Towards a Fuzzy Extension of the López de Mántaras Distance

Eva Armengol[1], Pilar Dellunde[1,2], and Àngel García-Cerdaña[1,3]

[1] Artificial Intelligence Research Institute (IIIA - CSIC),
Campus de la UAB, 08193 Bellaterra, Catalonia, Spain
{eva,pilar,angel}@iiia.csic.es
[2] Departament de Filosofia. Universitat Autònoma de Barcelona.
08193 Bellaterra, Catalonia, Spain
[3] Departament de Lògica, Història i Filosofia de la Ciència,
Universitat de Barcelona, Montalegre 6, 08001 Barcelona, Catalonia, Spain

Abstract. In this paper we introduce FLM, a divergence measure to compare a fuzzy and a crisp partition. This measure is an extension of LM, the López de Mántaras distance. This extension allows to handle domain objects having attributes with continuous values. This means that for some domains the use of fuzzy sets may report better results than the discretization that is the usual way to deal with continuous values. We experimented with both FLM and LM in the context of the lazy learning method called *Lazy Induction of Descriptions* useful for classification tasks.

Keywords: Machine learning, partitions, fuzzy partitions, entropy measures, López de Mántaras distance.

1 Introduction

There are machine learning techniques such as clustering or inductive learning methods, where the comparison of partitions plays an important role. In this paper we introduce FLM, an extension of the López de Mántaras (LM) distance to compare a fuzzy and a crisp partition. The LM distance was first introduced in [1] as a new attribute selection measure for ID3-like inductive algorithms. ID3 [3] is a well-known inductive learning algorithm to induce classification rules in the form of a decision tree. The LM measure is based on a distance between partitions such that the selected attribute in a node induces the partition which is closest to the correct partition of the subset of training examples corresponding to this node.

The advantage of the LM distance, compared with other selection measures such as the Quinlan's gain (see [3]), is that LM is not biased towards selecting attributes whith many values. The LM distance is defined using the measures of information of the different partitions involved. Given two partitions $\mathcal{P}$ and $\mathcal{Q}$ of a set X, the distance between them is computed as follows:

$$LM(\mathcal{P}, \mathcal{Q}) = 2 - \frac{I(\mathcal{P}) + I(\mathcal{Q})}{I(\mathcal{P} \cap \mathcal{Q})}$$

S. Greco et al. (Eds.): IPMU 2012, Part I, CCIS 297, pp. 81–90, 2012.

where $I(\mathcal{P})$ and $I(\mathcal{Q})$ measure the information contained in the partitions $\mathcal{P}$ and $\mathcal{Q}$ respectively and $I(\mathcal{P} \cap \mathcal{Q})$ is the mutual information of the two partitions.

In [4], a paradigm apparatus was introduced for the evaluation of clustering comparison techniques and distinguish between the goodness of clusterings and the similarity of clusterings by clarifying the degree to which different measures confuse the two. This evaluation shows that LM is one of the measures that exhibits the desired behaviour under each of the test scenarios.

In previous works (see for instance [5,6]) we used LM in the framework of a lazy learning method called *Lazy Induction of Descriptions* (LID). LID [2] is a method useful for classification tasks. Due to the characteristics of LM, LID can only deal with domain objects having attributes with nominal values. However knowledge representation of domain objects often involves the use of continuous values. Techniques dealing with continuous values usually use the discretization, consisting on building intervals of values that should be considered as equivalent. There are two kinds of discretization: crisp and fuzzy. In crisp discretization the range of the continuous value is split into several intervals. Elements of an interval are considered as equivalent and each interval is handled as a discrete value. In some domains, the crisp discretization shows some counter-intuitive behavior around the thresholds of the intervals: values around the threshold of two adjacent intervals are considered as different but may be they are not so. For this reason, sometimes is interesting to build a fuzzy discretization from a crisp one, as it is done for instance in [7]. In the context of Case-based Reasoning, the use of fuzzy sets to discretize attributes with continuous values could make the retrieval task more accurate.

