Lazy Induction of Descriptions using two fuzzy versions of the Rand Index

Eva Armengol and Àngel García-Cerdaña

IIIA - Artificial Intelligence Research Institute,
CSIC - Spanish Council for Scientific Research,
Campus UAB, 08193 Bellaterra, Catalonia (Spain)
email: eva,angel@iiia.csic.es

Abstract. In this paper we introduce an extension of the lazy learning method called *Lazy Induction of Descriptions* (LID). This new version is able to deal with fuzzy cases, i.e., cases described by attributes taking continuous values represented as fuzzy sets. LID classifies new cases based on the relevance of the attributes describing them. This relevance is assessed using a distance measure that compares the correct partition (i.e., the correct classification of cases) with the partitions induced by each one of the attributes. The fuzzy version of LID introduced in this paper uses two fuzzy versions of the Rand index to compare fuzzy partitions: one proposed by Campello and another proposed by Hüllermeier and Rifqi. We experimented with both indexes on data sets from the UCI machine learning repository.

1 Introduction

Case-based reasoning (CBR) is based on the idea that similar problems (cases) have similar solutions. Given a problem to be solved the first step of a CBR method [1] is to retrieve a subset of cases assessed as the most similar to the problem. Depending on the similarity criteria, the subset of retrieved cases will be different and, thus, the solution of the new problem will also be different. Notice that, differently than inductive learning methods (e.g., decision trees), CBR methods are *lazy* in the sense that the problem solving process depends on each new problem. *Lazy Induction of Descriptions* (LID) [3] is a lazy learning method useful for classification tasks. LID retrieves precedents based on the relevance of attributes. This relevance is assessed using a distance measure that compares the correct partition (i.e., the correct classification of cases) with the partitions induced by each one of the attributes.

Although LID is able to deal with relational objects represented as *feature terms* [2], we take here a version of LID that handles objects (cases) represented using *propositional representation*, that is, as a set of pairs attribute-value, where the values are nominal (i.e., they take values in a finite set of values). However, sometimes this representation is not appropriate (for instance to represent people weight, age or some physical measures) being common the necessity to give continuous values to attributes. There are a lot of approaches dealing with attributes taking continuous values. Some of these approaches discretize the continuous values and then they use the usual similarity measures on the discretized values [9, 11]. By means of the discretization, continuous values can be handled as nominal. However in this procedure there is a lost of information since the values near to the thresholds of the discretization interval are considered equal but, in fact, they are not. With the goal of reducing such lost of information, other approaches use fuzzy sets to deal with continuous values (see for instance [12]). Previous versions of LID handle cases with attributes having nominal values and, when cases have attributes taking continuous values, they are previously discretized. The distance measure to compare partitions, denoted here by LM, is the one introduced by López de Mántaras in [7]. In this paper we want to analyze the performance of LID when the continuous values of attributes are represented using fuzzy sets. Since the distance LM is not appropriate for this task, it must be replaced by some other measure able to deal with fuzzy partitions.

In this paper we use two fuzzy versions of the Rand index [10]: one proposed by Campello [5], which can compare a fuzzy partition with a crisp one, and another one proposed by Hüllermeier and Rifqi [8], which can compare two fuzzy partitions. In Section 2 we give a brief introduction of LID and the Rand index. In Section 3 we explain the fuzzy version of LID. In Section 4 we show the results of the experiments with fuzzy LID.

2 Lazy Induction of Descriptions

Lazy Induction of Descriptions (LID) is a lazy learning method for classification tasks. LID determines which are the most relevant attributes of a problem 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 *similitude term* the description formed by these relevant features and *discriminatory set* the set of cases satisfying the similitude term.

