Computación y Sistemas Vol. 9 Núm. 4, pp. 314-325

A Supervised Discretization Method for Quantitative and Qualitative Ordered Variables

Un método de Discretización Supervisada para Variables Cuantitativas y Cualitativas Ordenadas

Francisco J. Ruiz[1], Cecilio Angulo[1]and Núria Agell[2]

[1] Knowledge Engineering Research Group. Universitat Politècnica de Catalunya
Av. Víctor Balaguer s/n. 08800 Vilanova i la Geltrú (Spain)
{francisco.javier.ruiz, cecilio.angulo}@upc.edu
[2]Department of Quantitative Methods Management. ESADE-Universitat Ramon Llull
Av. Pedralbes 62-65. 08034 Barcelona (Spain)
nuria.agell@esade.edu

Article received on November 15, 2005; accepted on January 01,2006

Abstract

In this work, a new technique to define cut-points in the discretization process of a continuous attribute is presented. This method is used as a prior step in a regression problem, considered as a learning problem in which the output variable can be either quantitative (continuous or discreet) or qualitative defined over an ordinal scale. The proposed method emphasizes the concept of location to determine discretization cut-points. In the case of continuous outputs, the method is based on the maximization of the difference between distributions by using intervalar distances. In the case of qualitative outputs, a qualitative distance is defined over a structure of absolute orders of magnitude. The main characteristics of the method presented are illustrated through three examples, two for continuous outputs and the last for a qualitative output.
Keywords: Supervised Discretization, Regression, Qualitative Reasoning, Intervalar distance.

Resumen

En este trabajo se presenta una nueva técnica para definir las fronteras en el proceso de discretización de una variable continua. Este método es usado como paso previo en un problema de regresión, considerado como un problema de aprendizaje en el cual la variable de salida puede ser cuantitativa (continua o discreta) o cualitativa definida sobre una escala ordinal. El método propuesto enfatiza el concepto de "localidad" para determinar las fronteras de las discretización. En el caso de variables continuas, el método se basa en la maximización de la diferencia entre distribuciones usando distancias intercalares, y en el caso de salidas cualitativas, en una distancia definida sobre una estructura de órdenes de magnitud absolutos. La principal característica del método se ilustra con tres ejemplos, dos para salidas continuas y un último con salidas cualitativas.
Palabras Clave: Discretización Supervisada, Regresión, Razonamiento Cualitativo, Distancia Intervalar.

1 Introduction

Discretization is the process by which a continuous attribute is transformed into a discreet attribute. It is obtained by grouping different values from the continuous attribute and assigning each a unique label. One of the main reasons for following a discretization process is to allow the use of patterns with continuous variables in learning algorithms that, like decision trees, require discrete variables [Fayyad & Irani (1993)]. In addition, the use of discreet variables reduces computational costs in learning algorithms, as well as simplifying the interpretation of results [Liu et al. (2002)], [Dougherty et al. (1995)]. Sometimes these considerations justify the loss of information that can take place in the discretization process.

In this paper, a method of discretization is specifically set out for the problem of learning known as regression (or ordinal regression, if the output space is finite). This method takes into account the order existing in the output variable, whether this order comes from a real valued variable, or from a variable defined on an ordinal scale.

The new method is inspired by the idea of location, obtained through a distance. If the exit is continuous, the distance will be defined by a supervised kernel function [Agell et al. (2004b)] and the method will be specially adapted when it is used jointly with a learning process based on this type of kernel function. However, if the ordered output is based on orders of magnitude space, a distance based on the idea of location function proposed in [Rovira et al (2004)] will be used.

The traditional methods of discretization can be classified as bottom-up, that is to say, they have an incremental search for borders, or top-down, decreasing an initial set of borders of high cardinality. Unlike previous methods, this new method, finds the cut-points of the discretization in one single step, which drastically improves the computational speed with respect to others. In addition, the number of intervals does not have to be previously defined by the user, as in the stopping conditions of iterative algorithms. This algorithm will only depend on the nature of the handled data and the concept of the border location, which can be considered as a measurement of the border quality.

From the location concept, it is inferred that the effectiveness of a cut-point depends on the length of the intervals that this point separates. A cut-point can be a good fit for intervals of short length but not for intervals of greater length, in the same way that a geographic border can be a good fit for separate regions but not for separate countries.