The Rand index [8] is a common measure used to compare two clusterings. The Rand index, as it was originally formulated, allows uniquely the evaluation of crisp clustering partitions. In [9], Campello proposed a fuzzy extension of the Rand Index for clustering and classification assessment. This index is defined using basic concepts from fuzzy set theory. Hullermeier-Rifqi [10] introduced another extension of the Rand index suitable for comparing two fuzzy partitions. Since neither in [9] nor in [10] experimental results were conducted, in [11] we experimentally compared the two fuzzy versions of the Rand Index. From these experiments we saw that both measures had a high computational cost. In this context it seems natural to try to introduce an extension of the LM distance for dealing with fuzzy partitions.

In this paper we first introduce a fuzzy extension of the LM distance and we prove some basic properties of this extension. Then we report some experimental results comparing both LM and FLM when used by the LID method as measure to compare partitions.

2 A Fuzzy Version of the López de Mántaras Distance

In this section, first we define a fuzzy extension of the LM distance, that we call FLM. This measure allows to compare a fuzzy partition with respect to a crisp partition. We also prove some basic formal properties of this measure.

Definition 1 (Fuzzy n-partition, normal partition [12]). *Given a finite data set $X = \{x_1, \ldots, x_k\}$ and a positive integer $1 < n < k$, a* fuzzy n-partition *on X is any finite collection $\mathcal{P} = \{P_1, \ldots, P_n\}$ of fuzzy subsets on X such that:*

1) $\sum_{i=1}^{n} P_i(x_h) = 1, \quad 1 \leq h \leq k;$ 2) $0 < \sum_{h=1}^{k} P_i(x_h) < k, \quad 1 \leq i \leq n.$

A fuzzy n-partition on a set X is normal *if and only if for each set $P_i \in \mathcal{P}$, there exists an element $x \in X$ such that $P_i(x) = 1$. This element is called* prototypical *w.r.t. the class P_i.*

The number $\sum_{h=1}^{k} P_i(x_h)$ is the *scalar cardinality* of the fuzzy set P_i and it will be denoted by $|P_i|$.

Definition 2 (Fuzzy LM). *Let $X = \{x_1, \ldots, x_k\}$ be a given a data set, let $\mathcal{P} = \{P_1, \ldots, P_n\}$ be a fuzzy n-partition of X, and $\mathcal{Q} = \{Q_1, \ldots, Q_m\}$ a crisp partition of X. The measure $FLM(\mathcal{P}, \mathcal{Q})$ is computed as follows:*

$$FLM(\mathcal{P}, \mathcal{Q}) = 2 - \frac{I(\mathcal{P}) + I(\mathcal{Q})}{I(\mathcal{P} \cap \mathcal{Q})}, \quad \textit{where:}$$

$$I(\mathcal{P}) = -\sum_{i=1}^{n} p_i log_2 p_i, \textit{ with } p_i = \frac{|P_i|}{k}; \qquad I(\mathcal{Q}) = -\sum_{j=1}^{m} q_j log_2 q_j, \textit{ with } q_j = \frac{|Q_j|}{k};$$

$$I(\mathcal{P} \cap \mathcal{Q}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} log_2 r_{ij}, \textit{ with } r_{ij} = \frac{|P_i \cap Q_j|}{k},$$

where $P_i \cap Q_j : X \to [0, 1]$ is the fuzzy set defined as:

$$(P_i \cap Q_j)(x) = \begin{cases} P_i(x), & \textit{when } x \in Q_j, \\ 0, & \textit{otherwise.} \end{cases}$$

So defined, when $\mathcal{P}$ and $\mathcal{Q}$ are both crisp partitions, $FLM(\mathcal{P}, \mathcal{Q})$ is exactly $LM(\mathcal{P}, \mathcal{Q})$. Let us prove now some formal properties of FLM.

