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

J. Sanz 1 C. Lopez-Molina 1 J. Cerrón 1 R. Mesiar 2 H. Bustince 1

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 \( U = \{u_1, \ldots, u_n\} \) 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 \( \mathbb{M}(0, \ldots, 0) = 0 \) and \( \mathbb{M}(1, \ldots, 1) = 1 \).

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 the result of the Choquet integral is an associative, commutative, increasing function such that \( T(1, x) = x \) for all \( x \in [0, 1] \).

We say that \( m \) and satisfies

\[
\mathbb{M}(\emptyset) = 0 \quad \text{and} \quad \mathbb{M}(N) = 1.
\]

**Definition 4** [8, 9] Let \( N = \{1, \ldots, n\} \). A fuzzy measure is a function \( m : 2^N \rightarrow [0, 1] \) which is monotonic (i.e. \( m(A) \leq m(B) \) whenever \( A \subseteq B \)) 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 \( N = \{1, \ldots, n\} \) 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 \( (w_1, \ldots, w_n) \in [0, 1]^n \) with \( \sum_{i=1}^{n} w_i = 1 \).

- Weighted mean. We assign the following values for the fuzzy measure:
  \[
m(|\{i\}|) = w_i, \quad \text{for } |A| > 1
\]

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

- Exponential cardinality.

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

**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 \( \{1, \ldots, n\} \) such that \( 0 \leq x(1) \leq x(2) \leq \ldots \leq x(n) \) with the convention \( x(0) = 0 \) and \( A(i) = \{A(i), \ldots, A(n)\} \).

2.2. Fuzzy rule-based classification systems

A 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 \( i \)-th component \( (i = 1, 2, \ldots, n) \) 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 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:

\[
\text{Rule } R_j: \text{ If } x_1 = A_{j1} \text{ and } \ldots \text{ and } x_n = A_{jn}, \text{ then Class } = C_j \text{ with } RW_j
\]

(1)
where $R_j$ is the label of the $j$th 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 [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 $x_p = (x_{p1}, \ldots, x_{pn})$ 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 [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.

\[
\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})),
\]

\[j = 1, \ldots, L. \tag{2}\]

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

\[
b_{jk} = \mu_{A_j}(x_p) \times RW_j \quad k = \text{Class}(R_j), \quad j = 1, \ldots, L. \tag{3}\]

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

\[
Y_k = \sum_{j=1}^{L} b_{jk} \quad (j = 1, \ldots, L \text{ and } b_{jk} > 0), \quad k = 1, \ldots, M. \tag{4}\]

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 [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 $x_p = (x_{p1}, \ldots, x_{pn}, C_p)$ 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 [21] as:

\[
RW_j = \frac{\sum_{x_p \in \text{Class}(C_j)} \mu_{A_j}(x_p) - \sum_{x_p \in \text{Class}(C_j)} \mu_{A_j}(x_p)}{\sum_{p=1}^M \mu_{A_j}(x_p)} \tag{6}\]

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 $x_p = (x_{p1}, \ldots, x_{pn})$ 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 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.

   \[
   \mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})),
   \quad j = 1, \ldots, L. \tag{7}
   \]

   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.

   \[
   b^k_j = \mu_{A_j}(x_p) \times RW_j \quad k = \text{Class}(R_j), \quad j = 1, \ldots, L. \tag{8}
   \]

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

   \[
   Y_k = C_{m_k}(b^k_1 j = 1, \ldots, L \text{ and } b^k_j > 0), \quad k = 1, \ldots, M, \tag{9}
   \]

   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.

   \[
   F(Y_1, \ldots, Y_M) = \arg \max_{k=1,\ldots,M} Y_k \tag{10}
   \]

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 \( k = 1, \ldots, M \).