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 distance (the LM distance in [3]). Such distance is used to compare each partition $\mathcal{P}_f$ induced by an attribute f with the correct partition $\mathcal{P}_c$. The *correct partition* has 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 distance

```
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
```

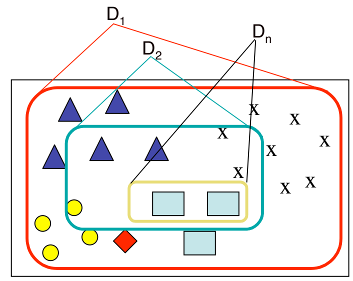

Fig. 1. The LID algorithm. On the right there is the intuitive idea of LID.

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}.

Now let us explain how to select the most discriminant attribute using the Rand index [10]. This index is used to compare clusterings being both classical partitions and it takes as basic unit of comparison the way in which two objects are clustered. The situation in which two objects are placed either together in the same cluster in both clusterings, or placed in different clusters in both clusterings, represents a similarity between the clusterings. Conversely, the situation in which two objects are in the same cluster in one clustering and in different clusters in the other, shows a dissimilarity between both clusterings. The Rand index assesses the similarity between clusterings based on the number of equal assignments of pairs of objects normalized by the total number of pairs. Inside LID, the Rand index is used to compare the partitions induced by each one of the attributes describing the objects with the correct partition. Let $X = \{x_1, \ldots, x_n\}$ be a finite set of objects, and let $\mathcal{P} = \{P_1, \ldots, P_k\}$ and $\mathcal{Q} = \{Q_1, \ldots, Q_h\}$ be two partitions of X in k and h sets, respectively. Given two objects x and x'

```
(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.

we say that both objects are *paired* in a partition when both objects belong to the same set of the partition. Otherwise, we say that both objects are *impaired*. Now let us consider the set $C := \{(x_i, x_j) \in X \times X : 1 \leq i < j \leq n\}$, which can be identified with the set of unordered pairs $\{x, y\}$, with $x, y \in X$. The Rand index between the partitions $\mathcal{P}$ and $\mathcal{Q}$ is defined as follows:

$$R(\mathcal{P}, \mathcal{Q}) = \frac{a + d}{a + b + c + d} \tag{1}$$

where $a = |\{(x, x') \in C : x \text{ and } x' \text{ paired in } \mathcal{P} \text{ and paired in } \mathcal{Q}\}|$,
$b = |\{(x, x') \in C : x \text{ and } x' \text{ paired in } \mathcal{P} \text{ and impaired in } \mathcal{Q}\}|$,
$c = |\{(x, x') \in C : x \text{ and } x' \text{ impaired in } \mathcal{P} \text{ and paired in } \mathcal{Q}\}|$,
$d = |\{(x, x') \in C : x \text{ and } x' \text{ impaired in } \mathcal{P} \text{ and impaired in } \mathcal{Q}\}|$.

Notice that in fact the Rand index gives a measure of the similarity between two partitions. Therefore we say that the attribute f inducing $\mathcal{P}_f$ is *more discriminatory* than the attribute g inducing $\mathcal{P}_g$ iff $1 - R(\mathcal{P}_f, \mathcal{P}_c) < 1 - R(\mathcal{P}_g, \mathcal{P}_c)$.

3 A fuzzy version of LID

In this section we explain a fuzzy version of LID using two fuzzifications of the Rand index: the one defined by Campello [5] and another one defined by Hüllermeier and Rifqi [8]. 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 so-called *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. For each one of the $n+1$ intervals $[\alpha_0, \alpha_1], \ldots, [\alpha_n, \alpha_{n+1}]$ corresponds a trapezoidal fuzzy set defined as follows, where $1 < i < n+1$:

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

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

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

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

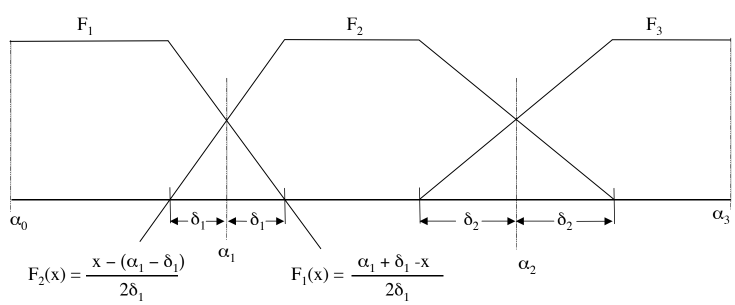

