Refining Discretizations of Continuous-Valued Attributes

Eva Armengol[1] and Ángel García-Cerdaña[1,2]

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

Abstract. The Rand index is a measure commonly used to compare crisp partitions. Campello (2007) and Hüllermeier and Rifqi (2009) respectively, proposed two extensions of this index capable to compare fuzzy partitions. These approaches are useful when continuous values of attributes are discretized using fuzzy sets. In previous works we experimented with these extensions and compared their accuracy with the one of the crisp Rand index. In this paper we propose the ε-procedure, an alternative way to deal with attributes taking continuous values. Accuracy results on some known datasets of the Machine Learning repository using the ε-procedure as crisp discretization method jointly with the crisp Rand index are comparable to the ones given using the crisp Rand index and its fuzzifications with standard crisp and fuzzy discretization methods respectively.

Keywords: Machine learning, Classification, Discretization methods.

1 Introduction

Knowledge representation of domain objects often involves the use of continuous values. One of the most widely used techniques to deal with continuous values is 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 a continuous value is split into several intervals. Elements of an interval are considered as equivalent and each interval is handled as a discrete value. There are different methods of crisp discretization. For instance, some of them take into account the length of the interval, or the frequency of the values, while others are entropy-based (for more information see [1]). 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 it is interesting to build a fuzzy discretization from a crisp one, as it is done for instance in [2].

Given an attribute taking continuous values, let $\alpha_1, \ldots, \alpha_n$ be the thresholds determining the discretization intervals for that attribute. Let α_0 and α_{n+1} be

V. Torra et al. (Eds.): MDAI 2012, LNAI 7647, pp. 258–269, 2012.

the minimum and maximum of the values that this attribute takes in its range. To fuzzy discretize the attribute, assuming that the fuzzy sets are trapezoidal, the membership vector is calculated in the following way:

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

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

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

In these formulas, the parameters δ_i represents the overlapping degree between contiguous fuzzy sets and they are computed as follows: $\delta_i = p \cdot |\alpha_i - \alpha_{i-1}|$, being the factor p a percentage that we can adjust.

Intuitively, since the representation using fuzzy sets is smooth around the thresholds of discretization intervals, it seems more appropriate than the crisp discretization in some domains. Nevertheless, the use of fuzzy sets implies that for each attribute value it is necessary to deal with the membership of this value to each fuzzy set representing the attribute. As we will see later, for the fuzzy extensions of the Rand index, this situation produces an increment of the run time with respect to the crisp version of the Rand index. For this reason, we have searched for an alternative method of discretization such that:

a) it retains as much as possible the advantages of the discretization methods using fuzzy sets,
b) it can be used with crisp measures such as the Rand index and, therefore,
c) it involves a reduction of the run time associated to the fuzzy measures.

As a concrete alternative, in the current paper we propose the ε-procedure, a method of discretization that induces classical partitions from continuous values of attributes. It consists on a refinement of the crisp discretization obtained by any standard discretization method. From a set of discretization thresholds $\alpha_1, \alpha_2, \ldots, \alpha_n$, the ε-procedure introduces the intervals $(\alpha_i - \delta_i, \alpha_i + \delta_i]$ being δ_i a parameter that depends on the length of the interval $(\alpha_{i-1}, \alpha_i]$.

In [3] we experimented with the Rand Index and two fuzzy extensions of it: one proposed by Campello [4] and the other proposed by Hüllermeier and Rifqi [5]. These experiments have been performed in the framework of a lazy learning method called LID. For this reason in the present paper we have carried out similar experiments with the ε-procedure in order to compare its behavior with the one obtained using standard discretization methods.

```
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
```

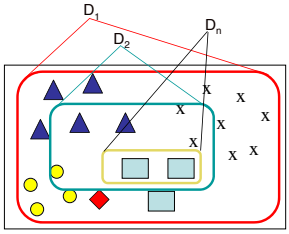

Fig. 1. The LID algorithm. On the right there is the intuitive idea of LID.

The paper is organized as follows. Section 2 contains preliminary concepts. In Sec. 3 the ε-procedure is presented. Section 4 contains the explanation of the experiments and a discussion of the results. The last section contains conclusions and future work.

