Article

Using the Choquet Integral in the Fuzzy Reasoning Method of Fuzzy Rule-Based Classification Systems

Edurne Barrenechea, Humberto Bustince, Javier Fernandez, Daniel Paternain and José Antonio Sanz *

Universidad Publica de Navarra, Campus Arrosadia 31006, Pamplona, Spain;
E-Mails: edurne.barrenechea@unavarra.es (E.B.); bustince@unavarra.es (H.B.);
fcojavier.fernandez@unavarra.es (J.F.); daniel.paternain@unavarra.es (D.P.)

* Author to whom correspondence should be addressed; E-Mail: joseantonio.sanz@unavarra.es;
  Tel.: +34-948-16-60-48; Fax: +34-948-16-89-24.

Abstract: In this paper we present a new fuzzy reasoning method in which the Choquet integral is used as aggregation function. In this manner, we can take into account the interaction among the rules of the system. For this reason, we consider several fuzzy measures, since it is a key point on the subsequent success of the Choquet integral, and we apply the new method with the same fuzzy measure for all the classes. However, the relationship among the set of rules of each class can be different and therefore the best fuzzy measure can change depending on the class. Consequently, we propose a learning method by means of a genetic algorithm in which the most suitable fuzzy measure for each class is computed. From the obtained results it is shown that our new proposal allows the performance of the classical fuzzy reasoning methods of the winning rule and additive combination to be enhanced whenever the fuzzy measure is appropriate for the tackled problem.

Keywords: fuzzy rule-based classification systems; Choquet integral; fuzzy measure; genetic algorithm

1. Introduction

A classification problem [1,2] consists of assigning objects into predefined groups or classes based on the observed variables related to the objects. To do so, a learning algorithm, which uses the available information, is used to give some decision function to determine the class to which the objects belong.

Fuzzy Rule-Based Classification Systems (FRBCSs) [3] aside from their good performance provide a model close to the one used by humans, since it is composed of a set of rules formed of linguistic labels. Due to these reasons, they are widely used to deal with real world problems [4]. The two main components of FRBCSs are the knowledge base, where it is stored the information about the problem, and the Fuzzy Reasoning Method (FRM).

The FRM is an inference procedure that uses the information in the knowledge base to determine the class in which new examples are classified. To do so, in first place the local information is computed, that is, the compatibility between the example and each fuzzy rule in the system. Then, this local information is aggregated to generate global information that is associated with each class of the problem and finally, the example is classified in the class having the maximum global information.

The FRM of the winning rule is traditionally used in the specialized literature [5–9]. It uses the maximum as aggregation function [10,11] to obtain the global information. This FRM only considers per each class the information given by a single fuzzy rule having the greatest compatibility with the example, and consequently it ignores the available information given by the remainder fuzzy rules of the system.

In this paper, we propose a new FRM that takes into account the information given by several or even all the fuzzy rules in the system. To do so, we consider the use of the Choquet integral [12,13] as the aggregation operator in the FRM. The Choquet integral is related to a fuzzy measure [12,14], which models the interaction among the elements to be aggregated (the information given by the rules of the system in this case). Therefore, a key point is the choice of an appropriate fuzzy measure for each problem we want to deal with. To perform this choice, we propose to use a genetic algorithm [15] in order to learn the most suitable fuzzy measure for each class to carry out the aggregation stage.

In order to study the usefulness of the new proposal, we apply the well-known Chi *et al.*'s algorithm [8] to accomplish the fuzzy rule learning process. We compare the performance of the classic FRMs of the winning rule and additive combination with respect to both the ones obtained when using the Choquet integral related to several fuzzy measures and the Choquet integral when the fuzzy measure is genetically learned. The behaviour of the approaches is tested in seventeen numerical dataset selected from the KEEL data-set repository [16,17], and in order to support our conclusions, we use some non-parametric statistical tests [18,19].

This paper is arranged as follows. In Section 2 we recall some preliminary concepts that are necessary to understand the paper. The new proposal is described in detail in Section 3, including the new FRM, the fuzzy measures considered in the paper and the method to genetically learn the fuzzy measure. Next, the experimental set-up and the corresponding analysis of the results are presented in Sections 4 and 5 respectively. Finally, the main conclusions are drawn in Section 6.

2. Preliminaries

This section is aimed at introducing the background necessary to understand the new proposal. In first place we recall some theoretical concepts, next we introduce basic concepts about FRBCSs and finally we describe the evolutionary model considered in this paper.