Proposition 1 (Basic facts). *Let X, $\mathcal{P}$, and $\mathcal{Q}$ be as in Definition 2. The following conditions hold ($1 \leq i \leq n$, $1 \leq j \leq m$):*

1) $p_i, q_j \in (0, 1)$, 3) $\sum_{j=1}^{m} r_{ij} = p_i$, 5) $\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} = 1$,
2) $r_{ij} \in [0, 1)$, 4) $\sum_{i=1}^{n} r_{ij} = q_j$, 6) $\sum_{i=1}^{n}\sum_{j=1}^{m} p_i.q_j = 1$.

Proof: 1) and 2) are clear by Definition 1.

3) Let $1 \leq i \leq n$ and $1 \leq h \leq k$. Since $\mathcal{Q}$ is a crisp partition of X, $Q_l(x_h) = 1$ for some equivalence class Q_l of the partition. Then we have:

$$\sum_{j=1}^{m} (P_i \cap Q_j)(x_h) = (P_i \cap Q_l)(x_h) = P_i(x_h) \tag{1}$$

and therefore, by (1) and by definition of r_{ij},

$$\sum_{j=1}^{m} r_{ij} = \frac{1}{k}\sum_{j=1}^{m}\sum_{h=1}^{k}(P_i \cap Q_j)(x_h) = \frac{1}{k}\sum_{h=1}^{k}\sum_{j=1}^{m}(P_i \cap Q_j)(x_h) = \frac{1}{k}\sum_{h=1}^{k} P_i(x_h) = p_i.$$

4) Let $1 \leq j \leq m$ and $1 \leq h \leq k$. Since $\mathcal{Q}$ is a crisp partition of Xwe have:

$$\sum_{i=1}^{n}(P_i \cap Q_j)(x_h) = \begin{cases} \sum_{i=1}^{n} P_i(x_h), & \text{when } Q_j(x_h) = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, since $\mathcal{P}$ is a fuzzy n-partition of X, $\sum_{i=1}^{n} P_i(x_h) = 1$ and thus,

$$\sum_{i=1}^{n}(P_i \cap Q_j)(x_h) = Q_j(x_h) \tag{2}$$

and now, using (2), we obtain:

$$\sum_{i=1}^{n} r_{ij} = \frac{1}{k}\sum_{i=1}^{n}\sum_{h=1}^{k}(P_i \cap Q_j)(x_h) = \frac{1}{k}\sum_{h=1}^{k}\sum_{i=1}^{n}(P_i \cap Q_j)(x_h) = \frac{1}{k}\sum_{h=1}^{k} Q_j(x_h) = q_j.$$

5) $\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} = \sum_{i=1}^{n}(\sum_{j=1}^{m} r_{ij}) = \sum_{i=1}^{n} p_i = 1.$

6) $\sum_{i=1}^{n}\sum_{j=1}^{m} p_i \cdot q_j = \sum_{i=1}^{n}(p_i \sum_{j=1}^{m} q_j) = \sum_{i=1}^{n} p_i \cdot 1 = \sum_{i=1}^{n} p_i = 1.$ □

Proposition 2. *Given a fuzzy n-partition $\mathcal{P}$ and a crisp m-partition $\mathcal{Q}$ on a finite set $X = \{x_1, \ldots, x_k\}$, it holds that $LM(\mathcal{P}, \mathcal{Q}) \in [0, 1]$.*

Proof: First, let us see that $I(\mathcal{P} \cap \mathcal{Q}) \geq I(\mathcal{P})$. By item 3) of Proposition 1, for every $1 \leq i \leq n$ and $1 \leq j \leq m$, $r_{ij} \leq p_i$, and since the logarithm function is increasing, we have that $\log r_{ij} \leq \log p_i$. Therefore,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} log_2 r_{ij} \leq \sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} log_2 p_i = \sum_{i=1}^{n}(\sum_{j=1}^{m} r_{ij}) log_2 p_i = \sum_{i=1}^{n} p_i log_2 p_i.$$