2 Preliminary Concepts

In this section we explain the algorithm of the method LID used in the experiments. LID uses a mesure Δ to compare partitions. In the experiments we have used as Δ the Rand index and two of its fuzzifications, one proposed by Campello in [4] an the other by Hüllermeier and Rifqi in [5]. In this section we also explain these three measures in some detail.

2.1 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 (i.e., an object 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 *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 $\mathfrak{p}$, LID (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 the 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* 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 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, $\mathfrak{p}$ 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 $\mathfrak{p}$ is classified in the class of the majority of cases in S_{D_j}. When there is a tie between two classes, LID gives a multiple solution proposing both classes as the classification for $\mathfrak{p}$.

In our experiments we have taken the Rand index as the measure Δ that supports the selection of relevant attributes. This index is introduced in the next section.

2.2 The Rand Index

The Rand index [6] was conceived to compare clusterings produced by several automatic methods. The basic assumptions for using the Rand index are the following: 1) the clusterings to be compared are crisp in the sense that the set of clusters is a crisp partition of the domain; 2) the clusters are defined by both the objects that they contain and the objects that they do not contain; and 3) all objects are equally important in determining the clustering. From these assumptions it follows that a basic unit of comparison between two clusterings is how pairs of objects are clustered. If a pair of objects are placed together in a class in each one of the two clusterings, or if they are assigned to different classes in both clusterings, this represents a similarity between the clusterings. The opposite case is the one in which a pair of objects are in the same class in one clustering and in different classes in the other one. From this point of view, a measure of the similarity between two clusterings of the same data set can be defined as the number of equal assignments of object pairs normalized by the total number of object pairs.

Let X be a finite set $X = \{x_1, \ldots, x_n\}$; $\mathcal{P} = \{P_1, \ldots, P_k\}$ be a partition of X in k sets; and $\mathcal{Q} = \{Q_1, \ldots, Q_h\}$ a partition of the same set X in h sets. Given two objects x and x' we say that both objects are *paired* in a partition when

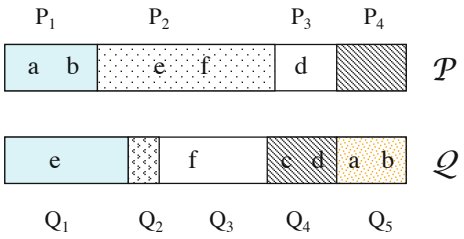

Fig. 2. Examples of paired and impaired objects: 1) a and b are paired in both partitions; 2) e and f are paired in $\mathcal{P}$ and impaired in $\mathcal{Q}$; 3) c and d are paired in $\mathcal{Q}$ and impaired in $\mathcal{P}$; 4) d and f are impaired in both partitions.

both objects belong to the same class of the partition (see Fig. 2). 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 among the partitions $\mathcal{P}$ and $\mathcal{Q}$ is defined as follows:

$$R(\mathcal{P}, \mathcal{Q}) = \frac{a + d}{a + b + c + d} \tag{2}$$

where,

- a is the number of pairs $(x, x') \in C$ such that x and x' are paired in both partitions.
- b is the number of pairs of objects $(x, x') \in C$ such that x and x' are paired in $\mathcal{P}$ and impaired in $\mathcal{Q}$.
- c is the number of pairs of objects $(x, x') \in C$ such that x and x' are impaired in $\mathcal{P}$ and paired in $\mathcal{Q}$.
- d is the number of pairs of objects $(x, x') \in C$ such that x and x' are impaired in both partitions.

The Rand index is commonly used to compare clusterings formed by automatic systems. It gives a measure of how similar are two clusterings. 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. However, when continuous values are represented by means of fuzzy sets the partitions induced are fuzzy and in this situation the Rand index is not appropriate to make comparisons involving fuzzy partitions. In the next sections we will explain two proposals of fuzzification of the Rand index.