2.1. Theoretical Concepts

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

Definition 1 *[20] A fuzzy set F defined on a finite and non-empty universe $U = \{u_1, ..., u_n\}$ is given by*

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

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

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

Definition 2 *[10,11] A triangular norm (t-norm) $T : [0, 1]^2 \to [0, 1]$, is an associative, commutative, increasing function such that $T(1, x) = x$ for all $x \in [0, 1]$.*

When several numerical values need to be combined into a single value, we use aggregation functions.

Definition 3 *[10,11] An aggregation function of dimension n (n-ary aggregation function) is a non-decreasing mapping $\mathbb{M} : [0, 1]^n \to [0, 1]$ such that $\mathbb{M}(0, \dots, 0) = 0$ and $\mathbb{M}(1, \dots, 1) = 1$.*

Finally, we recall the necessary concept that derives in the definition of the aggregation function known as the Choquet integral [12].

Definition 4 *[12,14] Let $N = \{1, ..., n\}$. A fuzzy measure is a function $m : 2^N \to [0, 1]$ which is monotonic (i.e., $m(A) \leq m(B)$ whenever $A \subset B$) and satisfies $m(\emptyset) = 0$ and $m(N) = 1$.*

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.

Definition 5 *[12] 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 $m(A) = m(B)$;*

Definition 6 *[12] 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:*

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

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

2.2. Fuzzy Rule-Based Classification Systems

FRBCSs are widely used in data mining, 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.
- Fuzzy Reasoning Method: It is the mechanism used to classify examples using the information stored in the knowledge base.

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

We use fuzzy rules of the following form:

$$\text{Rule } R_j : \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \text{ then Class} = C_j \text{ with } RW_j \tag{1}$$

where R_j is the label of the jth rule, $x = (x_1, \ldots, x_n)$ 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 [21].

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 $x_p = (x_{p1}, \ldots, x_{pn})$ 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 [22] are the following:

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.

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \qquad j = 1, \ldots, L \tag{2}$$

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

$$b_j^k = \mu_{A_j}(x_p) \times RW_j^k \qquad k = Class(R_j), \quad j = 1, \ldots, L \tag{3}$$

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

$$Y_k = \mathbb{M}(b_j^k | j = 1, \ldots, L \text{ and } b_j^k > 0), \qquad k = 1, \ldots, M \tag{4}$$

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.

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

2.2.2. Chi *et al*. Rule Generation Algorithm

Chi *et al*. fuzzy rule learning method [8] is the extension of the Wang and Mendel algorithm [23] 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 Ruspini's fuzzy partitions are computed using triangular shaped membership functions in this paper.

2. *Generation of a fuzzy rule for each example* $x_p = (x_{p1}, \ldots, x_{pn}, C_p)$ 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 [24] as:

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

    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.

2.3. Evolutionary Model

In this paper, we consider the evolutionary model of CHC [25] to accomplish the learning of the fuzzy measure. CHC is a GA that presents a good trade-off between exploration and exploitation, being a good choice in problems with complex search spaces.
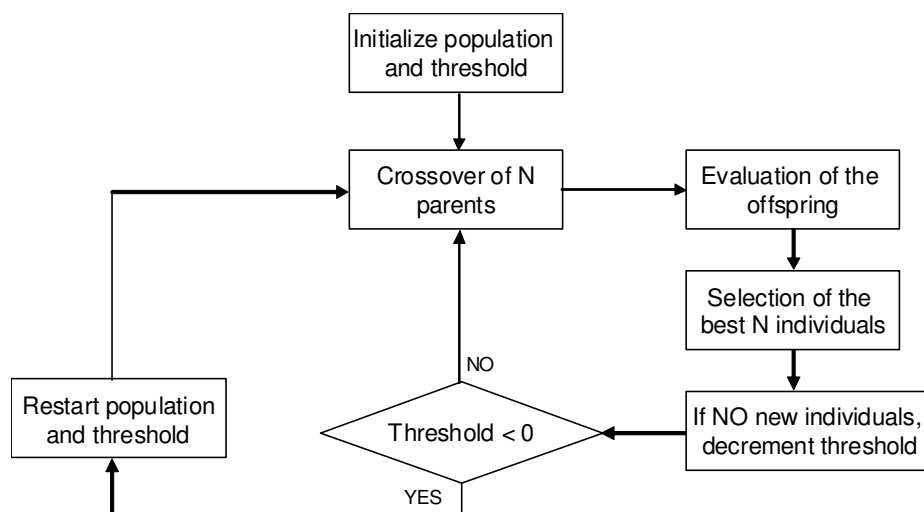
The CHC evolutionary model considers a population-based selection approach in order to perform a suitable global search. It makes use of a "Population-based Selection" approach, where *N* parents and their corresponding offspring are combined to select the best *N* individuals to form the next population. 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 algorithm restarts when Th is below zero. The scheme of this model is depicted in Figure 1.