Consequently, $I(\mathcal{P} \cap \mathcal{Q}) \geq I(\mathcal{P})$. Secondly, we show that $I(\mathcal{P} \cap \mathcal{Q}) \geq I(\mathcal{Q})$. By item 4) of Proposition 1 we have that, for every $1 \leq i \leq n$ and $1 \leq j \leq m$, $r_{ij} \leq q_j$, and thus $\log r_{ij} \leq \log q_j$. Therefore,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} log_2 r_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} r_{ij} log_2 r_{ij} \leq \sum_{j=1}^{m}\sum_{i=1}^{n} r_{ij} log_2 q_j =$$

$$= \sum_{j=1}^{m}(\sum_{i=1}^{n} r_{ij}) log_2 q_j = \sum_{i=1}^{n} q_j log_2 q_j.$$

Consequently, we also have $I(\mathcal{P} \cap \mathcal{Q}) \geq I(\mathcal{Q})$. Thus, $2 \cdot I(\mathcal{P} \cap \mathcal{Q}) \geq I(\mathcal{P}) + I(\mathcal{Q})$, and then:

$$LM(\mathcal{P}, \mathcal{Q}) = 2 - \frac{I(\mathcal{P}) + I(\mathcal{Q})}{I(\mathcal{P} \cap \mathcal{Q})} \geq 0.$$

Finally, we need to prove that

$$\frac{I(\mathcal{P}) + I(\mathcal{Q})}{I(\mathcal{P} \cap \mathcal{Q})} \geq 1.$$

It will be sufficient to prove that $I(\mathcal{P} \cap \mathcal{Q}) \leq I(\mathcal{P}) + I(\mathcal{Q})$. Indeed, by using the definitions, items 3) and 4) of Proposition 1, and some properties of the logarithm function, we have:

$$I(\mathcal{P} \cap \mathcal{Q}) - I(\mathcal{P}) - I(\mathcal{Q}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} log_2 r_{ij} + \sum_{i=1}^{n} p_i log_2 p_i + \sum_{j=1}^{m} q_j log_2 q_j =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} log_2 \frac{1}{r_{ij}} + \sum_{i=1}^{n}(\sum_{j=1}^{m} r_{ij}) log_2 p_i + \sum_{j=1}^{m}(\sum_{i=1}^{n} r_{ij}) log_2 q_j = \sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} log_2 \frac{p_i q_j}{r_{ij}}.$$

Now we use the well known fact that $\ln x \leq x - 1$. For base 2 we have $log_2 x = \frac{\ln x}{\ln 2} \leq \frac{x-1}{\ln 2}$. Now, using this fact and items 5) and 6) of Proposition 1, we have:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} log_2 \frac{p_i \cdot q_j}{r_{ij}} \leq \frac{1}{\ln 2} \cdot \sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} (\frac{p_i \cdot q_j}{r_{ij}} - 1) = \frac{1}{\ln 2} \cdot \sum_{i=1}^{n}\sum_{j=1}^{m} (p_i \cdot q_j - r_{ij}) =$$

$$= \frac{1}{\ln 2} (\sum_{i=1}^{n}\sum_{j=1}^{m} p_i \cdot q_j - \sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij}) = \frac{1}{\ln 2} \cdot (1 - 1) = \frac{1}{\ln 2} \cdot 0 = 0.$$

□

3 Experiments

The experimentation with the extended version of the LM distance has been carried out by including it into a lazy learning method called *Lazy Induction of Descriptions* (LID in short). In this section we explain LID in some detail and then we report the experiments and the results obtained with both crisp and fuzzy versions of LM.

