A Novel Approach to Robust Fuzzy Clustering of Relational Data

Mario G.C.A. Cimino, Beatrice Lazzerini, Francesco Marcelloni
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni
University of Pisa
Via Diotisalvi 2, 56122 Pisa
Fax: +39 050 568522; Tel: +39 050 568678
{mario.cimino, b.lazzerini, f.marcelloni}@iet.unipi.it

Abstract **- In this paper, we propose a new approach to robust fuzzy clustering of relational data, which does not require any particular restriction on the relation matrix. More precisely, we adopt an algorithm based on the Fuzzy C-Means (FCM) algorithm, improved with Davé's concept of Noise Cluster, and suitable for data which are expressed in terms of mutual numerical relationships among patterns. In this way, we tackle a relational clustering problem taking advantage of the stability and effectiveness of object data clustering algorithms. We also exploit the concept of prototype as representative of the mutual relationships of a group of similar patterns. We show that our approach is more scalable and less sensitive to cluster initialization and parameter variations than the robust Non-Euclidean Fuzzy Relational data Clustering algorithm (robust-NE-FRC), one of the most efficient recently proposed relational algorithms, on both real and synthetic data sets.**

Keywords **– robust fuzzy relational clustering, robust fuzzy C-means, noise prototype.**

I. INTRODUCTION

Relational data are frequently encountered in fields where the characterisation of objects by means of attributes is not possible or appropriate, whereas it is easier to identify the mutual relationships among them. To partition relational data, several fuzzy relational clustering algorithms have been proposed. Most algorithms determine a fuzzy partition of the data set using an alternating optimisation scheme to iteratively minimize an appropriate objective function based on a (usually Euclidean) dissimilarity relation R [1-5]. When R is non-Euclidean, the algorithms can still be applied but no theoretical proof for their convergence is provided [6].

As far as fuzzy object clustering is concerned, one of the most popular algorithms is the fuzzy C-means (FCM) algorithm, which can be applied if the objects of interest are represented as points in a multi-dimensional space. FCM relates the concept of object similarity to spatial closeness and finds cluster centres as prototypes. Several examples of application of FCM to real clustering problems have proved the good characteristics of this algorithm with respect to stability and partition quality. Further, its convergence has been formally demonstrated. As it is well known in the literature, FCM is not robust against noise and outliers. To overcome these problems, Davé proposed a modified version of FCM, denoted robust-FCM in the following, based on the concept of noise cluster [7].

In this paper, we present a new fuzzy relational algorithm which has resulted to be stable without requiring any particular restrictions on the relation matrix. After representing each object by the vector of its relation strengths with the other objects in the data set, we apply the robust-FCM to this vector space. In this way, though working on relational data, we can exploit the good properties of stability and effectiveness of object data clustering algorithms. Finally, we experimentally show the high scalability and low sensitivity to cluster initialization and parameter variations of our method on both real and synthetic data sets. To this aim, we compare our method with the robust Non-Euclidean Fuzzy Relational data Clustering algorithm (robust-NE-FRC), recently proposed in [8].

II. OUR APPROACH

Our approach is founded on the following observation: in relational clustering, each pattern $\underline{x}_i$ is defined by the values of the relations between $\underline{x}_i$ and all patterns in the data set. If the data set is composed of M patterns, each pattern $\underline{x}_i$ can be represented as a vector $\underline{x}_i = [r_{i,1},...,r_{i,M}]$ in $\Re^M$, where $r_{i,j}$ is the extent to which $\underline{x}_i$ is related to $\underline{x}_j$. Since a relational clustering algorithm should group patterns that are "closely related" to each other, and "not so closely" related to patterns in other clusters, as indicated by their relative relational degrees, we can obtain clusters by grouping patterns based on their closeness in the space $\Re^M$. Representing the $M \times M$ relation matrix as M vectors defined in the feature space $\Re^M$ allows us to transform a relational clustering problem into an object clustering problem. In previous papers [9][10], we have applied classical FCM to partition the M vectors of the data set. We have shown that our approach is more stable than some of the most popular fuzzy relational clustering algorithms such as RFCM and NERFCM [6]. Further, we have discussed how this conversion allows us to determine the optimal number of clusters using some popular indexes like the Xie and Beni's index [11] and the Fukuyama and Sugeno's index [6]. Finally, we have highlighted the importance of having a prototype for each cluster. Here, a prototype is a (possibly virtual) pattern whose relationship with all patterns of the data set is representative of the mutual relationships of a group of similar patterns.