Fig. 3. Trapezoidal fuzzy sets. The values α_1 and α_2 are given by the domain expert as the thresholds of the discretization intervals for a given attribute.

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 in Section 2 but using the particular representation for the fuzzy cases and replacing the Rand index by one of its fuzzy versions.

3.1 The Campello's fuzzy Rand index

In [5] Campello extends the Rand index to make it feasible to compare fuzzy partitions. To this end, he first rewrite the original formulation of the Rand index in an equivalent form by using basic concepts from set theory. Given the partitions $\mathcal{P}$ and $\mathcal{Q}$ of a set of objects X, and the set C of pairs of elements in X defined in Sec. 2, Campello defines the following subsets of C:

$V = \{(x, x') : x \text{ and } x' \text{ paired in } \mathcal{P}\}$, $W = \{(x, x') : x \text{ and } x' \text{ impaired in } \mathcal{P}\}$,
$Y = \{(x, x') : x \text{ and } x' \text{ paired in } \mathcal{Q}\}$, $Z = \{(x, x') : x \text{ and } x' \text{ impaired in } \mathcal{Q}\}$.

According to these definitions, the coefficients in Eq. (1) can be rewritten as follows: $a = |V \cap Y|$, $b = |V \cap Z|$, $c = |W \cap Y|$, $d = |W \cap Z|$. When we consider fuzzy partitions, the sets above are fuzzy sets. Let $P_i(x) \in [0, 1]$ be the degree of membership of the object $x \in X$ to the set P_i. Campello defines the fuzzy binary relations V, W, Y and Z on the set C by using the following expressions involving a t-norm $\otimes$ and a t-conorm $\oplus$:

$$V(x, x') = \bigoplus_{i=1}^{k}(P_i(x) \otimes P_i(x')), \quad W(x, x') = \bigoplus_{1 \le i \ne j \le k}(P_i(x) \otimes P_j(x')),$$
$$Y(x, x') = \bigoplus_{i=1}^{h}(Q_i(x) \otimes Q_i(x')), \quad Z(x, x') = \bigoplus_{1 \le i \ne j \le h}(Q_i(x) \otimes Q_j(x')).$$

As it is usually done, Campello takes the intersection of fuzzy binary relations as the t-norm of the membership degrees of the pairs, and he uses the sigma-count principle for defining the fuzzy set cardinality (see [6]). Thus, the coefficients a, b, c, d are obtained as follows:

$$a = |V \bigcap Y| = \sum_{(x,x') \in C}(V(x, x') \otimes Y(x, x'))$$
$$b = |V \bigcap Z| = \sum_{(x,x') \in C}(V(x, x') \otimes Z(x, x'))$$
$$c = |W \bigcap Y| = \sum_{(x,x') \in C}(W(x, x') \otimes Y(x, x'))$$
$$d = |W \bigcap Z| = \sum_{(x,x') \in C}(W(x, x') \otimes Z(x, x'))$$

Then, the fuzzy version of the Rand index is also defined by the equation (1) giving a measure of the similarity between two partitions. Since LID uses a normalized distance measure, we have to take $1 - R(\mathcal{P}, \mathcal{Q})$. The Campello's fuzzy formulation of the Rand index is appropriated to compare a crisp partition with a fuzzy partition. Notice that the correct partition in classification problems is commonly crisp, therefore the use of the distance associated to the Rand index of Campello inside LID is justified. We will denote as CI such distance.

3.2 The Hüllermeier-Rifqi's fuzzy Rand index

When CI is used to compare two fuzzy partitions, it presents an important problem since the property of reflexivity is not satisfied. For this reason Hüllermeier and Rifqi proposed in [8] a different fuzzy version for the Rand index which allows the comparison of two fuzzy partitions and that satisfies all the desirable metric properties. Let us recall their definition.

Given a fuzzy partition $\mathcal{P} = \{P_1, P_2, \ldots, P_k\}$, each object x is characterized by its membership vector $\mathcal{P}(x) = (P_1(x), P_2(x), ..., P_k(x)) \in [0,1]^k$ where $P_i(x)$ is the membership degree of x to the cluster P_i. Given two objects x and x' and two fuzzy partitions $\mathcal{P}$ and $\mathcal{Q}$, the *degree of concordance* of both objects in these partitions is defined by means the expression $1-|E_{\mathcal{P}}(x,x')-E_{\mathcal{Q}}(x,x')|$ where $E_{\mathcal{P}}$ is the fuzzy equivalence relation on X defined by $E_{\mathcal{P}}(x,x') := 1-\|\mathcal{P}(x)-\mathcal{P}(x')\|$ being $\|.