Figure 1. CHC scheme.

3. A Novel Fuzzy Reasoning Method Using the Choquet Integral

This section is aimed at describing our new FRM making use of the Choquet integral to aggregate the local information given by the rules in the RB, that is, the values of b_j^k computed with Equation (3). Specifically, we propose to modify the third step of the general FRM introduced in Section 2.2.1. by applying Equation (7) instead of Equation (4).

$$Y_k = C_{m_k}(b_j^k | j = 1, \ldots, L \text{ and } b_j^k > 0), \qquad k = 1, \ldots, M \tag{7}$$

where m_k is the fuzzy measure considered for the k-th class of the problem, M is the number of classes of the classification problem and L is the number of rules composing the RB.

In the remainder of this section, we introduce the fuzzy measures considered in this paper in first place, and then we provide a learning proposal in which we optimize the fuzzy measure for each class of the problem.

3.1. Fuzzy Measures

According to Equation (7) we apply the Choquet integral to obtain the global information from the local information given by each rule of the system. A key point on the success of the Choquet integral is the definition of the fuzzy measure related to it. Let $N = \{1, ..., n\}$ and $A \subseteq N$, we consider the use of the following five fuzzy measures:

(1) Cardinality or uniform measure.

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

(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 $(w_1, ..., w_n) \in [0, 1]^n$ with $\sum_{i=1}^{n} w_i = 1$.

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

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

We must point out that all these fuzzy measures are additive (the exponential cardinality is additive only when $q = 1$) and the cardinality and exponential cardinality are also symmetric.

3.2. Fuzzy Measure Learning Method

The definition of an appropriate fuzzy measure plays an essential role in the success of the Choquet integral. In the proposed FRM, the first attempt is to apply the Choquet integral with the same fuzzy measure for each class of the problem. However, the set of rules of each class can interact in a different way. This fact could be taken into account by taking different values of the parameter q for each class using the exponential cardinality. 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 $k = 1, \ldots, M$. Consequently, we

propose a learning method to compute the most appropriate fuzzy measure for each class of the problem, since it can provoke an increase on the system's accuracy.

In order to carry out this optimization problem, we consider the use of the CHC evolutionary model [25] (see Section 2.3). In the remainder of this section, we describe the specific features of our evolutionary model.

- **Coding scheme**. We have a set of real parameters to be optimized (q_k, with $k = 1, ..., M$), 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:

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

where $G_k \in [0.01, 1.99]$ with $k = 1, ..., M$. In order to compute their real values (in the range $[0.01, 100]$) we apply Equation (8).

$$q_k = \begin{cases} G_k, & \text{if } 0 < G_k \leq 1 \\ \frac{1}{2 - G_k}, & \text{if } 1 < G_k < 2 \end{cases} \tag{8}$$

The change of range is provoked because we need to give the same chances to produce offspring in the ranges $[0.01, 1]$ and $[1, 100]$ after applying the crossover operator. Looking at how the crossover operator works, if we encoded the parameters in the range $[0.01, 100]$ we would favour the generation of offspring in the range $[1, 100]$ and consequently, we would reduce the probability of the generation of offspring in the range $[0.01, 1]$. For this reason, we adapt the range in order to solve this undesirable situation.
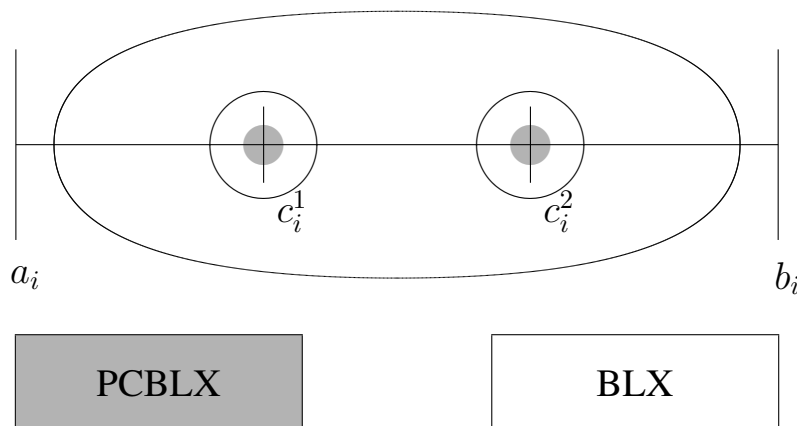
- **Initial Gene Pool**. We include an individual having all genes with value 1. In this manner, at least we obtain the results provided by the cardinality measure.

- **Chromosome Evaluation**. We use the most common metric for classification, *i.e.*, the accuracy rate that is the percentage of correctly classified examples.