LID is a lazy learning method for classification tasks. LID determines which are the most relevant attributes of a problem (i.e., a case to be classified) and searches in a case base for cases sharing these relevant attributes. The problem is classified when LID finds a set of relevant attributes shared by a subset of cases all of them belonging to the same class. We call the description formed by these relevant features *similitude term* and the set of cases satisfying the similitude term *discriminatory set*.

```
Function LID (p, Di, SDi, C)
 if stopping-condition (SDi) then return class (SDi)
    else  fd := Select-attribute (p, SDi, C)
          Di+1 := Add-attribute (fd, Di)
          SDi+1 := Discriminatory-set (Di+1, SDi)
          LID (p, Di+1, SDi+1, C)
 end-if
end-function
```

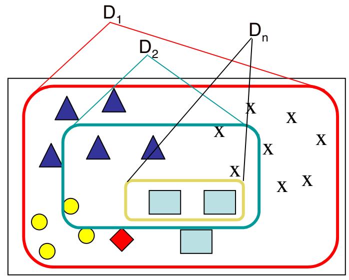

Fig. 1. The LID algorithm. On the right there is the intuitive idea of LID.

Given a problem for solving p, the LID algorithm (Fig. 1) initializes D_0 as a description with no attributes, the discriminatory set S_{D_0}, as the set of cases satisfying D_0, i.e., all the available cases, and C as the set of solution classes into which the known cases are classified. Let D_i be the current similitude term and S_{D_i} be the set of all the cases satisfying D_i. When the stopping condition of LID is not satisfied, the next step is to select an attribute for specializing D_i. The specialization of D_i is achieved by adding attributes to it. Given a set F of attributes candidate to specialize D_i, the next step of the algorithm is the selection of an attribute $f \in F$. Selecting the most discriminatory attribute in F is heuristically done using a measure Δ to compare each partition $\mathcal{P}_f$ induced by an attribute f with the correct partition $\mathcal{P}_c$. The *correct partition* is the one having as many sets as solution classes. Each attribute $f \in F$ induces in the discriminatory set a partition $\mathcal{P}_f$ with as many sets as the number of different values that f takes in the cases.

Given a measure Δ and two attributes f and g inducing respectively partitions $\mathcal{P}_f$ and $\mathcal{P}_g$, we say that f is *more discriminatory* than g iff $\Delta(\mathcal{P}_f, \mathcal{P}_c) < \Delta(\mathcal{P}_g, \mathcal{P}_c)$. This means that the partition $\mathcal{P}_f$ is closer to the correct partition than the partition $\mathcal{P}_g$. LID selects the most discriminatory attribute to specialize D_i. Let f_d be the most discriminatory attribute in F. The specialization of D_i defines a new similitude term D_{i+1} by adding to D_i the attribute f_d. The new similitude term $D_{i+1} = D_i \cup \{f_d\}$ is satisfied by a subset of cases in S_{D_i}, namely $S_{D_{i+1}}$. Next, LID is recursively called with $S_{D_{i+1}}$ and D_{i+1}. The recursive call of LID has $S_{D_{i+1}}$ instead of S_{D_i} because the cases that are not satisfied by D_{i+1} will not satisfy any further specialization. Notice that the specialization reduces the discriminatory set at each step, i.e., we get a sequence $S_{D_n} \subset S_{D_{n-1}} \subset \ldots \subset S_{D_0}$. LID has two stopping situations: 1) all the cases in the discriminatory set S_{D_j} belong to the same solution class C_i, or 2) there is no attribute allowing the specialization of the similitude term. When the stopping condition 1) is satisfied, p is classified as belonging to C_i. When the stopping condition 2) is satisfied, S_{D_j} contains cases from several classes; in such situation the *majority criteria* is applied, and p is classified in the class of the majority of cases in S_{D_j}.

```
(define (object :id OBJ-50)
   (Sepallength 7.0)
   (Sepalwidth 3.2)
   (Petallength 4.7)
   (Petalwidth 1.4))
```

```
(define (object :id OBJ-50)
   (Sepallength (define (fuzzy-value)
                  (Value 7.0)
                  (Membership 0 0 1)))
   (Sepalwidth (define (fuzzy-value)
                  (Value 3.2)
                  (Membership 0 1 0)))
   (Petallength (define (fuzzy-value)
                  (Value 4.7)
                  (Membership 0 0.6087 0.3913)))
   (Petalwidth (define (fuzzy-value)
                  (Value 1.4)
                  (Membership  0 1 0))))
```

Fig. 2. On the left there is a propositional representation of an object. On the right there is the representation of the same object extended with the membership vector.