\|$ a distance on $[0,1]^k$ yielding values in $[0,1]$. Thus, two objects are equivalent to a degree 1 when both have the same membership degrees in all the sets of the partition. This fuzzy equivalence is used to define the notion of *concordance* as a fuzzy binary relation, which generalizes the crisp binary relation (induced by a crisp partition) defined on the set C of unordered pairs of objects of X using the notions of *paired* and *unpaired*. Then, a distance measure on fuzzy partitions using the *degree of discordance* is defined as $|E_{\mathcal{P}}(x,x') - E_{\mathcal{Q}}(x,x')|$. Thus given a data set X of n elements, and two fuzzy partitions $\mathcal{P}$ and $\mathcal{Q}$ on X, the distance between both partitions is the normalized sum of degrees of discordance:

$$d(\mathcal{P},\mathcal{Q}) = \frac{\sum_{(x,x')\in C} |E_{\mathcal{P}}(x,x') - E_{\mathcal{Q}}(x,x')|}{n(n-1)/2}. \tag{2}$$

Since the Rand index measures similarity, by using the expression $1-d(\mathcal{P},\mathcal{Q})$ we can asses the similarity of two fuzzy partitions $\mathcal{P}$ and $\mathcal{Q}$. In [8] the authors prove that this similarity is a generalization of the Rand index, and they prove also that the distance (2) is a *pseudometric*, i.e., it satisfies the properties of *reflexivity*, *symmetry*, and the *triangular inequality*. Let us recall that a fuzzy partition $\mathcal{P} = \{P_1, \ldots, P_k\}$ is called *normal* if *a*) for each $x \in X$, $P_1(x) + \cdots + P_k(x) = 1$, and *b*) it has a prototypical element, i.e., for every $P_i \in \mathcal{P}$, there exists an $x \in X$ such that $P_i(x) = 1$. In their paper Hüllermeier and Rifqi also show that for normal partitions, and taking the equivalence relation on X defined by

$$E_{\mathcal{P}}(x,x') = 1 - \frac{1}{2}\sum_{i=1}^{k} |P_i(x) - P_i(x')|, \tag{3}$$

the distance defined by the equation (2) is a *metric*, i.e., it also satisfies the property of *separation* ($d(\mathcal{P},\mathcal{Q}) = 0$ implies $\mathcal{P} = \mathcal{Q}$). We have taken this metric as measure of the distance in our experiments. From now on, we will call HR the distance proposed by Hüllermeier and Rifqi using (3).

4 Experiments

We conducted several experiments on four data sets coming from the UCI Repository [4] using the fuzzy versions of the Rand index inside LID. We used four data sets: *iris, heart-statlog, glass* and *thyroids*. For the evaluation of the crisp Rand index we taken the discretization intervals provided by Weka [13], and the same thresholds have been used for defining fuzzy sets. Thus, for instance, for the *Iris* data set, Weka gets the following intervals:

– Attribute Petalwidth: $(\infty, 0.8], (0.8, 1.75], (1.75, \infty)$
– Attribute Petallength: $(\infty, 2.45], (2.45, 4.75], (4.75, \infty)$
– Attribute Sepalwidth: $(\infty, 2.95], (2.95, 3.35], (3.35, \infty)$
– Attribute Sepallength: $(\infty, 5.55], (5.55, 6.15], (6.15, \infty)$

We performed three kinds of experiments: 1) using the crisp Rand index (with the discretization proposed by Weka); 2) using the fuzzyfication proposed by Campello (CI); and 3) using the fuzzyfication proposed by Hüllermeier-Rifqi (HR). The experiments with the crisp Rand index are considered as the baseline of the LID performance. In the fuzzy experiments, to calculate the values δ_i (see Sec. 3) we experimented with $p = 0.05, 0.10, 0.15, 0.20$. Moreover when using the Campello's fuzzyfication we also need to choose a t-norm and a t-conorm. In our experiments we taken the Minimum and the Maximum, respectively.