- **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. Specifically, we consider the PCBLX operator that is based on the BLX-α [26].

  The PCBLX is described as follows. Assuming that $X = (x_1 \cdots x_n)$ and $Y = (y_1 \cdots y_n)$, ($x_i, y_i \in [a_i, b_i] \subset \Re, i = 1 \cdots n$) are two real-coded chromosomes that are going to be crossed. The PCBLX operator generates the two following offsprings:

  - $O_1 = (o_{11} \cdots o_{1n})$, 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 $I_i = \mid x_i - y_i \mid$.
  - $O_2 = (o_{21} \cdots o_{2n})$, 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 $u_i^2 = \min\{b_i, y_i + I_i\}$.

- **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.

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

4. Experimental Framework

In this section, we first present the real world classification data-sets selected for the experimental study. Next, we introduce the parameter set-up considered along this study. Finally, we introduce the statistical tests that are necessary to compare the results achieved throughout the experimental study.

4.1. Data-Sets

We have selected seventeen numerical data-sets selected from the KEEL data-set repository [16,17]. Table 1 summarizes the properties of the selected data-sets, showing for each data-set the number of examples (#Ex.), the number of attributes (#Atts.) and the number of classes (#Class.). We must point out that the *magic*, *ring* and *twonorm* data-sets have been stratified sampled at 10% in order to reduce their size for training and examples with missing values have been removed like in the *wisconsin* data-set.

A *5-fold cross-validation model* was considered in order to carry out the different experiments. That is, we split the data-set into 5 random partitions of data, each one with 20% of the examples, and we use a combination of 4 of them (80%) to train the system and the remaining one to test it. This process is repeated five times by using a different partition to test the system each time. We consider the average result of the five partitions as the final classification rate of the algorithm. This procedure is a standard for testing the performance of classifiers [27,28].

Table 1. Summary Description for the employed data-sets.

Id.	Data-set	#Ex.	#Atts.	#Class.
bal	Balance	625	4	3
ban	Banana	5300	2	2
eco	Ecoli	336	7	8
gla	Glass	214	9	6
iri	Iris	150	4	3
led	Led7digit	500	7	10
mag	Magic	1902	10	2
new	Newthyroid	215	5	3
pho	Phoneme	5404	5	2
pim	Pima	768	8	2
rin	Ring	740	20	2
seg	Segment	2310	19	7
tit	Titanic	2201	3	2
two	Twonorm	740	20	2
veh	Vehicle	846	18	4
win	Wine	178	13	3
wis	Wisconsin	683	11	2

4.2. Configuration of the Proposals and Notation

We will apply the following configuration for the Chi *et al*. rule generation algorithm:

- Conjunction operator: Product *t*-norm.
- Rule weight: Penalized Certainty Factor.
- Number of linguistic labels: 3.

For the new proposal using the Dirac's fuzzy measure, the value selected as i is the one associated with the median, that is, if the number of elements is odd we take $i = \frac{n+1}{2}$, whereas if the number of elements is even we take $i = \frac{n}{2} + 1$. We must stress that when using the Dirac's measure taking $i = n$ we obtain the same results provided by the maximum (Max.). In addition, if we used $i = 1$, we would obtain the results provided by the minimum as aggregation function [29] but we do not include them since the achieved performance is poor.

Regarding the genetic process, we have used the values suggested in [30], which are:

- Population Size: 50 individuals.
- Number of evaluations: 20,000.
- Bits per gene for the Gray codification (for incest prevention): 30 bits.

Finally, for the sake of clarity, Table 2 shows the names given to the different approaches considered along the experimental study.

Table 2. Names given to the seven approaches used in the paper.

Name	Aggregation function	Fuzzy Measure
Max.	Maximum	-
AC	Sum (This function is not an aggregation function as introduced in Definition 3 because it does not provide a result in $[0, 1]$.)	-
Card.	Choquet integral	Cardinality
Dirac.	Choquet integral	Dirac's measure
WMean.	Choquet integral	Weighted mean
OWA	Choquet integral	OWA
Card_GA	Choquet integral	Exponential cardinality

4.3. Statistical Tests for Performance Comparison

In this paper, we use some hypothesis validation techniques in order to give statistical support to the analysis of the results [31,32]. We will use non-parametric tests because the initial conditions that guarantee the reliability of the parametric tests cannot be fulfilled, which implies that the statistical analysis loses credibility with these parametric tests [18].

Specifically, we use the Friedman aligned ranks test [33] to detect statistical differences among a group of results and the Holm post-hoc test [34] to find the algorithms that reject the equality hypothesis with respect to a selected control method.