In the following section, the state-of-the-art of classic discretization methods is briefly introduced, and the motivation of the present methodology is given. Then the distances that the proposed method uses are defined and described. In section four, the method of discretization is explained, based on the previously defined distances. Next, the main characteristics of the method presented are illustrated with three examples. Finally, the conclusions and indications for future works are given.

2 Discretization

Discretization has to be considered as a prior stage in the global process of inductive learning. This stage can be conducted directly by an expert, or automatically by means of a suitable methodology. In any event, the discretization process entails an implicit knowledge of the data. This knowledge is explicitly introduced into the learning process when considering experts' opinion, or implicitly extracted from data when discretization is done automatically.

The existing methods of discretization can be classified into two main categories: supervised and not supervised. Supervised methods do not consider the class to which the training patterns belong. Among these, the most significant are: the equal width intervals method, and the equal frequency intervals method [Dougherty et al. (1995)]. These methods are very simple to implement, and have a low computational cost; on the other hand, the results obtained are not very satisfactory in most cases.

Nevertheless, supervised methods consider the interdependence between the variable to be discretized and the class to which the patterns belong.

Some representative methods in this category are based on the entropy concept, such as D2 [Catlett (1991)] and MDLP [Fayyad & Irani (1993)], which optimizes a measurement of the entropy on the discretization considered. On the other hand, some techniques use statistical methods such as test or classical clustering techniques, ChiMerge [Kerber 1992], Chi2 and ConMerge [Wang & Liu (1998)]. Finally, other methods are based on the mutual information measure between class and variable, for example Zeta [Ho et al. (1997)], CADD [Ching et al. (1995)], CAIR [Wong & Liu (1975)], and CAIM [Kurgan & Cios (2004)].

Regarding the way the interval cut-points are obtained, discretization methods can also be considered as top-down or bottom-up methods.

The former begin with an empty list of cut-points that is increased iteratively. To this type belongs, for instance, the CAIM method. Nevertheless, the second group of methods begins with the complete list of variable values as possible cut-points, which are eliminated by joining the corresponding intervals. This is how the ChiMerge method operates.

In pattern recognition problems, the discretization methods used are mainly based on the maximization or minimization of a coefficient. This coefficient depends exclusively on the contingency table (joint frequencies table) defined from the discrete output and input variables.

In these methods, the order existing in the output variable is never considered. However, they take advantage of the fact that values of the patterns in the same interval belong to the same class. This paper considers that when information related to the ordering of the output variable exists, it must be regarded as a crucial point for improving the discretization

process. In addition, the method introduced highlights the fact that distributions of two contiguous intervals must have significant differences, which allows classes from the same interval to be distributed with a greater deviation. For instance, the variable *age* as input variable and *income* as an output variable. What allows us to distinguish, for example, between the age intervals (14, 25] and (25, 37], is not that the former have a low income and the later a high income, but that most people in the first interval have a low income while those in the second interval have either low, average or high incomes, that is to say, the deviation of incomes in the first and the second intervals are very different.

3 Distance Construction

3.1 Intervalar Distances over Continuous Variables

When considering real numbers, as in any Euclidean space, distances can be directly defined from the Euclidean product; but when considering a non-Euclidean space, as is the case of intervals, a different way of introducing distance functions must be found.

The natural way of introducing a concept similar to a distance between intervals is to consider the minimum between the two intervals. Nevertheless, this application is not a distance.

On the other hand, considering intervals as subsets of the real line the Hausdorff distance, defined as the maximum between the distances from any point of the interval to another interval (where the distance from a given point to an interval is defined as the minimum distance between this point and any one of the interval points) fulfils all the distance properties. In [González et al. (2004)] it is shown that this distance can be expressed as follows,

$$\mathrm{d}([\mathrm{a},\mathrm{b}],[\mathrm{c},\mathrm{d}]) = \Delta\mathrm{c} + \Delta\mathrm{r} \tag{1}$$

where,

$$\Delta\mathrm{c} = \left|\frac{\mathrm{a}+\mathrm{b}}{2} - \frac{\mathrm{c}+\mathrm{d}}{2}\right| \tag{2}$$

$$\Delta\mathrm{r} = \left|\frac{\mathrm{b}-\mathrm{a}}{2} - \frac{\mathrm{d}-\mathrm{c}}{2}\right| \tag{3}$$