2.3 The Campello's Fuzzy Rand Index

In [4] Campello extends the Rand index to make it feasible to compare fuzzy partitions. Given a finite data set $X = \{x_1, \ldots, x_n\}$, a *fuzzy partition* on X (in

the sense of Ruspini [7]) is any finite collection $\mathcal{P} = \{P_1, \ldots, P_k\}$ of fuzzy subsets on X such that $\sum_{i=1}^{k} P_i(x_j) = 1$, for each j, $1 \leq j \leq n$. To the end to define a fuzzy extension, Campello first rewrites the original formulation of the Rand index in an equivalent form by using basic concepts of set theory. Given the crisp partitions $\mathcal{P}$, with k sets, and $\mathcal{Q}$, with h sets, Campello defines the following sets of pairs:

V: pairs $(x, x') \in C$ paired in $\mathcal{P}$, W: pairs $(x, x') \in C$ impaired in $\mathcal{P}$,
Y: pairs $(x, x') \in C$ paired in $\mathcal{Q}$, Z: pairs $(x, x') \in C$ impaired in $\mathcal{Q}$,

where C is the set of pairs of elements of X defined in Sec. 2.2.

According to the sets above, the coefficients a, b, c and d of the Rand index in Eq. (2) can be rewritten in the following way: $a = |V \cap Y|$, $b = |V \cap Z|$, $c = |W \cap Y|$, and $d = |W \cap Z|$.

When we consider fuzzy partitions, the sets above are fuzzy sets. Let $P_i(x) \in [0, 1]$ denote the membership degree of the object $x \in X$ to the set P_i. Then, 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 $\oplus$:

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

Now, as it is usually done, Campello calculates the intersection of two fuzzy relations by using the t-norm (applied to the membership degrees of each pair to each relation). Then, using the sigma-count principle for defining the fuzzy set cardinality, he obtains the coefficients a, b, c, and d in the following way:

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

Then, the fuzzy version of the Rand index is also defined by Eq. (2), giving a measure of the similarity between two partitions. Since LID uses a normalized distance measure, we have to take $1 - R(\mathcal{P}, \mathcal{Q})$. Nevertheless, as Campello himself warns in [4], his fuzzy formulation of the Rand index does not satisfies some basic metric properties and it is properly defined only for the comparison of a fuzzy partition with a non-fuzzy reference partition (see also [5] for a discussion on this subject). However, notice that the correct partition in classification problems is commonly crisp; thus the use of the distance associated to the Rand index of Campello inside LID is justified. From now on, we denote as CI the distance associated with the Campello Rand index.

2.4 The Hüllermeier-Rifqi's Fuzzy Rand Index

Hüllermeier and Rifqi proposed in [5] a different fuzzy version for the Rand index which allows the comparison of two fuzzy partitions and satisfies all the desirable metric properties. In the next we recall the definition of this fuzzy version.

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 defined by $E_{\mathcal{P}}(x,x') := 1 - \|\mathcal{P}(x) - \mathcal{P}(x')\|$ being $\|.\|$ a distance on $[0,1]^k$. 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:

$$d(\mathcal{P},\mathcal{Q}) = \frac{\sum_{(x,x')\in C} |E_{\mathcal{P}}(x,x') - E_{\mathcal{Q}}(x,x')|}{n(n-1)/2}. \tag{3}$$

As it is shown in [5,11], the function (3) is a *pseudometric*, that is, it satisfies reflexivity, simmetry and triangular inequality, but it is not a *metric* because in general it does not satisfies the property of *separation* ($d(\mathcal{P},\mathcal{Q}) = 0$ implies $\mathcal{P} = \mathcal{Q}$). A fuzzy partition in the sense of Ruspini $\mathcal{P} = \{P_1, \ldots, P_k\}$ is called *normal* if it has a prototypical element, i.e., for every $P_i \in \mathcal{P}$, there exists an $x \in X$ such that $P_i(x) = 1$. Hüllermeier and Rifqi show that for normal partitions, taking the equivalence relation on X defined by

$$E_{\mathcal{P}}(x,x') = 1 - \frac{1}{2}\sum_{i=1}^{k} |P_i(x) - P_i(x')|, \tag{4}$$

the distance defined by Eq. (3) is a metric. From now on, we will call HRI (for Hüllermeier-Rifqui Index) this distance measure.

In addition to the extensions of the Rand Index presented by Campello and Hülermeier and Rifqi, other extensions of the Rand index have been proposed in the literature [8,9,10]. In the article [11] a comparative study of the indices presented in the mentioned papers in relation to the indexes proposed by Campello and Hülermeier and Rifqi is performed.