The post-hoc procedure allows us to know whether a hypothesis of comparison could be rejected at a specified level of significance α. Furthermore, we compute the adjusted p-value (APV) in order to take into account the fact that multiple tests are conducted. In this manner, we can directly compare the APV with respect to the level of significance α in order to be able to reject the null hypothesis.

Furthermore, we consider the method of aligned ranks of the algorithms in order to show graphically how good a method is with respect to its partners. The first step to compute this ranking is to obtain the average performance of the algorithms in each data set. Next, we compute the subtractions between the accuracy of each algorithm minus the average value for each data-set. Then, we rank all these differences in descending order and, finally, we average the rankings obtained by each algorithm. In this manner, the algorithm that achieves the lowest average ranking is the best one.

These tests are suggested in the studies presented in [18,31,35], where it is shown that their use in the field of machine learning is highly recommended.

5. Experimental Results
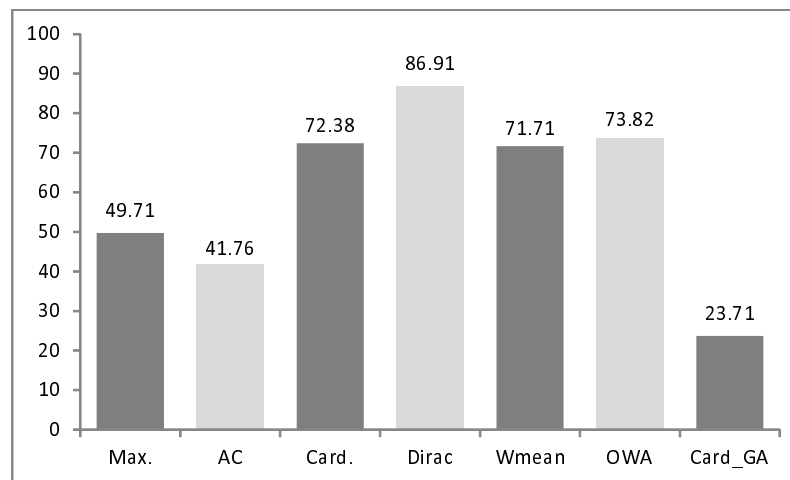
Table 3 shows the classification accuracy along with the standard deviation obtained both in training and in testing by the different approaches used in the experimental study, where the best global result for each data-set is emphasised in **bold-face**. From these results it can be observed that the behaviour of the new proposal using standard fuzzy measures (Card., Dirac, WMean and OWA) is

similar among themselves except the proposal associated with Dirac's measure, since it provides worse results. Regarding the behaviour of these proposals with respect the classical FRM of the winning rule, which uses the maximum as aggregation function (Max.), we can observe that although they provide a worse mean performance, the lack of accuracy is mainly due to three datasets, namely *balance*, *iris* and *twonorm*, the latter being especially bad for these proposals. However, when the fuzzy measure is appropriate for the specific problem we are dealing with, like the one that has been genetically learned (Card_GA), both the increase in the system's performance and the robustness of the method can be noted, since it provides the best result in eleven out of the seventeen datasets of the study. Finally, we also compare our new FRM with respect to the classical additive combination FRM (AC), which aggregates the positive association degrees by summing them and therefore does not provide a result in the range between the minimum and maximum of the aggregated values like the Choquet integral. In this comparison, we can stress that the Card_GA proposal obtains an average mean enhancement of 2.04%, which is based on the improvement of the performance of the AC FRM in more than half of the data-sets.

Table 3. Results in training (Tr.) and testing (Tst.) along with their standard deviations achieved by the seven approaches considered in this paper.