III. Improving clustering with the noise concept

As it is well known in the literature, the classical FCM is not robust against noise and outliers [7]. Thus, a pattern, which is strongly related to no pattern in the data set, is forced to belong to one or more of the clusters, possibly modifying the shape of these clusters. To overcome this problem, in this paper, we make our approach robust to noise and outliers by adopting the robust version of FCM (robust-FCM) proposed by Davé [7]. Robust-FCM is based on the concept of noise-prototype, that is, an entity that is always at the same distance δ from each pattern in the data set. The presence of the noise-prototype changes the constraint on the membership values imposed by the classical FCM: If C is the number of clusters, the constraint is $0 < \sum_{i=1}^{C} u_{i,k} \leq 1$ instead of $\sum_{i=1}^{C} u_{i,k} = 1$, where $u_{i,k}$ represents the membership value of pattern $\underline{x}_k$ to cluster i. Thus, the objective function becomes

$$J(V,U,X) = \sum_{i=1}^{C} \sum_{j=1}^{M} u_{ij}^{m} \cdot d^2(\underline{v}_i, \underline{x}_j) + \sum_{j=1}^{M} u_{*j}^{m} \cdot \delta^2 \quad (1)$$

where M is the number of patterns, m is the fuzzification coefficient, $d(\underline{v}_i, \underline{x}_j)$ is the distance between the prototype $\underline{v}_i$ of cluster i and $\underline{x}_j$, u_{ij} is the membership degree of point $\underline{x}_j$ to the cluster represented by $\underline{v}_i$, and the membership u_{*j} of point $\underline{x}_j$ to the noise cluster (indicated by the subscript *) is

$$u_{*j} = 1 - \sum_{i=1}^{C} u_{ij} \quad (2)$$

The noise distance δ, called the *resolution parameter*, is obviously a critical factor of the algorithm, and its value should be based on data set statistics: in particular, it is related to the concept of "scale" in robust statistics. In [7], Davé recommends to choose

$$\delta^2 = \lambda \left[\frac{\sum_{i=1}^{C} \sum_{j=1}^{M} d^2(\underline{v}_i, \underline{x}_j)}{MC} \right] \quad (3)$$

where λ is a positive real multiplier.

One serious problem is the stability with respect to variations of the resolution parameter. For this reason, as suggested in [7], λ is initialized to a maximum value and changes adaptively to a smaller value as the algorithm progresses.

Like for the majority of robust clustering algorithms, also the results produced by robust-FCM depend on the choice of the initial partition. To reduce this dependence, we adopt the following procedure. First, we apply the classical FCM with low termination accuracy (e.g., equal to 0.1 or greater if the data set is bigger) and random initial partition. Then, we use the resulting final partition as the initial partition of robust-FCM (the membership of each pattern to the noise cluster is initialized to 0). We experimentally verified that this procedure strongly reduces the effect of the initialization on the final partition.

IV. Experimental results

We applied both our algorithm and robust-NE-FRC to a real data set and to a synthetic data set. For the latter, we have randomly added increasing percentages of patterns, and tested the convergence speed of both the algorithms.