In [González et al. (2004)] another methodology is proposed to transfer a structure of Euclidean space to the intervals. Each point in the real line is associated to a point in R^2. The composition between this map and the usual Euclidean product in R^2 defines a kernel function in the set of intervals. This kernel allows the defining of a new distance. By associating to each interval the point $\mathbf{x} = (c, r) \in R^2$, where c is the centre and r the radio of the interval, the expression of this distance is,

$$d(\mathbf{x_1}, \mathbf{x_2}) = \sqrt{(c_1 - c_2)^2 + (r_1 - r_2)^2} \tag{4}$$

Finally, in the set of intervals we can also consider distance associated to the intersection. In [Agell 1998] it is demonstrated that the intersection of intervals is a kernel function in the set of intervals, and the induced distance is defined as,

$$\sqrt{b-a+d-c-2\cdot l(\mathbf{x_1}\cap\mathbf{x_2})} \tag{5}$$

where $\mathbf{x_1}\cap\mathbf{x_2}$ is the intersection between the intervals $\mathbf{x_1}=[a,b]$ and $\mathbf{x}_2=[c,d]$, and $l(\mathbf{x_1}\cap\mathbf{x_2})$ is the intersection length,

$$l([a,b]\cap[c,d])=\max(0,\min(b-c,d-a)) \tag{6}$$

3.2 Distance in OM(n)

The absolute orders of magnitude models [Travé (2003)] are defined via a finite set of ordered symbols or qualitative ordered labels. These models introduce a structure that unifies the sign algebra and the intervalar algebra throughout a continuum of qualitative structures. In this section, the orders of magnitude model is briefly introduced [Agell (1998)] and a methodology to define distances on the model is considered.

The absolute order of magnitude model with granularity n, OM(n), is defined via a symmetric partition of the real line in 2n+1 segments (Figure 1) where $N_i=[-a_i, -a_{i-1})$, $0=\{0\}$ and $P_i=(a_{i-1},a_i]$. In OM(n), the *basic elements* are denoted by a set of ordered labels from the $S_1=\{N_n, N_{n-1},..., 0, P_1, ..., P_{n-1}, P_n\}$.

The *Quantity Space* $S\equiv$OM(n), is the set of labels like $[X,Y]$ for all $X,Y\in S_1$, with $X\leq$Y, i.e. $x\leq y$ for all $x\in X$, $y\in Y$. The interval $[X,Y]$ can be considered as the union of all the basic labels between X and Y.

The binary relation $\leq_P$, *to be more precise than*, $X,Y\in S$, X is more precise than Y ($X\leq_P Y$) if $X\subseteq Y$, is an order relation defined in S (directly obtained from the properties of the inclusion).

For all $X\in S-\{0\}$ the *base* of X is considered as the set $B_X=\{B\in S_1-\{0\} : B\leq_P X\}$ and, given a basic element $U\in S_1$, the *U-expansion* of X is $X_U=min\{Y\in S : X\leq_P Y \wedge U\leq_P Y\}$, i.e. the smaller interval with respect to the inclusion containing X and U.

To be able to define a distance in OM(n), a strategy in two steps is considered:

- Firstly, a *location function* is considered to associate a k-dimensional real vector to each label in S.
- Then, a *metric* defined in R^k is considered.

The location function and the metric have to be chosen in such a way that they capture the significance of the labels in the qualitative space in relation to the scenario considered in each application.

For example, we can consider the location function defined in [Rovira et al (2004)]. Following this function, each element X in S is codified by means of a pair of integer numbers $\mathbf{l}(X)=(l_1(X), l_2(X))$ as follows:

$$l_1(X)=\#(B_{X_{N_n}})+\#(B_X) \tag{7}$$

$$l_2(X)=\#(B_{X_{P_n}})-\#(B_X) \tag{8}$$

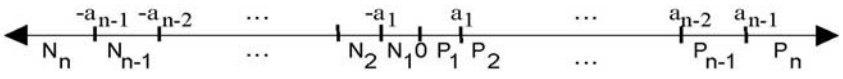

Fig. 1. Symmetric partition of the real line in 2n+1 basic elements.

so $l_1(X)$ is the number of basic elements in $S_1-\{0\}$ that are between the base of X and N_n, and $l_2(X)$ is the number of elements in $S_1-\{0\}$ that are between the base of X and P_n. This codification by points of R^2 takes into account not only the label's position but also its precision. The Euclidean distance between these points allows a distance in S to be

defined as follows:

$$\begin{aligned} d: \quad & S \times S \rightarrow [0,+\infty) \\ & (X,Y) \rightarrow \quad \delta \end{aligned} \tag{9}$$

with $\delta = \sqrt{\left(l_1(X) - l_1(Y)\right)^2 + \left(l_2(X) - l_2(Y)\right)^2}$.