Data Set	Max. Tr.	Max. Tst	AC Tr.	AC Tst	Card. Tr.	Card. Tst	Dirac Tr.	Dirac Tst	WMean Tr.	WMean Tst	OWA Tr.	OWA Tst	Card_GA Tr.	Card_GA Tst
Bal	91.52 ± 0.23	**90.56 ± 1.04**	91.52 ± 0.23	**90.56 ± 1.19**	89.96 ± 1.11	86.72 ± 2.57	89.48 ± 0.41	87.04 ± 1.04	90.20 ± 0.93	86.72 ± 1.93	90.24 ± 1.07	86.88 ± 2.37	91.52 ± 0.23	89.92 ± 0.91
Ban	60.36 ± 0.39	60.32 ± 1.33	59.83 ± 0.28	60.02 ± 1.30	60.68 ± 0.60	60.47 ± 1.62	62.00 ± 1.18	61.77 ± 0.79	60.34 ± 0.63	60.08 ± 1.44	60.55 ± 0.71	60.36 ± 1.34	63.11 ± 0.90	**62.36 ± 1.28**
Eco	76.27 ± 1.34	71.76 ± 6.69	78.42 ± 0.86	73.53 ± 5.40	75.97 ± 2.03	69.71 ± 5.94	73.36 ± 2.57	67.94 ± 6.01	75.60 ± 1.45	70.59 ± 6.33	75.45 ± 1.73	70.88 ± 6.36	83.26 ± 2.12	**75.00 ± 9.24**
Gla	66.01 ± 2.58	57.67 ± 1.04	66.47 ± 1.84	59.07 ± 2.65	62.61 ± 3.39	58.60 ± 3.82	62.15 ± 4.82	57.21 ± 2.08	61.09 ± 3.97	57.21 ± 4.53	63.31 ± 2.86	**59.53 ± 2.65**	68.23 ± 1.36	**59.53 ± 3.12**
Iri	93.00 ± 0.75	92.67 ± 1.49	95.33 ± 1.51	**94.67 ± 4.47**	88.50 ± 1.09	87.33 ± 1.49	84.67 ± 1.73	84.00 ± 4.35	88.67 ± 1.92	87.33 ± 2.79	87.00 ± 1.26	86.67 ± 2.36	96.67 ± 0.83	**94.67 ± 2.98**
Led	75.90 ± 2.96	**64.20 ± 5.63**	75.90 ± 2.96	**64.20 ± 5.63**	75.90 ± 2.96	**64.20 ± 5.63**	75.90 ± 2.96	**64.20 ± 5.63**	75.90 ± 2.96	**64.20 ± 5.63**	75.90 ± 2.96	**64.20 ± 5.63**	75.90 ± 2.96	**64.20 ± 5.63**
Mag	76.00 ± 0.61	74.75 ± 1.85	76.49 ± 0.56	75.38 ± 1.51	74.11 ± 0.22	72.65 ± 1.50	71.65 ± 0.43	70.81 ± 1.04	74.01 ± 0.31	72.97 ± 1.65	74.29 ± 0.32	73.07 ± 1.40	79.31 ± 0.93	**77.80 ± 3.62**
New	86.40 ± 1.06	85.12 ± 3.53	87.44 ± 1.40	86.51 ± 4.16	87.79 ± 1.23	86.05 ± 3.68	89.19 ± 0.88	86.98 ± 4.22	87.91 ± 1.61	87.44 ± 3.53	88.02 ± 1.52	86.51 ± 3.45	94.42 ± 1.27	**92.56 ± 4.47**
Pho	71.91 ± 0.11	71.91 ± 0.37	72.82 ± 0.09	72.62 ± 0.64	71.54 ± 0.18	71.23 ± 0.80	72.20 ± 0.36	72.16 ± 0.71	71.79 ± 0.27	71.56 ± 0.60	72.03 ± 0.26	71.93 ± 1.01	76.16 ± 0.25	**75.39 ± 1.28**
Pim	75.46 ± 0.70	72.99 ± 0.98	74.64 ± 0.50	73.25 ± 1.55	75.75 ± 0.40	73.77 ± 1.75	75.36 ± 0.49	73.38 ± 1.66	75.98 ± 0.48	74.55 ± 2.53	75.46 ± 0.28	74.03 ± 2.43	78.39 ± 0.71	**75.06 ± 1.18**
Rin	59.39 ± 0.44	52.70 ± 0.83	57.70 ± 0.44	52.03 ± 0.48	53.75 ± 0.44	51.08 ± 0.37	51.11 ± 0.28	50.68 ± 0.48	53.99 ± 0.64	51.08 ± 0.37	53.72 ± 0.41	51.22 ± 0.30	81.35 ± 1.69	**77.70 ± 1.85**
Seg	86.01 ± 1.31	85.02 ± 2.26	86.03 ± 0.95	84.81 ± 1.84	84.40 ± 0.92	83.51 ± 1.93	78.40 ± 1.93	77.92 ± 3.93	84.43 ± 1.22	83.38 ± 2.16	83.99 ± 0.91	82.94 ± 2.62	87.18 ± 0.74	**85.06 ± 2.23**
Tit	78.33 ± 0.41	**78.32 ± 1.71**	78.33 ± 0.41	**78.32 ± 1.71**	78.33 ± 0.41	**78.32 ± 1.71**	78.33 ± 0.41	**78.32 ± 1.71**	78.33 ± 0.41	**78.32 ± 1.71**	78.33 ± 0.41	**78.32 ± 1.71**	78.33 ± 0.41	**78.32 ± 1.71**
Two	87.13 ± 0.77	83.78 ± 1.72	94.97 ± 0.51	**93.24 ± 1.85**	73.34 ± 1.23	70.41 ± 4.39	63.68 ± 1.44	62.57 ± 4.91	73.01 ± 1.33	69.19 ± 4.29	74.02 ± 2.23	70.68 ± 3.59	91.45 ± 0.44	87.57 ± 1.83
Veh	66.11 ± 0.80	**61.41 ± 3.66**	64.16 ± 0.70	61.29 ± 3.26	64.21 ± 0.56	60.94 ± 4.29	57.54 ± 1.25	53.41 ± 3.29	64.18 ± 0.85	60.12 ± 3.56	63.95 ± 0.58	59.65 ± 3.26	67.85 ± 0.48	60.59 ± 3.53
Win	98.74 ± 0.58	92.78 ± 5.41	98.74 ± 0.59	**93.33 ± 5.76**	96.21 ± 0.62	91.11 ± 5.69	89.61 ± 1.32	85.56 ± 4.12	95.93 ± 0.76	91.11 ± 5.69	96.21 ± 1.07	90.00 ± 4.65	99.86 ± 0.31	92.22 ± 4.56
Wis	98.17 ± 0.29	95.62 ± 1.37	97.99 ± 0.45	95.77 ± 1.31	98.21 ± 0.27	**96.06 ± 1.42**	98.13 ± 0.24	95.91 ± 1.60	98.24 ± 0.28	95.77 ± 1.66	98.21 ± 0.27	**96.06 ± 1.42**	98.54 ± 0.13	95.33 ± 1.42
Mean	79.22 ± 0.90	75.98 ± 2.41	79.81 ± 0.84	76.98 ± 2.63	77.13 ± 1.04	74.24 ± 2.86	74.87 ± 1.33	72.34 ± 2.80	77.03 ± 1.18	74.21 ± 2.96	77.10 ± 1.11	74.29 ± 2.74	83.03 ± 0.93	**79.02 ± 2.99**