To assess the stability of the results with respect to the initial random partition, we evaluated the classification rates and the partition coefficient over 10 runs and computed the average and standard deviations. The partition coefficient $P(U) = \frac{1}{M} \left(\sum_{i=1}^{C} \sum_{j=1}^{M} u_{ij}^2 \right)$ is a further index commonly used to evaluate the goodness of a partition. P essentially measures the distance the partition U is from being crisp by assessing the fuzziness in the rows of U. P varies in the interval $[1/C, 1]$. Empirical studies show that maximizing P leads to a good interpretation of data. Thus, the closer P is to 1, the better the partition is. In the experiments described in sections IV.A and IV.B, when the robust-NE-FRC does not reach the convergence after 100 steps, it returns the final result. On the contrary, in the experiments presented in section IV.C, only runs of robust-NE-FRC that achieve convergence are taken into account.

To test the stability to the fuzziness variation (expressed through the fuzzification coefficient m) and the scalability, we repeated the experiments for m ranging in [1.1, 3.0], with step equal to 0.1. It is well-known in the literature that, when the number of objects and, consequently, the dimensionality of the problem increase, m has to be decreased to enforce the assignment of patterns to clusters. The lower the value of m necessary for the clustering algorithm to achieve a reasonable solution, the lower the robustness of the algorithm to the dimensionality curse. Finally, we show how the prototype-based analysis, possible with our algorithm, helps us interpret the results.

A. Countries data set

Countries Data (CD) set is a classical example of relational data [8]; dissimilarities between 12 countries are obtained by averaging the results of a survey among political science students; Fig. 1 shows the dissimilarity matrix in graphical form, where the degree of dissimilarity is painted as a grey level. A brief analysis of this structure reveals that Egypt is dissimilar to any of the three typical groups (numbered 0, 1 and 2 in the first column): its dissimilarity degrees are all grey-white (i.e., high) except (obviously) with itself. We can therefore consider it as an "outlier" to be put in the noise cluster (3). Also India has a strange silhouette, but Egypt is undoubtedly the "worst clusterable" country. This data set can be considered as an example of real, non-Euclidean relational data.

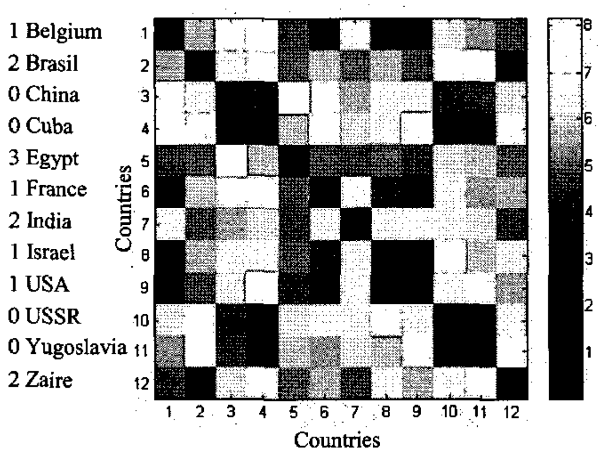

Fig. 1 Countries Data dissimilarity matrix.

Fig. 2 shows the three relational prototypes produced by robust-FCM: the continuous, dashed and dotted lines, respectively; the figure highlights that only Egypt (i.e., pattern 5) and, in part, India (pattern 7) have high dissimilarity with all prototypes. If we consider the value 4.6 as the threshold of dissimilarity to belong to a cluster, we can distinguish a first cluster (continuous line) including patterns 3, 4, 10 and 11, a second cluster (dashed line) consisting of patterns 2, 7 and 12, and a third cluster (dotted line) formed by patterns 1, 6, 8 and 9. Let us note that pattern 5 has a high average dissimilarity with all clusters, thus it belongs to the noise cluster. This qualitative analysis based on the observation of the prototypes is confirmed by the membership values computed by the algorithm. In particular, the memberships of Egypt and India to the three clusters are 0.18, 0.28, and 0.25, and 0.21, 0.31 and 0.18, respectively.