3 The ε-Procedure

Our goal is to design a procedure allowing the discretization of continuous values. The idea is, on the one hand, to keep the advantages provided by the fuzzy set representation and, on the other hand, to generate crisp partitions in order to use standard crisp measures to compare them. There are related works as the one from Ishibuchi and Yamamoto [12], that propose a method to construct fuzzy

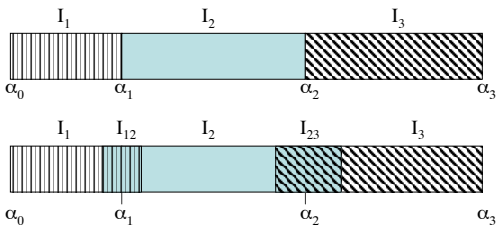

Fig. 3. The upper part shows the effect of a crisp discretization using three intervals. The lower part shows the discretization proposed by the ε-procedure.

discretizations from crisp ones. The authors consider that sometimes experts cannot give the discretization thresholds and they propose to dynamically obtain them. In [2] Kuajima et al. analyze the effects of fuzzy discretization on rule-based classifiers performance. The authors define the *fuzzification grade* as the overlap between adjacent fuzzy sets and they conduct experiments taking several of these grades.

The ε-procedure we introduce in the present paper discretizes continuous values by considering intermediate intervals that could be interpreted as the overlapping of two contiguous fuzzy sets. The partition generated in this way is crisp, therefore the standard Rand index can be used. Let f be an attribute taking continuous values, $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the discretization values for f, and α_0 and α_{n+1} be the minimum and maximum respectively of the values of f. The ε-procedure considers the following intervals: $[\alpha_0, \alpha_1 - \delta_1], (\alpha_1 - \delta_1, \alpha_1 + \delta_1], (\alpha_1 + \delta_1, \alpha_2 - \delta_2], (\alpha_2 - \delta_2, \alpha_2 + \delta_2] \ldots (\alpha_n + \delta_n, \alpha_{n+1}]$, where $\delta_i = p \cdot |\alpha_i - \alpha_{i-1}|$, being p an adjustable percentage. In order to avoid some undesired overlappings, the parameter p must respect the following constraint:

$$p \leq \frac{1}{2} \cdot \frac{|\alpha_{i+1} - \alpha_i|}{|\alpha_i - \alpha_{i-1}|}. \tag{5}$$

Notice that whereas with the usual discretization the values $v = \alpha_i - \varepsilon$ and $v' = \alpha_i + \varepsilon$, being ε sufficiently small, belong to different intervals, using the ε-procedure both v and v' belong to the same partition when $\varepsilon < \delta$. Figure 3 shows an example where a crisp discretization produces three intervals and for the same range of values, the ε-procedure introduces the intervals I_{12} and I_{23} that contain values around the thresholds α_1 and α_2. These new intervals join values that in the first discretization belong to different intervals.

4 Experiments

We have performed experiments with the goal of proving the feasibility of the ε-procedure as an alternative to fuzzy discretization. In previous works [3] we have shown that the fuzzifications of the Rand index have good predictivity but

Table 1. The left column shows the discretization method used in each one of the four situations. The central column shows the index used as Δ-measure. The right column corresponds to the label assigned to each situation.

DISCRETIZATION METHOD	Δ-MEASURE	LABEL
Crisp intervals given by an standard method	Rand Index	CRI
Fuzzy intervals built over the crisp ones	CI	FCI
Fuzzy intervals built over the crisp ones	HRI	FHRI
Crisp intervals built with the ε-procedure	Rand Index	εRI

they also have a high computational cost mainly due to the fact that they need to operate with all the membership degrees of all pairs of objects. Instead, the ε-procedure uses the crisp Rand index.