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:

\[
Th = (\#Genes \cdot \text{BITSGENE})/4.0
\]

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.](image-url)

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 \((q_k, \text{ with } k = 1, \ldots, M)\), 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:

  \[
  C_{\text{CHOQUET}} = \{G_1, \ldots, G_M\}.
  \]
where \( G_k \in [0.01, 1.99] \) with \( k = 1, \ldots, M \). In order to compute their real values (in the range \([0.01, 100]\)) we apply Eq. (11).

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

- **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 (the offspring are generated around their parents). Specifically, we consider the PCBLX operator which is described as follows:

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

- \( O_1 = (o_{11} \cdots o_{1n}) \), where \( o_{1i} \) is a randomly (uniformly) chosen number from the interval \([l^1_i, u^1_i]\), with \( l^1_i = \max\{a_i, x_i - I_i\} \), \( u^1_i = \min\{a_i, x_i + I_i\} \), and \( I_i = |x_i - y_i| \).

- \( O_2 = (o_{21} \cdots o_{2n}) \), where \( o_{2i} \) is a randomly (uniformly) chosen number from the interval \([l^2_i, u^2_i]\), with \( l^2_i = \max\{a_i, y_i - I_i\} \) and \( u^2_i = \min\{b_i, y_i + I_i\} \).

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 was to show the usefulness of the 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.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Data-set</th>
<th>#Ex.</th>
<th>#Atts.</th>
<th>#Class.</th>
</tr>
</thead>
<tbody>
<tr>
<td>app</td>
<td>Appendicitis</td>
<td>106</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>aus</td>
<td>Australian</td>
<td>690</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>bal</td>
<td>Balance</td>
<td>625</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>ban</td>
<td>Banana</td>
<td>5300</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>bre</td>
<td>Breast</td>
<td>277</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>bup</td>
<td>Bupa</td>
<td>945</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>car</td>
<td>Car</td>
<td>1728</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>cle</td>
<td>Cleveland</td>
<td>297</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>crx</td>
<td>Crx</td>
<td>653</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>eco</td>
<td>Ecoli</td>
<td>336</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>fla</td>
<td>Flare</td>
<td>1066</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>ger</td>
<td>German</td>
<td>1000</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>gla</td>
<td>Glass</td>
<td>214</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>hab</td>
<td>Haberman</td>
<td>306</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>hay</td>
<td>Hayes-Roth</td>
<td>160</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>hea</td>
<td>Heart</td>
<td>270</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>hou</td>
<td>Housevotes</td>
<td>232</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>ion</td>
<td>Ionosphere</td>
<td>351</td>
<td>33</td>
<td>10</td>
</tr>
<tr>
<td>iri</td>
<td>Iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>led</td>
<td>Led7digit</td>
<td>500</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>mag</td>
<td>Magic</td>
<td>1902</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>mam</td>
<td>Mammographic</td>
<td>830</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>mon</td>
<td>Monk-2</td>
<td>412</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>new</td>
<td>New-Thyroid</td>
<td>215</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>pag</td>
<td>Page-blocks</td>
<td>548</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>pen</td>
<td>Penbased</td>
<td>1992</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>pho</td>
<td>Phoneme</td>
<td>5404</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>pim</td>
<td>Pima</td>
<td>768</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>rin</td>
<td>Ring</td>
<td>7400</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>sat</td>
<td>Satimage</td>
<td>6435</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>seg</td>
<td>Segment</td>
<td>2310</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>shu</td>
<td>Shuttle</td>
<td>5800</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>spe</td>
<td>Spectfheart</td>
<td>267</td>
<td>44</td>
<td>2</td>
</tr>
<tr>
<td>tae</td>
<td>Tae</td>
<td>151</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>tic</td>
<td>Tictactoe</td>
<td>958</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>tit</td>
<td>Titanic</td>
<td>2201</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>two</td>
<td>TwoNorm</td>
<td>740</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>veh</td>
<td>Vehicle</td>
<td>846</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>win</td>
<td>Wine</td>
<td>178</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>wis</td>
<td>Wisconsin</td>
<td>683</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>zoo</td>
<td>Zoo</td>
<td>101</td>
<td>16</td>
<td>7</td>
</tr>
</tbody>
</table>

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 par-
tionate.