The function d inherits all the properties of the distance in R^2, measuring the similarity between two labels based on the distance between its codifications.

4 Supervised discretization algorithm

Let us initially consider that there is a set of L sample patterns, each characterized by means of a set of attributes or input variables and an output variable, numerical or qualitative, but ordered. During the process of discretization, as usual, only uni-dimensional input variables will be considered jointly with the output variable. Let both X be one of the continuous input variables to be discretizated and Y the output set, whether continuous, discrete or qualitative. The training set can partially be represented by the set of values,

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_L, y_L)\} \subset X \times Y \tag{10}$$

Definition. A discretization D with granularity n of the variable X, obtained from the above training set, is defined as a set of disjoint intervals,

$$D=\{[d_0,d_1] , (d_1,d_2] , \ldots, (d_{n-1},d_n]\} \tag{11}$$

$d_0=min\{x_1,\ldots,x_L\}$, $d_n=max\{x_1,\ldots,x_L\}$ and $d_i \in \{x_1,\ldots,x_L\}$ with $d_0<d_1<d_2<\ldots<d_n$.

As a general criterion to be applied, any point $d_i \in \{x_1,\ldots,x_L\}$ will be an adequate frontier when it separates intervals on the variable X with different behaviour according to the specific learning problem being considered. Let us be more precise.

When d_i is being considered as a good candidate to be a frontier for the discretization, do not explicitly consider both adjacent intervals $(d_{i-1}, d_i]$ and $(d_i, d_{i+1}]$, or else the fitness of each candidate will be analyzed independently from the others, by using two intervals defined with the same number of data, Δ. This parameter, indicating the number of values of the variable X on the left and right of d_i to be considered when comparing both intervals, attempts to capture the concept of *localisation* that has been previously announced. The Δ value is also associated with the desired intensity of discretization: a high value will provoke a reduced number of frontiers, and similarly a low Δ value will increase their number. In the following, let us suppose that the L patterns in the training set, $\{(x_1,y_1),\ldots,(x_L,y_L)\}$ have been ordered such that $x_i<x_{i+1}$ $\forall i$.

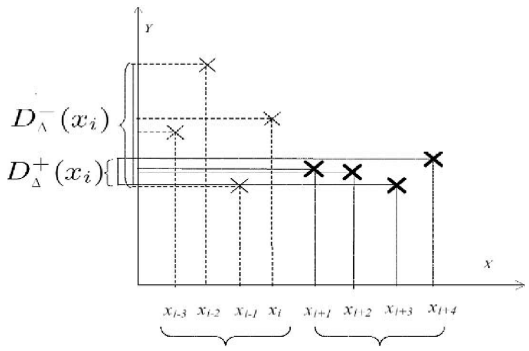

Fig. 2. Meaning of intervals $D_\Delta^-(x_i)$ and $D_\Delta^+(x_i)$ for Δ=4**.**

For a fixed parameter Δ, a pair of sets will be associated for each of the values x_i, $i \geq \Delta$ and $i<L-\Delta$,

$$D_-(x_i)=\{x_{i-\Delta+1}, \ldots, x_i\} \tag{12}$$

$$D_+(x_i)=\{x_{i+1}, \ldots, x_{i+\Delta}\} \tag{13}$$

corresponding to the Δ patterns with their value on the input variable closer to and both before and after x_i, respectively. A couple of sets will also be associated,

$$D^-_\Delta(x_i)=\{y_{i-\Delta+1}, \ldots, y_i\} \tag{14}$$

$$D^+_\Delta(x_i)=\{y_{i+1}, \ldots, y_{i+\Delta}\} \tag{15}$$

corresponding to the outputs of the patterns in $D_-(x_i)$ and $D_+(x_i)$ respectively. Note that these sets are not defined for the first Δ–1 and the last Δ patterns of the ordered series, so these patterns will never be selected as frontiers in the discretization.