In the experiments we used four data sets, *iris, heart-statlog, glass,* and *thyroids* coming from the UCI Repository [13], where objects are described by attributes having continuous values. The discretization thresholds have been obtained with the MDL discretization method proposed by Fayyad and Irani's [14] and we have used the implementation of it given by Weka [15,16]. These thresholds have been taken as basis to define the fuzzy sets used by CI and HRI, and also for the ε-procedure to induce the new discretization intervals. The target of the experiments is to compare the predictivity of LID in the following four situations, which are summarized in Table 1:

1. The attributes with continuous values are discretized by using the thresholds given by a standard discretization method. Then LID runs using the Rand index (RI) as measure Δ to compare the partitions induced by the attibutes with the correct partition. So, in this case we have a crisp discretization and Δ = RI. We denote this procedure by CRI (the "C" stands by *Crisp*).
2. The attributes with continuous values are discretized by using the fuzzy sets built from the thresholds given by a standard discretization method. Then LID runs using CI as measure Δ with the Minimum and the Maximum as t-norm and t-conorm, respectively. We denote this procedure with the label FCI (the "F" stands by *Fuzzy*).
3. The attributes with continuous values are discretized by using the fuzzy sets built as in the previous procedure. Then LID runs using the Hüllermeier-Rifqi index (HRI) as measure Δ. So, in this case we have a fuzzy discretization and Δ = HRI. We denote this procedure by FHRI.
4. The attributes with continuous values are discretized by using the thresholds given by a standard discretization method and refined by the ε-procedure. Then LID runs using the Rand index (RI) as measure Δ. So, in this case we have the crisp discretization given by the ε-procedure, and Δ = CI. We denote this procedure with the label εRI.

As we have mentioned before, in the formulas (1) each δ_i is computed by $\delta_i = p \cdot |\alpha_i - \alpha_{i-1}|$, being the factor p a percentage that we can adjust and that must

Table 2. Mean accuracy of LID corresponding to the procedures labeled by CRI, FCI, FHRI, and εRI after seven trials of 10-fold cross-validation. The table shows the best results; the corresponding p is indicated between parenthesis.

DATASET	CRI	FCI	FHR	εRI
glass	35.460	25.696 (0.10)	38.644 (0.10)	**49.373** (0.10)
heart-statlog	65.397	62.857 (0.05)	65.026 (0.05)	**65.555** (0.10)
iris	88.780	91.917 (0.10)	94.482 (0.10)	**96.286** (0.10)
thyroids	**86.562**	81.447 (0.05)	82.331 (0.10)	84.935 (0.15)

satisfy the constraint (5). We have experimented with $p = 0.05, 0.10$ and 0.15. All these values of p satisfy (5) for all the considered data sets. 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. The algorithm of the fuzzy LID is the same explained in Sec. 2.1 but using a particular representation for the fuzzy cases.

Table 2 shows the mean accuracy of LID corresponding to the procedures labeled by CRI, FCI, FHRI, and εRI after seven trials of 10-fold cross-validation. The accuracy depends on the value of p and this may be different for each domain. In the table we show the best results and we indicate the corresponding p between parenthesis. Thus, for instance, the best result on *iris* for FCI is taking $p = 0.10$ whereas the best accuracy for FCI on *heart-statlog* is taking $p = 0.05$. In our experiments we have seen that for each data set there is a value of p that represents an inflection point on the accuracy. For instance, on the *iris* data set using εRI, the accuracy taking $p = 0.05$ is 93.720 and taking $p = 0.15$ is 95.333. The best value for p is different for each data set and also for each method. So, in practice, we should try with several values of p in order to find the best one.

Experiments show that the ε-procedure gives good predictive results outperforming all the fuzzy versions (FCI and FHRI) in all domains. Moreover, εRI also outperforms the procedure CRI except for the *thyroids* data set and for the *heart-statlog* where the accuracy of both is not significantly different.

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, i.e., most of the time it could be considered as a *no answer* of the system. This explains the low accuracy percentage given by all measures on the *glass* domain since most of times LID produces multiple classifications.

The computational complexity of the crisp Rand index is $\mathcal{O}((k + h) \cdot n^2)$, where k and h are the cardinalities of the partitions $\mathcal{P}$ and $\mathcal{Q}$ respectively; the cost of CI is $\mathcal{O}((max(k, h))^2 \cdot n^2)$; and the cost of the HRI is $\mathcal{O}(max(k, h) \cdot n^2)$. Notice that, in the worst case, both the Rand index and HRI have the same cost whereas the cost of CI is higher than them. In practice, the Rand index has lower

Table 3. Mean runtime necessary (in seconds) to evaluate a complete trial of 10-fold cross-validation on the datasets *iris* and *thyroids*

Dataset	CRI	FCI	FHRI	εRI
iris	2.624	226.773	77.463	14.358
thyroids	27.219	1092.055	383.037	165.348

cost than HRI mainly due to the lower complexity of the input data since fuzzy representations have to take into account membership degrees. Table 3 shows the mean runtime necessary to evaluate a complete trial of 10-fold cross-validation on the data sets *iris* and *thyroids*. These times have been obtained using a Mac with a processor Intel Core 2 Duo of 2.93 GHz.