Table 1 shows the results of LID after seven trials of 10-fold cross-validation. For each index, there are three columns C, I and M corresponding respectively to the percentage of correct, incorrect and multiple answers. LID produces multiple answers when the last similitude term cannot be further specialized and the cases included in its associated discriminatory set belong to several solution classes. In such situation, LID is not able to classify the new problem and, depending on the domain, this can be interpreted as *no solution*. For this reason we counted them separately. We chosen to show the results obtained taking $p = 0.15$ since this is the value producing the least percentage of incorrect classifications. Results obtained with the values 0.05 and 0.10 are not significantly different from those with $p = 0.15$. Worst results are those obtained with $p = 0.20$. The parameter p is a measure of the overlapping degree between two fuzzy sets. In our experiments, the error percentage is not largely influenced by this degree.

It is difficult to extract a clear conclusion about which is the best method since none of them is better than others in all the domains, however the fuzzy versions of LID seems to be better than the crisp version. Our interpretation of this is that the use of fuzzy sets probably supports a more finest classification since, compared with the crisp version, the use of both CR and HR produce a lower percentage of both incorrect and multiple classifications (this happens for all domains except *thyroids*). Thus, when domains have classes with unclear frontiers (i.e., it is difficult to find a discriminant description for them), the use of fuzzy sets can correct these frontiers. Notice also that the percentage of multiple

	Rand			CI			HR		
Data	C	I	M	C	I	M	C	I	M
iris	88.78	5.61	7.89	91.73	5.32	2.95	93.72	1.33	4.95
glass	35.46	9.50	55.04	9.56	6.26	84.18	30.63	13.97	55.40
thyroids	86.56	4.60	8.84	79.15	5.37	15.48	81.19	5.37	13.44
heart-statlog	65.40	16.19	18.41	54.55	14.97	30.48	56.40	16.67	26.93

Table 1. Percentage of correct classifications (C), incorrect classifications (I) and multiple classifications (M) of LID using the Rand index, CI and HR. Results are the mean of 7 trials of 10-fold cross-validation and they correspond to $p = 0.15$.

classifications produced by the Rand index is clearly lower than the one produced by both fuzzy versions (except for the *iris* domain). Since the percentage of incorrect classifications is also higher for the Rand index, we conclude that some of the objects that have not been classified using CI and HR (i.e., they produced multiple classifications) have been incorrectly classified using the Rand index. Therefore, the choice of a method has to be done taking into account the characteristics of the application domain. Sometimes it is preferable do not have answer in front of having an incorrect one; however, for some domains, to have more than one answer could be a valuable clue for classifying an object (for instance, when performing knowledge discovery).

Concerning the two fuzzy versions of LID, HR produces a lower percentage of multiple classifications than CI (except for the *iris* domain), however the percentage of incorrect classifications is higher in two of the domains (*glass* and *heart-statlog*). This means that CI is "more sure" in the classification of cases although a lot of times it cannot give a unique classification. Notice that for the *glass* domain the percentage of incorrect classifications is the lowest one; however the percentage of multiple classifications (i.e., no answer) is the highest one. We also conducted some experiments with the *bal* data set, also from the UCI repository, and the results of both HR and CI are not significantly different. Both indexes produce a percentage of incorrect classifications (3.52%) clearly lower than the produced by the Rand index (25.80%). Nevertheless, the Rand index produces a higher percentage of correct answers than the fuzzy indexes (65.53% in front of 60.54%).