Conditions of the Experiments. We have conducted several experiments on data sets coming from the UCI Repository [13] using LID with LM and FLM as the Δ measure. We have used the following data sets: *iris, bal, heart-statlog, glass, wdbc, glass,* and *thyroids*. For the evaluation we have taken the discretization intervals provided by Weka [14]. Thus, for instance, for the *Iris* data set, Weka gets the following intervals:

– Attribute Petalwidth: $[0.00, 0.80], (0.80, 1.75], (1.75, 2.25]$
– Attribute Petallength: $[1.00, 2.45], (2.45, 4.75], (4.75, 6.90]$
– Attribute Sepalwidth: $[2.20, 2.95], (2.95, 3.35], (3.35, 4.40]$
– Attribute Sepallength: $[4.40, 5.55], (5.55, 6.15], (6.15, 7.90]$

These intervals have been directly used by the LM distance. When using the FLM measure we define fuzzy sets. Firstly, we will explain how to represent the fuzzy cases handled by fuzzy LID. The left of Fig. 2 shows an example of an object from the *Iris* data set represented as a set of pairs attribute-value. The right of Fig. 2 shows the fuzzy representation of the same object. Notice that the value of each attribute is an object that has in turn two attributes: Value and Membership. The attribute Value takes the same value v that in the crisp version (for instance, 7.0 in the attribute Sepallength). The attribute Membership takes as value the *membership vector* associated to v, that is, a n-tuple μ, being n the number of fuzzy sets associated to the continuous range of an attribute. Each position i of μ represents the membership of the value v to the corresponding fuzzy set F_i. In the next we will explain how to compute the membership vector.

Given an attribute taking continuous values, let us suppose that the domain expert has given $\alpha_1, \ldots, \alpha_n$ as the thresholds determining the discretization intervals for that attribute. Let α_0 and α_{n+1} be the minimum and maximum respectively of the values that this attribute takes in its range. To each one of the $n+1$ intervals $[\alpha_0, \alpha_1], (\alpha_1, \alpha_2], \ldots, (\alpha_n, \alpha_{n+1}]$ corresponds a trapezoidal fuzzy set defined as follows, where $1 < i < n+1$:

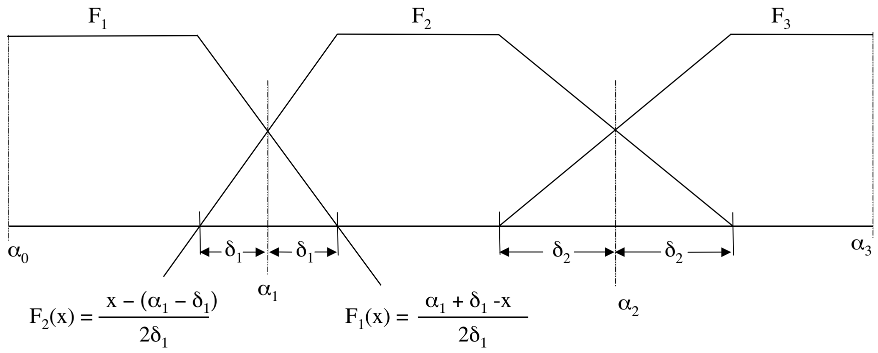

Fig. 3. Trapezoidal fuzzy sets. The values α_1 and α_2 are given by the domain expert as the thresholds of the discretization intervals for a given attribute.