Sets $D_-(x_i)$ and $D_+(x_i)$ must be represented by intervals. One possibility is to associate each of these sets to their complete range, as can be appreciated in Figure 2 for Δ=4. Using the complete range as associated interval is a more direct way to establish the association; however this presents two problems. First, the entire range is very influenced by possible *outliers*. Secondly, and related to the former disadvantage, the complete range will generate very similar associated intervals, so comparison between them is more difficult. Therefore, it would be preferable to associate an interval for each distribution with shorter length than the range; for instance, an interval centred at the mean, like $(\bar{y}-\sigma, \bar{y}+\sigma)$, where $\bar{y}$ is the mean of the distribution and σ its standard deviation; or more generally, $(\bar{y}-k\sigma, \bar{y}+k\sigma)$, with a new parameter to be adjusted. It is also possible to associate an interval centred at the median, like the inter-quartile range.

Similarly, when outputs are qualitative, it is possible to define $S_P(D_-(x_i))$ (respectively, $S_P(D_+(x_i))$) as the qualitative expression with higher precision for the outputs corresponding to, at least, a P % of the patterns with values in $D_-(x_i)$ (resp. $D_+(x_i)$), with $P<100$ a positive parameter previously fixed, as discussed above. When different qualitative expressions with equal precision satisfy this condition, those with the larger number of patterns are chosen.

It is important to note that in this methodology, the set of the pair of patterns being separated as the candidate to be a frontier, d_i, is not represented by its input, the variables to be discretized, but by its output values. In this way, patterns are not represented by themselves, but by what they represent in the specific learning problem to be solved.

Maximization of the *interval distance* between $D_-(x_i)$ and $D_+(x_i)$, for both continuous and ordinal discrete cases, or

the *qualitative distance* between $S_P(D_-(x_i))$ and $S_P(D_+(x_i))$, for the qualitative case, will be the criterion that allows the selection of frontiers for the discretization. Hence, the criterion selects as frontier those values of x_i that are significant local maxima for the distance function.

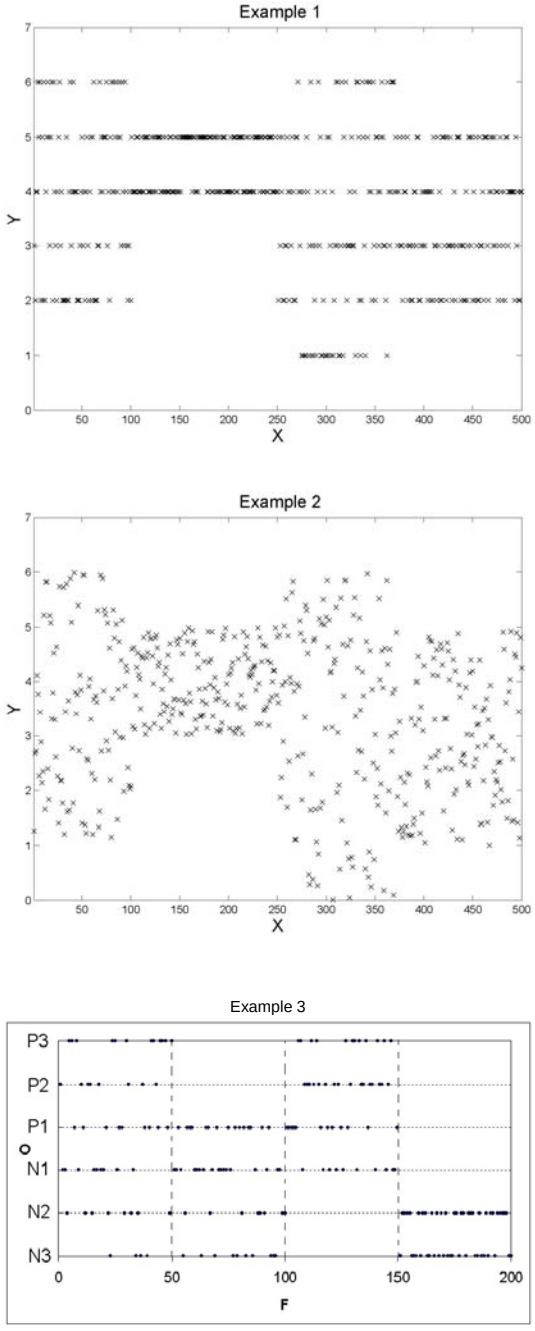

Fig. 3. Data distribution along the different examples.