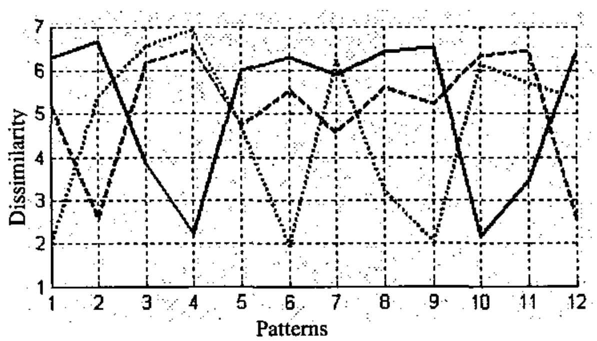

Fig. 2 Prototypes produced by robust-FCM on Countries Data.

Figures 3 and 4 show the goodness of our algorithm with respect to robust-NE-FRC. The performance of our algorithm in terms of classification rate and partition coefficient (Fig. 3) is very stable for all values of m. On the contrary, robust-NE-FRC (Fig. 4) requires values of m below 1.7 to achieve a satisfactory result.

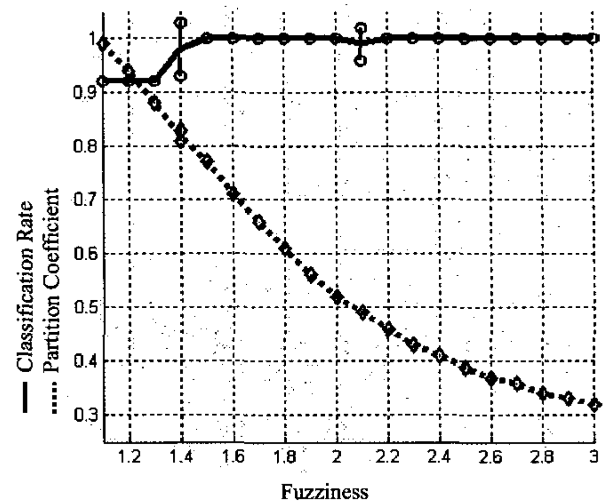

Fig. 3 Effectiveness of robust-FCM on Countries Data.

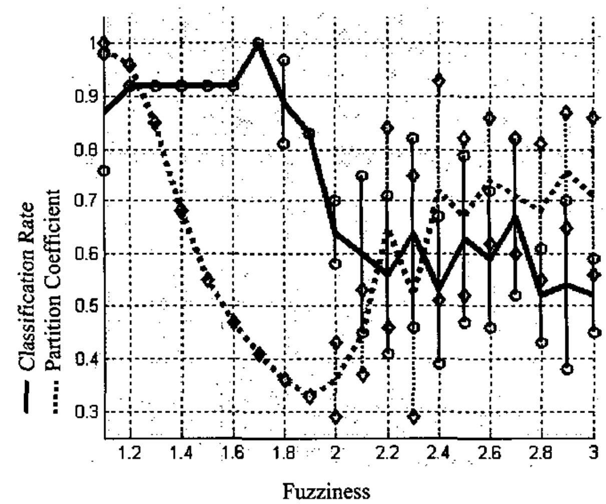

Fig. 4 Effectiveness of robust-NE-FRC on Countries Data.

B. Synthetic data set

Fig. 5 shows the synthetic data set extracted from [7]: three compact clusters generated using uniform random distribution of points centred around three prototypes, plus some random added noise.

For the sake of simplicity, we compute the relational patterns applying the Euclidean distance (or 2-norm distance) as dissimilarity measure; however, our approach does not require any particular metric on the relation matrix. We denote

R-synthetic-L_2 the relational synthetic data set built with L_2 norm.