$$F_1(x) = \begin{cases} 1 & \text{when } \alpha_0 \leq x \leq \alpha_1 - \delta_1 \\ \frac{\alpha_1+\delta_1-x}{2\delta_1} & \text{when } \alpha_1 - \delta_1 < x < \alpha_1 + \delta_1 \\ 0 & \text{when } \alpha_1 + \delta_1 \leq x \end{cases}$$

$$F_i(x) = \begin{cases} 0 & \text{when } x \leq \alpha_{i-1} - \delta_{i-1} \\ \frac{x-(\alpha_{i-1}-\delta_{i-1})}{2\delta_{i-1}} & \text{when } \alpha_{i-1} - \delta_{i-1} < x < \alpha_{i-1} + \delta_{i-1} \\ 1 & \text{when } \alpha_{i-1} + \delta_{i-1} \leq x \leq \alpha_i - \delta_i \\ \frac{\alpha_i+\delta_i-x}{2\delta_i} & \text{when } \alpha_i - \delta_i < x < \alpha_i + \delta_i \\ 0 & \text{when } \alpha_i + \delta_i \leq x \end{cases}$$

$$F_{n+1}(x) = \begin{cases} 0 & \text{when } x \leq \alpha_n - \delta_n \\ \frac{x-(\alpha_n-\delta_n)}{2\delta_n} & \text{when } \alpha_n - \delta_n < x < \alpha_n + \delta_n \\ 1 & \text{when } \alpha_n + \delta_n \leq x \leq \alpha_{n+1} \end{cases}$$

The parameters δ_i are computed as follows: $\delta_i = p \cdot |\alpha_i - \alpha_{i-1}|$, where the factor p corresponds to a percentage that we can adjust. Figure 3 shows the trapezoidal fuzzy sets defined when $n = 2$. For instance, for the *Iris* data set the values of α_i for the Petallength attribute are: $\alpha_0 = 1$, $\alpha_1 = 2.45$, $\alpha_2 = 4.75$, $\alpha_3 = 6.9$. The value 4.7 taken by the object *obj-50* in the attribute Petallength (Fig. 2) has associated the membership vector $(0, 0.6087, 0.3913)$, meaning that such value belongs to a degree 0 to the fuzzy set F_1 corresponding to the interval $[1, 2.45]$, to a degree 0.6087 to the fuzzy set F_2 corresponding to $(2.45, 4.75]$, and to a degree 0.3913 to the fuzzy set F_3 corresponding to $(4.75, 6.9]$.

In the fuzzy version of LID, the correct partition is the same than in the crisp case since each object belongs to a unique solution class. However, when the partitions induced by each attribute are fuzzy, an object can belong (to a certain degree) to more than one partition set. Thus the algorithm of the fuzzy LID is the same explained before but using the particular representation for the fuzzy cases and FLM as the Δ measure. In the fuzzy experiments, to calculate the values δ_i we have experimented with $p = 0.05$ and 0.10.

Table 1. The left part shows the percentage of correct classifications of LID using LM and FLM. The right part shows the percentage of incorrect classifications of LID using LM and FLM. Results are the mean of 7 trials of 10-fold cross-validation and they correspond to $p = 0.10$.

Dataset	LM	FLM	significant	LM	FLM	significant
bal	**70.8387**	66.6465	yes	28.9769	**25.5450**	yes
glass	**78.2703**	63.3519	yes	**21.7297**	34.5825	yes
heart-statlog	66.5608	**76.0317**	yes	33.4381	**20.0529**	yes
iris	93.8155	**95.7143**	yes	6.1845	**3.8095**	yes
thyroids	95.4660	94.3692	no	**4.5340**	4.8268	yes

Results. Table 1 shows the results of LID after seven trials of 10-fold cross-validation taking $p = 0.10$. Experiments show that the fuzzy version of LID gives good predictive results and in some domains (*heart statlog* and *iris*) outperforms the crisp version. LID can produce two kinds of outputs: the classification in one (correct or incorrect) class or a multiple classification. Multiple classification means that LID has not been capable to classify the input object in only one class. The utility of a multiple classification depends on the application domain, so it is the expert who decides what is better to force the method to give a classification (even incorrect) or to accept a 'no classification". The percentage of correct classifications is similar in $p = 0.05$ and in $p = 0.10$ but with $p = 0.10$, LID gives lower percentage of incorrect solutions and also a higher percentage of multiple solutions than with $p = 0.05$.