5 Examples

The proposed methodology will be illustrated in this Section through three artificially generated examples. They will allow the frontiers selection process and the global features of the entire method to be shown. For the first example, a set of data with discrete output will be generated, in such a way that results obtained can be compared with the state-of-the-art discretization method CAIM. The second example will consider a continuous output variable, the kind of problem that standard discretization methods are not able to deal with. Thirdly, a set of patterns with ordered qualitative output in OM(3) will be considered.

5.1 Quantitative Examples

The first example is based on a training set with 500 patterns characterized by one continuous input variable and one ordinal discrete output variable taking values on six classes labelled as {1,2,3,4,5,6}. Data is displayed such that, ordered by the input variable, the first 100 patterns are distributed into classes 2-6, patterns from 101 to 250 are distributed in classes 4-5, patterns through 251 to 370 in classes 1-6 and, finally, the last 80 are in classes from 2 to 5. This distribution is illustrated in Figure 3(a).

Ideal frontiers were artificially laid out on patterns 1, 100, 250, 370 and 500. The CAIM algorithm was initially applied with granularity level 4. The discretization obtained is shown in Table 1. It can be noted that the result obtained is not exactly as expected. The same Table also shows results based on the presented method for several values of the parameter Δ. The interval Euclidean distance has been employed. It can be observed that increasing the parameter value results in a decrease in the number of frontiers: for Δ=30 the expected frontiers are obtained.

Table 1. Results for Example 1.

Method	Time	Frontiers
CAIM	52.1 s	{1, 149, 270, 486, 500}
Δ=10	0.03 s	{1,27,37,…*(n=11)*,...,392,500}
Δ=20	0.03 s	{1,41,61,100,231,340,369,500}
Δ=30	0.03 s	{1,100,245,369,500}

Table 2. Results for Example 2.

Method	Time	Frontiers
Δ=10	0.03 s	{1,10,24,.(n=26).,482,500}
Δ=20	0.03 s	{1,22,100,..(n=10),..475,500}
Δ=30	0.03 s	{1,31,100,252,372,467,500}
Δ=40	0.03 s	{1,100,251,372,500}

The same Table displays the time calculation for each of the two algorithms. The new method improves this feature in three orders of magnitude with respect to CAIM, so overcoming the need for evaluating several values of Δ until the definitive selected result.

Figure 4 shows the distance function between adjacent intervals for Δ equal to 10, 20 and 30, respectively. A low value in this parameter allows many frontiers to be considered and determined with high precision. Increasing the parameter value implies fewer frontiers with lower precision. Detection of local maxima is performed with a voting algorithm that extracts those points that are absolute maxima to local range with length Δ. The right-hand graphics in Figure 4 represent the result of this votation in Example 1.

For the second example, a 500-pattern set has also been generated, in this case with the continuous output variable having values between 0 and 6 (Figure 3(b)). The first 100 patterns are distributed on the interval [1,6], the next 150 patterns in [3,5],

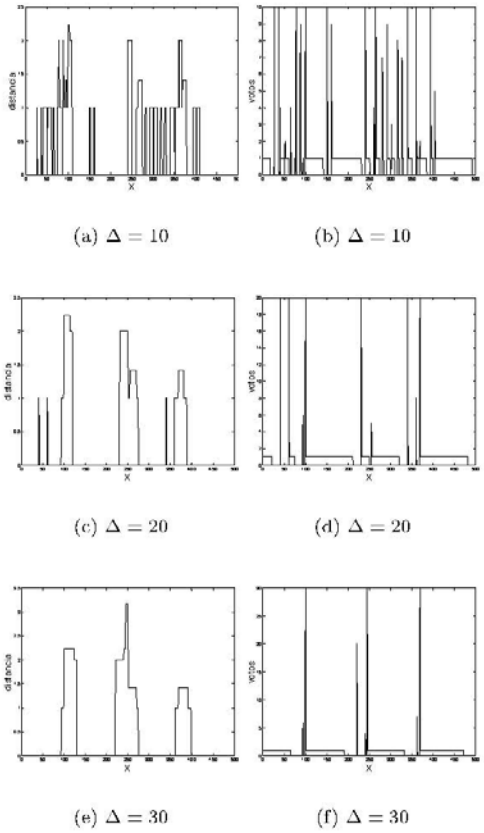

Fig. 4. Distance between intervals image of adjacent sets (left) and locating local maxima (right) for Example 1.