These facts are confirmed in Figure 3, where it is clearly shown that Card_GA is the best ranking method. The *p*-value obtained with the Friedman aligned ranks test is 0.02, which confirms the existence of statistical differences among these seven approaches. For this reason, we perform the Holm post-hoc test to check whether the best ranking method (Card_GA) is able to statistically enhance the remainder methods. From results in Table 4, the goodness of the proposal using the Choquet integral with a suitable fuzzy measure is clearly determined, since it outperforms both the proposals using a standard fuzzy measure and the classical FRM of the winning rule. Furthermore, the obtained APV shows that the Card_GA allows the performance of the additive combination FRM to be clearly enhanced. Therefore, it can be concluded that the best approach is the one that makes use of the Choquet integral with the fuzzy measure genetically learned.

Figure 3. Rankings of the seven approaches considered in the study.

Table 4. Holm test to compare Card_GA with respect to the different approaches.

i	Algorithm	APV
1	Dirac	5.52E–7
2	OWA	1.14E–4
3	Card.	1.56E–4
4	WMean	1.56E–4
5	Max.	0.06
6	AC	0.13

6. Conclusions

In this paper we have proposed a novel FRM in which the Choquet integral is used to aggregate the local information of the rules. The Choquet integral is associated with a fuzzy measure, which allows us to model the relationship among the rules. For this reason, we have applied several standard fuzzy measures in order to take into account such an interaction. However, the definition of an appropriate fuzzy measure is a complex problem and consequently, we have proposed a genetic learning method in which a fuzzy measure is computed to aggregate the information related to the different classes of the problem.

In the experimental study, we have used the Chi *et al.*'s algorithm to generate the fuzzy rules. In order to test the goodness of our method, we have used a wide benchmark of numerical data-sets to compare the behaviour of the classical FRMs of both the winning rule and the additive combination with respect to our new approach using both several well-known fuzzy measures and the fuzzy measure genetically learned. From this comparison, it can be concluded that the use of our new approach is advisable to face classification problems when the fuzzy measure is learnt to suit the features of each specific problem, since it statistically outperforms the results of the FRM of the winning rule and it clearly enhances the performance of the additive combination FRM.