5 Conclusions and Future Work

In this paper we have introduced a method, called ε-procedure, that constructs classical partitions on the range of an attribute taking continuous values. These partitions can be seen as refinements of the ones given by the expert or the ones given by a standard method of discretization. Moreover, the method can be seen as "similar" to the fuzzy methods of discretization since the ε-procedure takes into account the neighborhood of the thresholds given by the crisp discretization methods. The ε-procedure runs inside the LID method allowing it to deal with cases having attributes that take continuous values. We have carried on experiments with LID comparing its performance when dealing with both cases whose continuous attributes have been discretized and cases whose continuous attributes have been represented as fuzzy sets.

As future work we plan to conduct experiments to analyze in depth the effect of the value of δ in the accuracy of the ε-procedure. We think that this effect could also depend on the particular characteristics of the domain at hand. A different research line is to experiment with different Δ measures after the discretization produced by the ε-procedure. In particular we plan to experiment with the López de Mántaras (LM) distance [17] and to compare the results with those produced by a fuzzification of LM proposed in [18].

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), and the grants 2009-SGR-1433 and 2009-SGR-1434 from the Generalitat de Catalunya. The authors also want to thank Félix Bou and Pilar Dellunde for their helpful comments and suggestions.

References

1. Yang, Y., Webb, G.I., Wu, X.: Discretization Methods. In: The Data Mining and Knowledge Discovery Handbook, ch. 6, pp. 113–130. Springer (2005)
2. 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)
3. 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)
4. 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)
5. Hüllermeier, E., Rifqi, M.: A fuzzy variant of the Rand index for comparing clustering structures. In: Proceedings of the 2009 IFSA/EUSFLAT Conference, pp. 1294–1298 (2009)
6. Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66(336), 846–850 (1971)
7. Ruspini, E.H.: A new approach to clustering. Information and Control 15(1), 22–32 (1969)
8. Frigui, H., Hwang, C., Rhee, F.C.H.: Clustering and aggregation of relational data with applications to image database categorization. Pattern Recognition 40, 3053–3068 (2007)
9. Brower, R.: Extending the Rand, adjusted Rand and Jaccard indices to fuzzy partitions. Journal of Intelligent Informtion Systems 32, 213–235 (2009)
10. Anderson, D.T., Bezdek, J.C., Popescu, M., Keller, J.M.: Comparing fuzzy, probabilistic, and possibilistic partitions. IEEE Transactions on Fuzzy Systems 18(5), 906–918 (2010)
11. Hüllermeier, E., Rifqi, M., Henzgen, S., Senge, R.: Comparing Fuzzy Partitions: A Generalization of the Rand Index and Related Measures. IEEE Transactions on Fuzzy Systems 20(3), 546–556 (2012)
12. Ishibuchi, H., Yamamoto, T.: Deriving fuzzy discretization from interval discretization. In: Proceedings of FUZZ-IEEE 2003, vol. 1, pp. 749–754 (2003)
13. Asuncion, A., Newman, D.J.: UCI machine learning repository (2007)
14. Fayyad, U.M., Irani, K.B.: Multi-interval discretization of continuous-valued attributes for classification learning. In: Proceedings of IJCAI 1993, pp. 1022–1029 (1993)
15. Witten, I., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S.: Weka: Practical machine learning tools and techniques with java implementations (1999)
16. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. SIGKDD Explorations 11(1), 10–18 (2009)
17. López de Mántaras, R.: A distance-based attribute selection measure for decision tree induction. Machine Learning 6, 81–92 (1991)
18. Armengol, E., Dellunde, P., García-Cerdaña, À.: Towards a Fuzzy Extension of the López de Mántaras Distance. In: Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R. (eds.) IPMU 2012, Part I. CCIS, vol. 297, pp. 81–90. Springer, Heidelberg (2012)