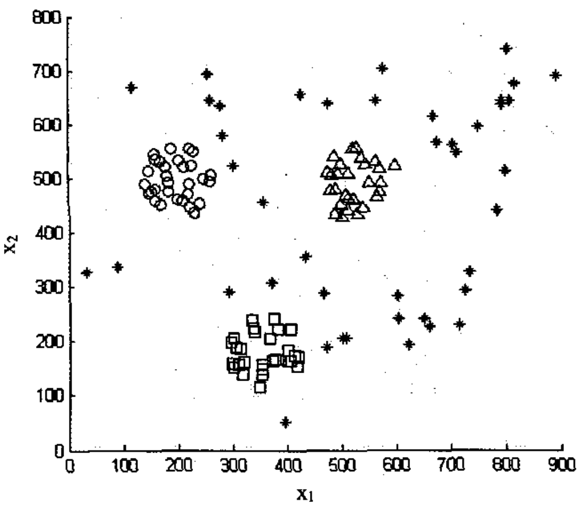

Fig. 5 The synthetic data set in the feature space.

Fig. 6 shows the relational prototypes produced by our method; we have ordered the patterns by class for better visualization; only the noise patterns, from 87 to 128, have high dissimilarity (>100) with all prototypes. If we assume the value 100 as the threshold to determine the membership to a cluster, we can see a first cluster (dotted line) formed by the patterns from 1 to 30, a second cluster (dashed line) with patterns from 31 to 58, and a third cluster (continuous line) with patterns from 59 to 86. We can also observe that some noise patterns (from 87 to 128) are very close to the chosen threshold. This qualitative analysis is confirmed by the membership matrix computed by the algorithm.

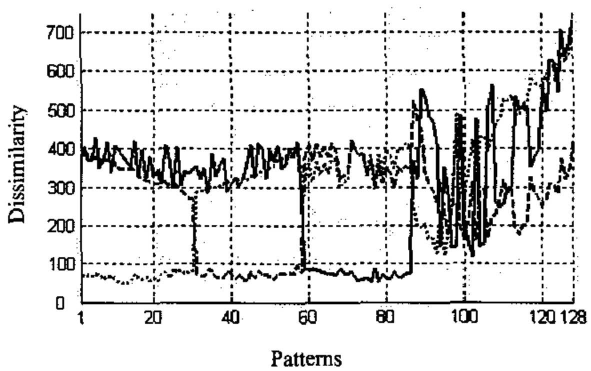

Fig. 6 Prototypes produced by robust-FCM on R-synthetic-L_2 data.

Figures 7 and 8 show the advantage of robust-FCM initialised with the final partition of FCM over the robust-FCM initialised with a randomly generated partition. Actually, in the latter case, the algorithm is less stable for low values of fuzziness.

Finally, Fig. 9 shows that robust-NE-FRC achieves lower performance on R-synthetic-L_2 data.

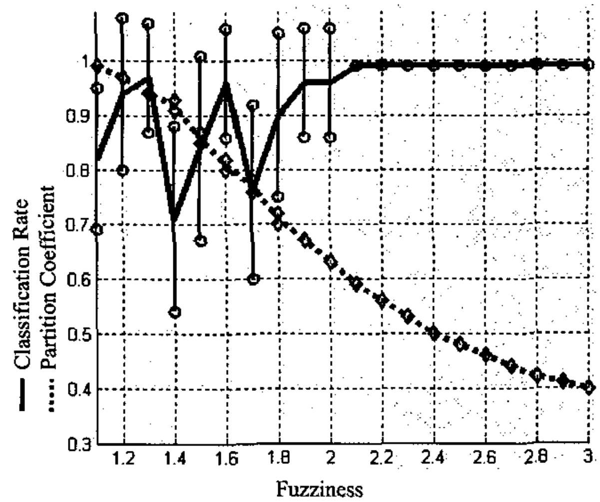

Fig. 7 Effectiveness of robust-FCM on R-synthetic-L_2 data, without FCM initialisation.