from 251 to 370 in [0,6] and the remainder in [1,5]. Previously, data were ordered according to their input variable. The same methodology as in the first example is used. Results are presented in Table 2 for several values of parameter Δ. A similar discussion is possible for this example: a low value of Δ deserves a high number of frontiers. A fitted Δ value selects the expected frontiers.

No comparison is possible in this example with the CAIM method because this method cannot deal with continuous output variables.

5.2 Qualitative output example

A set of 200 patterns is generated: patterns from 1 to 50 are associated to labels from N_3 to P_3; patterns from 51 to 100 are labelled between N_3 and P_1; from 101 to 150, outputs run between N_1 and P_3, and the last 50 patterns take values from N_3 to N_2 (Figure 3(c)). The discretization method should be able to determine frontiers 50, 100 and 150.

Several values of the parameter P, related to the frequency of patterns associated with qualitative expressions, generate results shown in Figure 5 for the same value Δ=15. Best results are obtained for the range from 70 to 100. However, a value of 100 signifies considering the less precise qualitative expression, hence taking into consideration outliers during the discretization process.

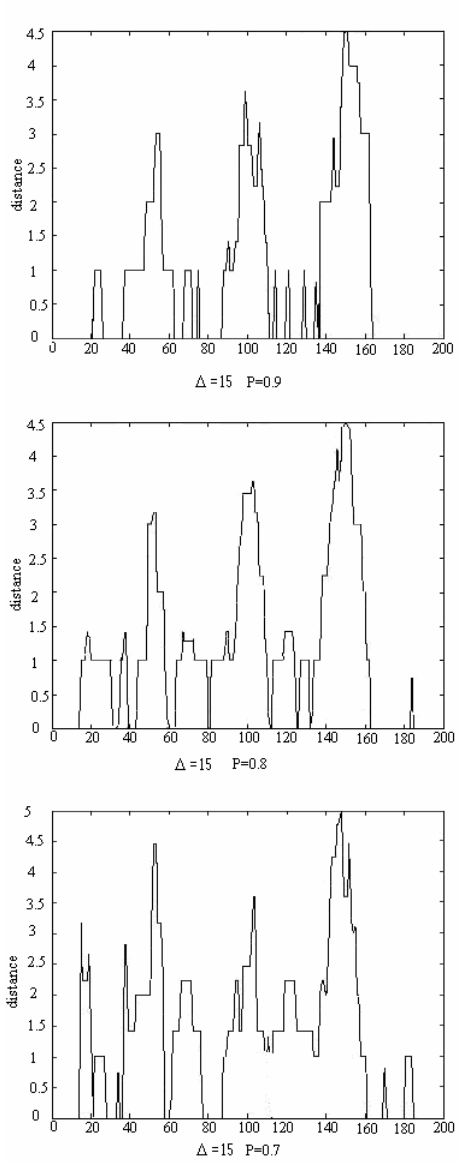

Fig. 5. Results varying the value of the parameter P.

6 Conclusions and future work

This work offers a new methodology to deal with the problem of supervised discretization for continuous variables taking into consideration ordination in the output variable; in this way, it is possible to make a discretization of variables for regression problems, whether the output variable is continuous, discrete (ordinal regression) or qualitative. The main feature of the method is that it considers that it is not mandatory for data from the same discretizated interval to be associated to similar data for the output variable. It has been proved that more valuable information can be captured even when frontiers are separating intervals with different distributions in the output variable. The new method selects how to cluster information based on a measure of the interval distance or a qualitative distance; other alternatives could be

considered. Artificial examples have allowed the efficiency of the algorithm to be analysed with respect to a Standard discretization algorithm.

Future work will be carried out on new distances to measure the difference between data distributions on adjacent intervals. It could also be noted that the new methodology is a starting point for dealing with the problem of discretization on several simultaneous variables on real data.

Acknowledgements

This work has been partially supported by the coordinated project MERITO (analysis and development of innovative soft-computing techniques integrating expert knowledge: an application to the measure of financial credit risk), funded by the Spanish Ministry of Science and Technology (TIC2002-04371-C02).