Acknowledgements

References

1. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; John Wiley: Hoboken, NJ, USA, 2001.
2. Alpaydin, E. *Introduction to Machine Learning*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2010.
3. Ishibuchi, H.; Nakashima, T.; Nii, M. *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*; Springer-Verlag: Berlin/Heidelberg, Germany, 2004.
4. Samantaray, S.R.; El-Arroudi, K.; Joós, G.; Kamwa, I. A fuzzy rule-based approach for islanding detection in distributed generation. *IEEE Trans. Power Deliv.* **2010**, *25*, 1427–1433.
5. Kuncheva, L.I. On the equivalence between fuzzy and statistical classifiers. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **1996**, *4*, 245–253.
6. Mandal, D.; Murthy, C.; Pal, S. Formulation of a multivalued recognition system. *Syst. Man Cybern. IEEE Trans.* **1992**, *22*, 607–620.
7. Ishibuchi, H.; Nozaki, K.; Tanaka, H. Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy Sets Syst.* **1992**, *52*, 21–32.
8. Chi, Z.; Yan, H.; Pham, T. *Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition*; World Scientific: Singapore, 1996.
9. Ray, K.S.; Ghoshal, J. Approximate reasoning approach to pattern recognition. *Fuzzy Sets Syst.* **1996**, *77*, 125–150.
10. Beliakov, G.; Pradera, A.; Calvo, T. *Aggregation Functions: A Guide for Practitioners. What is an aggregation function; Studies in Fuzziness and Soft Computing*; Springer: San Mateo, CA, USA, 2007; pp. 1–37.
11. Calvo, T.; Kolesarova, A.; Komornikova, M.; Mesiar, R. *Aggregation Operators New Trends and Applications: Aggregation Operators: Properties, Classes and Construction Methods*; Physica-Verlag: Heidelberg, Germany, 2002; pp. 3–104.
12. Choquet, G. Theory of capacities. *Ann. I'Inst. Fourier* **1953**, *5*, 131–295.
13. Klement, E.P.; Mesiar, R. Discrete integrals and axiomatically defined functionals. *Axioms* **2012**, *1*, 9–20.
14. Sugeno, M. Theory of Fuzzy Integrals and It's Applications. Ph.D. Thesis, Tokyo Institute of Techonology, Tokyo, Japan, 1974.
15. Holland, J.H. Genetic algorithms. *Sci. Am.* **1992**, *267*, 66–72.
16. Alcalá-Fdez, J.; Sánchez, L.; García, S.; del Jesus, M.J.; Ventura, S.; Garrell, J.; Otero, J.; Romero, C.; Bacardit, J.; Rivas, V.; *et al.* KEEL: A software tool to assess evolutionary algorithms for data mining problems. *Soft Comput.* **2009**, *13*, 307–318.

17. Alcalá-Fdez, J.; Fernández, A.; Luengo, J.; Derrac, J.; García, S.; Sánchez, L.; Herrera, F. KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *J. Mult.-Valued Logic Soft Comput.* **2011**, *17*, 255–287.

18. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.

19. García, S.; Fernández, A.; Luengo, J.; Herrera, F. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Inf. Sci.* **2010**, *180*, 2044–2064.

20. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353.

21. Ishibuchi, H.; Nakashima, T. Effect of rule weights in fuzzy rule-based classification systems. *IEEE Trans. Fuzzy Syst.* **2001**, *9*, 506–515.

22. Cordón, O.; del Jesus, M.J.; Herrera, F. A proposal on reasoning methods in fuzzy rule-based classification systems. *Int. J. Approx. Reason.* **1999**, *20*, 21–45.

23. Wang, L.X.; Mendel, J.M. Generating fuzzy rules by learning from examples. *IEEE Trans. Syst. Man Cybern.* **1992**, *25*, 353–361.

24. Ishibuchi, H.; Yamamoto, T. Rule weight specification in fuzzy rule-based classification systems. *IEEE Trans. Fuzzy Syst.* **2005**, *13*, 428–435.

25. Eshelman, L. The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional Genetic Recombination. In *Foundations of Genetic Algorithms*; Morgan Kaufman: San Francisco, CA, USA, 1991; pp. 265–283.

26. Herrera, F.; Lozano, M.; Sánchez, A.M. A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study. *Int. J. Intell. Syst.* **2003**, *18*, 309–338.

27. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* **2012**, *42*, 463–484.

28. Galar, M.; Fernndez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognit.* **2011**, *44*, 1761–1776.

29. Bardossy, A.; Duckstein, L. *Fuzzy Rule-Based Modeling with Applications to Geophysical, Biological, and Engineering Systems*; CRC Press: Boca Raton, FL, USA, 1995.

30. Sanz, J.; Fernandez, A.; Bustince, H.; Herrera, F. Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning. *Inf. Sci.* **2010**, *180*, 3674–3685.

31. García, S.; Fernández, A.; Luengo, J.; Herrera, F. A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. *Soft Comput.* **2009**, *13*, 959–977.

32. Sheskin, D. *Handbook of Parametric and Nonparametric Statistical Procedures*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.

33. Hodges, J.L.; Lehmann, E.L. Ranks methods for combination of independent experiments in analysis of variance. *Ann. Math. Stat.* **1962**, *33*, 482–497.

34. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.

35. García, S.; Herrera, F. An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *J. Mach. Learn. Res.* **2008**, *9*, 2677–2694.