4 Conclusions and Future Work

So far we have defined a fuzzy version of the LM distance, called FLM, in order to compare a fuzzy and a crisp partition. Further research will be devoted to explore different definitions based in different t-norms to extend the LM distance for comparing two fuzzy partitions. In this paper we have proved only some basic facts about the FLM measure, a systematic study of its formal properties is needed and it will be our immediate research objective.

In [15] the notion of "measure of the degree of fuzziness" or "entropy" of a fuzzy set was introduced using no probabilistic concepts. Based on this definition, some classes of divergence measures between fuzzy partitions were presented in [16]. Since the LM distance is an information theoretic approach to the comparison of crisp partitions, it could be interesting to study the relationship of our fuzzy measure with all these divergence measures. In the future we would also like to conduct more experiments to compare the Rand index and its two fuzzy extensions introduced in [9] and [10] with the LM distance and the measure FLM.

Acknowledgments. This research is partially funded by the Spanish MICINN projects Next-CBR (TIN 2009-13692-C03-01), MTM2011-25747, ARINF

(TIN2009-14704-C03-03), TASSAT TIN2010-20967-C04-01, and CONSOLIDER (CSD2007-0022), the European project ESF Eurocores-LogICCC/MICINN (FFI2008-03126-E/FILO), and the grants 2009-SGR-1433 and 2009-SGR-1434 from the Generalitat de Catalunya.

References

1. López de Mántaras, R.: A distance-based attribute selection measure for decision tree induction. Machine Learning 6, 81–92 (1991)
2. Armengol, E., Plaza, E.: Lazy Induction of Descriptions for Relational Case-Based Learning. In: Flach, P.A., De Raedt, L. (eds.) ECML 2001. LNCS (LNAI), vol. 2167, pp. 13–24. Springer, Heidelberg (2001)
3. Quinlan, J.R.: Induction of decision trees. Machine Learning 1, 81–106 (1986)
4. Pfitzner, D., Leibbrandt, R., Powers, D.M.W.: Characterization and evaluation of similarity measures for pairs of clusterings. Knowledge Information Systems 19(3), 361–394 (2009)
5. Armengol, E.: Discovering plausible explanations of carcinogenecity in chemical compounds. In: Perner, P. (ed.) MLDM 2007. LNCS (LNAI), vol. 4571, pp. 756–769. Springer, Heidelberg (2007)
6. Armengol, E., Puig, S.: Combining two lazy learning methods for classification and knowledge discovery. a case study for malignant melanoma diagnosis. In: Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, pp. 200–207 (2011)
7. Kuwajima, I., Nojima, Y., Ishibuchi, H.: Effects of constructing fuzzy discretization from crisp discretization for rule-based classifiers. Artificial Life and Robotics 13(1), 294–297 (2008)
8. Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66(336), 846–850 (1971)
9. Campello, R.J.G.B.: A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment. Pattern Recognition Letters 28(7), 833–841 (2007)
10. Hüllermeier, E., Rifqi, M.: A fuzzy variant of the Rand index for comparing clustering structures. In: Proceedings of IFSA/EUSFLAT Conference, pp. 1294–1298 (2009)
11. Armengol, E., García-Cerdaña, À.: Lazy Induction of Descriptions Using Two Fuzzy Versions of the Rand Index. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010, Part I. CCIS, vol. 80, pp. 396–405. Springer, Heidelberg (2010)
12. Zimmermann, H.: Fuzzy Set Theory and its applications, 2nd edn. Kluver Academic Publishers (1992)
13. Asuncion, A., Newman, D.J.: UCI machine learning repository (2007)
14. Witten, I., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S.: Weka: Practical machine learning tools and techniques with java implementations (1999)
15. de Luca, A., Termini, S.: A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. Information and Control 20(4), 301–312 (1972)
16. Montes, S., Couso, I., Gil, P., Bertoluzza, C.: Divergence measure between fuzzy sets. International Journal Approximate Reasoning 30(2), 91–105 (2002)