References

1. Núria Agell. *Estructures matemàtiques per al model qualitatiu d'ordres de magnitud absoluts*. Ph. D. Thesis. Universitat Politècnica de Catalunya, 1998.
2. Núria Agell, F.J. Ruiz, and Cecilio Angulo**.** A kernel intersection defined on intervals. In *Proc del Congrés Català d'Intel·ligència Artificial*, 2004.
3. Núria Agell, Xari Rovira, Francisco Ruiz, and Cecilio Angulo. Kernel machines for continuous and discrete variables: An application to credit risk measurement. In *Proc. of the Learning04*, Elche, Spain, 2004.
4. Catlett J. On changing continuous attributes into ordered discrete attributes. *Proc. Fifth European Working Session on Learning*. Berlin: SpringerVerlag, pp. 164-177, 1991.
5. J.Y. Ching, A.K.C. Wong, and K.C.C. Chan. Class-dependent discretization for inductive learning from continuous and mixed mode data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):641–651, 1995.
6. James Dougherty, Ron Kohavi, and Mehran Sahami. Supervised and unsupervised discretization of continuous features. In *International Conference on Machine Learning*, pages 194–202, 1995.
7. U. M. Fayyad and K. B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *Proc. of the 13th IJCAI*, pages 1022–1027, Chambery, France, 1993.
8. Luis González, Francisco Velasco, Cecilio Angulo, J.A. Ortega, and F.J. Ruiz. Sobre núcleos, distancias y similitudes entre intervalos. *Revista Iberoamericana de Inteligencia Artificial*, 8(23):111–117, 2004.
9. Ho, K.M and Scott, P.D. Zeta: A global method for discretization of continuous variables. In *KDD97: 3rd International Conference of Knowledge Discovery and Data mining*. Newport Beach, CA, pp. 191-194, 1997.
10. Kerber, R. ChiMerge: Discretization of Numeric Attributes. *Proc. 10th National Conference on Artificial Intelligence*. MIT Press, pp. 123-128, 1992.
11. Kurgan, L.A and Cios, K.J CAIM Discretization Algorithm. *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 2. pp. 145-153, 2004.
12. H. Liu, F. Hussain, C. Lim Tam, and M. Dash**.** Discretization: An enabling technique. *Data Mining and Knowledge Discovery*, 6(4):393–423, 2002.
13. Rovira, X., Agell, N., Sánchez, M., Prats, F. and Parra, X. An Approach to Qualitative Radial Basis Function Networks over Orders of Magnitude. *Proceedings of 18th International Workshop on Qualitative Reasoning*. 2004.
14. Travé-Massuyès, L. and Dague, P. *Modèles et raisonnements qualitatifs*. Hermès, 2003.
15. Wang, K. and Liu, B. Concurrent discretization of multiple attributes. *Pacific-Rim International Conference* on AI. pp. 250-259, 1998.
16. Wong, A.K.C. and Liu, T.S.: Typicality, diversity and feature pattern of an ensemble, *IEEE Trans. Computers*, vol. 24, pp. 158-181, 1975.

Francisco J. Ruiz *received the M.Sc. degree in Physics in 1988. He is a PhD candidate in Control, Vision and Robotics programs at the Automatic Control Department (Universitat Politècnica de Catalunya) where he is an Assistant Lecturer. His research focuses on Qualitative Reasoning and Kernel Methods with applications in control systems and finances.*

Dr. Cecilio Angulo Bahón *received an MSc degree in Mathematics in 1993 from the Universitat de Barcelona, and a PhD. in Science from the Universitat Politècnica de Catalunya in 2001. He is a lecturer in the Automatic Control Department at the Universitat Politècnica de Catalunya. The subject of his research is statistical learning with applications in control systems and finances.*

Dra. Núria Agell Jané *Professor at the Quantitative Methods department, ESADE, University Ramon Llull. Director of the PhD Programme in Management Sciencesl. Graduated in Mathematics from the University of Barcelona and PhD. in Applied Mathematics from the Technical University of Catalonia. Her main research activities are currently related to: (1) development of soft-computing models and technologies based on qualitative and fuzzy reasoning, (2) application of Artificial Intelligence techniques to finances, marketing and knowledge management.*