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 data-
sset the best results is highlighted in bold-face. The
last column (#FiredRules) shows the average num-
ber of fired rules in each data-set, that is, we com-
pute 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 perfor-
mance 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 ap-
plicated the Friedman aligned ranks test [27] in or-
der to check whether there are statistical differ-
ences 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 out-
performs 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 dif-
ferent 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 condi-
tions there are not statistical differences among the
different proposals.

<table>
<thead>
<tr>
<th>$i$</th>
<th>Algorithm</th>
<th>Hypothesis</th>
<th>APV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Delta_Dirac</td>
<td>Not Rejected</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>WMean</td>
<td>Not Rejected</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>OWA</td>
<td>Not Rejected</td>
<td>0.67</td>
</tr>
<tr>
<td>4</td>
<td>Card.</td>
<td>Not Rejected</td>
<td>0.67</td>
</tr>
</tbody>
</table>

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

Figure 3: Rankings of the approaches tackling
data-sets with an average number of fired rules per
class lower 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 ob-
tained 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 re-
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 statisti-
cally outperformed.

5. Conclusions

In this work we have presented a new FRM in which
the aggregation step is performed by using the Cho-
quet integral. We have considered five fuzzy mea-
sures in order to model the interaction among the
Table 2: Results in testing achieved by the different approaches.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Delta_Dirac</th>
<th>Card.</th>
<th>WMean</th>
<th>OWA</th>
<th>Card_GA</th>
<th>#FiredRules</th>
</tr>
</thead>
<tbody>
<tr>
<td>breast</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>1.00</td>
</tr>
<tr>
<td>car</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>1.00</td>
</tr>
<tr>
<td>flare</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>1.00</td>
</tr>
<tr>
<td>horsevotes</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>1.00</td>
</tr>
<tr>
<td>led7digit</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>1.00</td>
</tr>
<tr>
<td>tic tactoe</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>1.00</td>
</tr>
<tr>
<td>zoo</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>1.00</td>
</tr>
<tr>
<td>german</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>1.00</td>
</tr>
<tr>
<td>hayes-roth</td>
<td>0.66</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
<td>0.67</td>
<td>1.26</td>
</tr>
<tr>
<td>monk-2</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>1.29</td>
</tr>
<tr>
<td>crx</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>1.42</td>
</tr>
<tr>
<td>titanic</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>1.46</td>
</tr>
<tr>
<td>glass</td>
<td>0.58</td>
<td>0.59</td>
<td>0.57</td>
<td>0.60</td>
<td>0.60</td>
<td>1.74</td>
</tr>
<tr>
<td>cleveland</td>
<td>0.55</td>
<td>0.54</td>
<td>0.53</td>
<td>0.54</td>
<td>0.53</td>
<td>1.78</td>
</tr>
<tr>
<td>tae</td>
<td>0.55</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>1.84</td>
</tr>
<tr>
<td>banana</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.62</td>
<td>1.94</td>
</tr>
<tr>