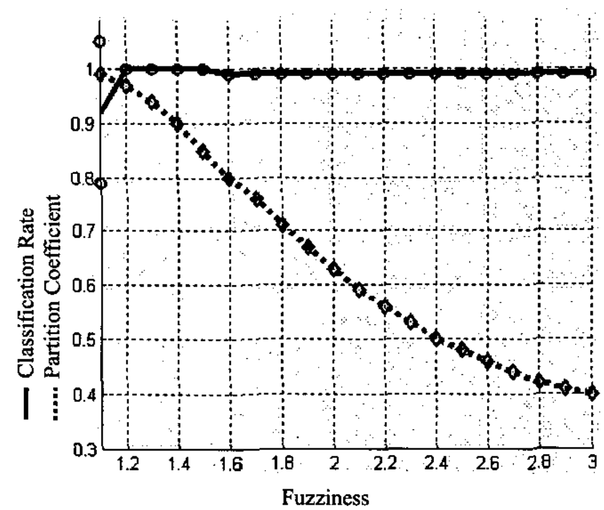

Fig. 8 Effectiveness of robust-FCM on R-synthetic-L_2 data, with FCM initialisation.

C. Scalability

To compare the two algorithms over a significant number of patterns, we increased the size of the synthetic data set (consisting of 128 points) as shown in Fig. 10.a-d. The new patterns of each class (including the noise class) are generated by randomly adding a number of points equal to 50% (i.e., 64 points), 100% (i.e., 128 points), 1000% (i.e., 1280 points) of the original size of the data set, and maintaining the ratio of each class to the total set of points. The added points are extracted, respectively, from the following circles shown in Fig. 10.a, where the centres are C_0 = [367.967, 181.5], C_1

=[200.421, 497.772] and C_2=[528.351, 491.446], and the radius is 80. The noise patterns are randomly extracted from the complementary area in the domain [0, 900]x[0, 800].

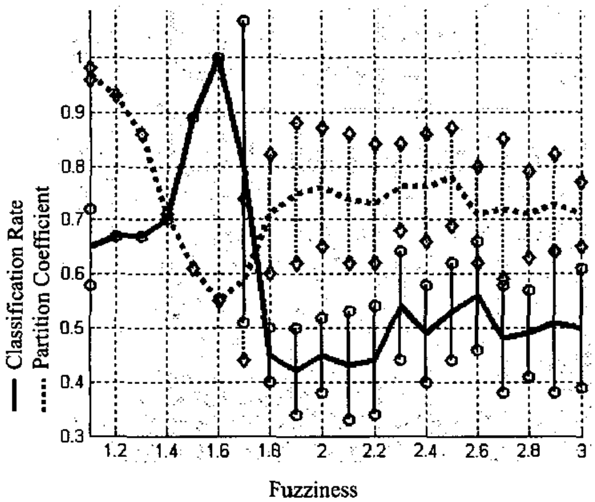

Fig. 9 Effectiveness of robust-NE-FRC on R-synthetic-L_2 data.

Tables I and II show the numerical results of the two algorithms on the three data sets. We chose, as optimal combination of parameters, the values corresponding to the best classification rate (see Fig. 8 and Fig. 9) and (with smaller importance) to the best partition coefficient; then we executed 10 trials of the two algorithms for each data set, adopting the following termination criteria: 0.0001 (Fig. 10.b), 0.001 (Fig. 10.c) and 0.01 (Fig. 10.d).

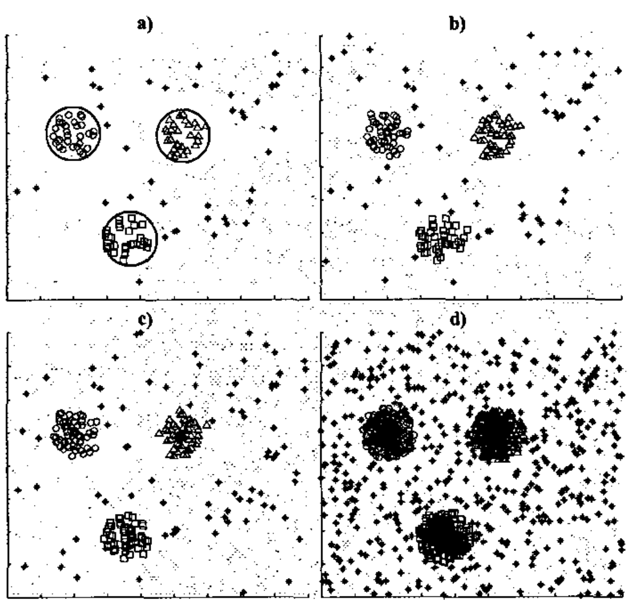

Fig. 10 The synthetic data set with added points in the feature space.

