993.
- [7] O. Cordon, F. Herrera, and A. Peregrin. Applicability of the fuzzy operators in the design of fuzzy logic controllers. *Fuzzy Sets and Systems*, 86(1):15–41, 1997.
- [8] G. Choquet. Theory of capacities. *Annals de l’Institut Fourier*, 5:131–295, 1953–1954.
- [9] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.
- [10] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. C. Fernández, and F. Herrera. Keel: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3):307–318, 2009.
- [11] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera. KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17(2–3):255–287, 2011.
- [12] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.
- [13] Salvador García, Alberto Fernández, Julián Luengo, and Francisco Herrera. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*, 180(10):2044–2064, 2010.
- [14] L. A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.
- [15] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners. What is an aggregation function*. Studies In Fuzziness and Soft Computing, Springer, San Mateo-California, 2007.
- [16] T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. *Aggregation Operators New Trends and Applications: Aggregation operators: properties, classes and construction methods*. Physica-Verlag, Heidelberg, 2002.
- [17] Vicenç Torra and Yasuo Narukawa. The interpretation of fuzzy integrals and their application to fuzzy systems. *International Journal of Approximate Reasoning*, 41(1):43–58, 2006.
- [18] Hisao Ishibuchi and T. Nakashima. Effect of rule weights in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 9(4):506–515, 2001.
- [19] Z. Chi, H. Yan, and T. Pham. *Fuzzy algorithms with applications to image processing and pattern recognition*. World Scientific, 1996.
- [20] L. X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(2):353–361, 1992.
- [21] Hisao Ishibuchi and T. Yamamoto. Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 13(4):428–435, 2005.
- [22] A. Jurio, H. Bustince, M. Pagola, A. Pradera, and R.R. Yager. Some properties of overlap and grouping functions and their application to image thresholding. *Fuzzy Sets and Systems*, (0):-, 2013.
- [23] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, and R. Orduna. Overlap functions. *Non-linear Analysis: Theory, Methods & Applications*, 72(3-4):1488–1499, 2010.
- [24] H. Bustince, M. Pagola, R. Mesiar, E. Hullermeier, and F. Herrera. Grouping, overlap, and generalized bintropic functions for fuzzy modeling of pairwise comparisons. *Fuzzy Systems, IEEE Transactions on*, 20(3):405–415, 2012.
- [25] L.J. Eshelman. *Foundations of Genetic Algorithms*, chapter The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, pages 265–283. Morgan Kaufman, 1991.
- [26] Francisco Herrera, Manuel Lozano, and Ana María Sánchez. A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study. *International Journal of Intelligent Systems*, 18(3):309–338, 2003.
- [27] J. L. Hodges and E. L. Lehmann. Ranks methods for combination of independent experiments in analysis of variance. *Annals of Mathematical Statistics*, 33:482–497, 1962.
- [28] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.