<td>balance</td>
<td>0.91</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.90</td>
<td>2.52</td>
</tr>
<tr>
<td>pageblocks</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>2.76</td>
</tr>
<tr>
<td>shuttle</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>2.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean (FR&lt;3)</th>
<th>0.70</th>
<th>0.69</th>
<th>0.69</th>
<th>0.69</th>
<th>0.70</th>
<th>3.03</th>
</tr>
</thead>
<tbody>
<tr>
<td>haberman</td>
<td>0.72</td>
<td>0.72</td>
<td>0.73</td>
<td>0.72</td>
<td>0.73</td>
<td>3.03</td>
</tr>
<tr>
<td>heart</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>3.07</td>
</tr>
<tr>
<td>iris</td>
<td>0.93</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.95</td>
<td>3.32</td>
</tr>
<tr>
<td>newthyroid</td>
<td>0.85</td>
<td>0.86</td>
<td>0.87</td>
<td>0.87</td>
<td>0.93</td>
<td>3.44</td>
</tr>
<tr>
<td>ecoli</td>
<td>0.72</td>
<td>0.70</td>
<td>0.71</td>
<td>0.71</td>
<td>0.75</td>
<td>3.62</td>
</tr>
<tr>
<td>australian</td>
<td>0.83</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.83</td>
<td>4.10</td>
</tr>
<tr>
<td>mammographic</td>
<td>0.81</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>5.32</td>
</tr>
<tr>
<td>wisconsin</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>5.42</td>
</tr>
<tr>
<td>appendicitis</td>
<td>0.87</td>
<td>0.85</td>
<td>0.87</td>
<td>0.85</td>
<td>0.87</td>
<td>6.54</td>
</tr>
<tr>
<td>ion</td>
<td>0.92</td>
<td>0.92</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
<td>7.26</td>
</tr>
<tr>
<td>satimage</td>
<td>0.47</td>
<td>0.50</td>
<td>0.50</td>
<td>0.51</td>
<td>0.51</td>
<td>7.42</td>
</tr>
<tr>
<td>ring</td>
<td>0.53</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.78</td>
<td>7.81</td>
</tr>
<tr>
<td>phone1me</td>
<td>0.72</td>
<td>0.71</td>
<td>0.72</td>
<td>0.72</td>
<td>0.75</td>
<td>8.67</td>
</tr>
<tr>
<td>segment</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
<td>0.85</td>
<td>8.85</td>
</tr>
<tr>
<td>penbased</td>
<td>0.94</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>10.64</td>
</tr>
<tr>
<td>bupa</td>
<td>0.58</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.63</td>
<td>11.37</td>
</tr>
<tr>
<td>wine</td>
<td>0.93</td>
<td>0.91</td>
<td>0.91</td>
<td>0.90</td>
<td>0.92</td>
<td>11.68</td>
</tr>
<tr>
<td>vehicle</td>
<td>0.61</td>
<td>0.61</td>
<td>0.60</td>
<td>0.60</td>
<td>0.61</td>
<td>12.01</td>
</tr>
<tr>
<td>pima</td>
<td>0.73</td>
<td>0.74</td>
<td>0.75</td>
<td>0.74</td>
<td>0.75</td>
<td>16.70</td>
</tr>
<tr>
<td>magic</td>
<td>0.75</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.78</td>
<td>21.39</td>
</tr>
<tr>
<td>twonorm</td>
<td>0.84</td>
<td>0.79</td>
<td>0.69</td>
<td>0.71</td>
<td>0.88</td>
<td>26.68</td>
</tr>
<tr>
<td>spectfheart</td>
<td>0.63</td>
<td>0.62</td>
<td>0.61</td>
<td>0.62</td>
<td>0.63</td>
<td>46.92</td>
</tr>
<tr>
<td>FR&gt;3</td>
<td>0.77</td>
<td>0.76</td>
<td>0.76</td>
<td>0.75</td>
<td>0.80</td>
<td>69.7</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>i</th>
<th>Algorithm</th>
<th>Hypothesis</th>
<th>APV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OWA</td>
<td>Rejected for Card_GA</td>
<td>5.51E-8</td>
</tr>
<tr>
<td>1</td>
<td>Card.</td>
<td>Rejected for Card_GA</td>
<td>1.06E-6</td>
</tr>
<tr>
<td>1</td>
<td>WMean</td>
<td>Rejected for Card_GA</td>
<td>1.06E-6</td>
</tr>
<tr>
<td>1</td>
<td>Delta_Dirac</td>
<td>Rejected for Card_GA</td>
<td>0.01</td>
</tr>
</tbody>
</table>

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