It can be observed that the percentages of correct classifications are near 100% for all tests. However, the partition coefficient (mainly due to the different adopted fuzziness) is much higher in our approach. Finally we note that the numbers of iterations in the two algorithms are similar, but robust NE-FRC has a higher standard deviation.

TABLE I – Results of robust-FCM on the R-"synthetic with added points"-L_2 data sets.

Size of Data Set	Correctly classified points	Part. Coef.	Iterations	Iterations (FCM)	Iterations (robust-FCM)
192	100.0%±0.0%	0.93±0.0	15.1±1.2	6.1±1.2	9.0±0.0
256	99.3%±0.1%	0.91±0.0	22.8±5.7	6.1±0.7	16.7±5.9
1408	98.7%±0.1%	0.93±0.0	19.8±6.5	6.5±1.3	13.3±7.1

TABLE II – Results of robust- NE-FRC on the R-"synthetic with added points"-L_2 data sets.

Data Set	Correctly classified points	Partition Coefficient	Iterations
192	99.5%±0.0%	0.53±0.0	22.3±4.3
256	99.6%±0.0%	0.55±0.0	25.1±11.8
1408	99.9%±0.0%	0.56±0.0	20.9±8.3

V. Conclusions

In this paper we have proposed a robust relational clustering technique based on the fuzzy C-means object clustering algorithm and the concept of Davé's noise cluster. We have shown experimentally that our method compares favourably with the robust-NE-FRC as regards scalability, cluster initialization and sensitivity to parameter variation. Furthermore, our method ensures convergence in all cases.

References

[1] R.J. Hathaway, J.C. Bezdek, "NERF c-means: Non-Euclidean relational fuzzy clustering", Pattern Recognition 27: 429-437, 1994.

[2] R.J. Hathaway, J.C. Bezdek, J.W. Davenport, "On relational data versions of *c*-means algorithms", Pattern Recognition Letters 17: 607-612, 1996.

[3] R.J. Hathaway, J.W. Davenport, J.C. Bezdek, "Relational dual of the c-means clustering algorithms", Pattern Recognition 22 (2): 205-212, 1989.

[4] M. Roubens, "Pattern classification problems and fuzzy sets", Fuzzy Sets and Systems 1: 239-253, 1978.

[5] M.P. Windham,"Numerical classification of proximity data with assignment measures", Journal of Classification 2: 157-172, 1985.

[6] J.C. Bezdek, J. Keller, R. Krishnapuram, N.R. Pal, "Fuzzy Models and Algorithms for Pattern Recognition and Image Processing", Kluwer Academic Publishing, Boston, 1999.

[7] R.N. Davé, "Robust fuzzy clustering algorithms", Second IEEE International Conference on Fuzzy Systems, 28 March-1 April 1993, vol. 2, pp. 1281-1286.

[8] R.N. Davé, S. Sen, "Robust fuzzy clustering of relational data", IEEE Transactions on Fuzzy Systems, Vol. 10, No. 6, pp. 713-727, 2002.

[9] P. Corsini, B. Lazzerini, F. Marcelloni, "Combining Supervised and Unsupervised Learning for Data Clustering", Neural Computing and Applications, Springer Verlag, in press.

[10] P. Corsini, B. Lazzerini, F. Marcelloni, "A new fuzzy relational clustering algorithm based on the fuzzy C-means algorithm", Soft Computing, Springer-Verlag, in press.

[11] X.L. Xie, G. Beni, "A validity measure for fuzzy clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence 13(8): 841-847, 1991.