Our conclusion is that the difference among the results using crisp and fuzzy indexes is strongly influenced by domain characteristics. Therefore it is necessary to perform an accurate analysis of the application domain (for instance, separability of classes, range of the values, etc.) in order to clearly determine the situations in which an index is better than others.

5 Conclusions and Future Work

We have introduced a new version of the method LID able to deal with fuzzy cases. Thus, cases have attributes taking continuous values which have been represented using fuzzy sets. In the current paper we experimented with LID using two different fuzzyfications of the Rand index one proposed by Campello and the other one proposed by Hüllermeier and Rifqi. From our experiments we concluded that it is difficult to assess a clear judgement about which measure is the best one in terms of classification accuracy. We performed experiments with different overlapping degrees of the fuzzy sets representing the values of the attributes, and we seen that this degrees do not significantly influence accuracy results. Our main conclusion is that the choice among the measures has to be made from an accurate analysis of the characteristics of the application domain.

All measures have a high computational cost, however we plan to experiment with the fuzzy extension proposed by Campello in order to exploit two parameters of the method: the t-norm and the t-conorm. In the experiments we used respectively the Minimum and the Maximum. In the future we plan to experiment with the t-norms of Łukasiewicz and Product and their dual t-conorms.

We also plan to use the similitude term generated by LID as a partial description of the solution classes as we have already done for the crisp version of LID. Now, this similitude term is fuzzy and this opens new opportunities to describe classes. In particular, we are thinking on knowledge discovery processes where the domain experts cannot define clearly the classes. In such domains, a fuzzy description of the classes could be very useful.

Acknowledgments

The authors acknowledge support of the Spanish projects MULOG2 (TIN 2007-68005-C04-01/04), ARINF (TIN 2009-14704-C03-03), Next-CBR (TIN 2009-13692-C03-01), and the grants 2009-SGR-1433/1434 from the Generalitat de Catalunya. The authors also want to thank Pilar Dellunde for her helpful comments and suggestions.

References

1. A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1):39–59, 1994.
2. E. Armengol and E. Plaza. Bottom-up induction of feature terms. *Machine Learning*, 41(1):259–294, 2000.
3. E. Armengol and E. Plaza. Lazy induction of descriptions for relational case-based learning. In L. De Reaedt and P. Flach, editors, *ECML-2001*, number 2167 in Lecture Notes in Artificial Intelligence, pages 13–24. Springer, 2001.
4. A. Asuncion and D.J. Newman. UCI machine learning repository, 2007.
5. R.J.G.B. Campello. A fuzzy extension of the rand index and other related indexes for clustering and classification assessment. *Pattern Recognition Letters*, 28(7):833–841, 2007.
6. A. de Luca and S. Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, 20(4):301–312, 1972.
7. R. López de Mántaras. A distance-based attribute selection measure for decision tree induction. *Machine Learning*, 6:81–92, 1991.
8. E. Hüllermeier and M. Rifqi. A fuzzy variant of the Rand index for comparing clustering structures. In *IFSA/EUSFLAT Conference*, pages 1294–1298, 2009.
9. M. Lebowitz. Categorizing numeric information for generalization. *Cognitive Science*, 9(3):285–308, 1985.
10. W.M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, Dec. 1971.
11. J.C. Schlimmer. Learning and representation change. In *Proceedings of the AAAI-87*, pages 511–515, 1987.
12. D.R. Wilson and T.R. Martinez. Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, 6:1–34, 1997.
13. I. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes, and S. Cunningham. Weka: Practical machine learning tools and techniques with java implementations, 1999.