Greater Knowledge Extraction Based on Fuzzy Logic And GKPFCM Clustering Algorithm

BENJAMÍN OJEDA-MAGAÑA[1], RUBÉN RUELAS[1], FULGENCIO S. BUENDÍA BUENDÍA[2], DIEGO ANDINA[2]
[1]Departamento de Ingeniería de Proyectos – CUCEI
Universidad de Guadalajara
José Guadalupe Zuno No. 48. C.P. 45101 Zapopan, Jalisco
MÉXICO.
[2]Departamento SSR, E.T.S.I Ingenieros de Telecomunicación
Universidad Politécnica de Madrid
ESPAÑA.

Abstract: - This work proposes how to generate a set of fuzzy rules from a data set using a clustering algorithm, the GKPFCM. If we recommend a number of clusters, the GKPFCM identifies the location and the approximate shape of each cluster. These ones describe the relations among the variables of the data set, and they can be expressed as conditional rules such as "If/Then". The GKPFCM provides membership and typicality values from which a knowledge base is generated through fuzzy rules, which can be used for the classification and characterization of new data.

Key-Words: - Knowledge extraction, Fuzzy rules, GKPFCM Clustering, Fuzzy Clustering, Possibilistic Clustering, Gustafson-Kessel Clustering.

1 Introduction

There are a lot of problems in the real life where it is necessary to extract knowledge from a data set. Besides, a common way to express these relations is through conditional rules which can be used to interpret, to verify or even to make decisions about the identified states or characteristics of a system. The automatic extraction of knowledge from a data set is based on several methods, the clustering algorithms is one of them.

With a clustering algorithm, the membership functions of fuzzy sets can be identified, and a set of fuzzy rules can be defined according to them [1-3]. However, the problem with these algorithms is that they provide a partition of the characteristics space of the variables in such a way that every input data belongs to an identified class; and there is no distinction between data of the same class with high and small membership value.

The GKPFCM (Gustafson Kessel Possibilistic Fuzzy c-Means) algorithm [4] is used in this work because it allows a better identification of the clusters, it provides the membership and the typicality values which make a fuzzy and a possibilistic partition of the space respectively, and sets of fuzzy rules can be identified from the results

The membership functions provide a fuzzy partition of the space where data are defined. On the other hand, the typicality values have fewer restrictions among clusters and the sensitivity to noise can be adjusted. Using this characteristic of the possibilistic algorithm several clusters can be identified with their most representative data which gives the possibility to characterize the conclusion at the moment of new data classification. The proposed characterization corresponds to typical values, atypical values or noise.

Some popular clustering algorithm are the Fuzzy c-Means (FCM) [5, 6] and the Possibilistic c-Means (PCM) [7], which can be used to identify the membership functions that will be the support to build a set of fuzzy rules. These clustering algorithms are among the developed hybrid clustering algorithms that have been developed as a more efficient solution to clusters identification in a dataset.

Pal et al [8] proposed the Possibilistic Fuzzy c-Means (PFCM) which provides the membership as those obtained with the FCM and the typicality obtained with the PCM. The PFCM is based in the Euclidian distance, so Ojeda et al [4] proposed an improvement to the algorithm resulting in the Gustafson-Kessel Possibilistic Fuzzy c-Means

(GKPFCM) algorithm, where the Mahalanobis distance and the Gustafson Kessel algorithm [9] are used for the covariance matrix calculus, resulting in the identification of clusters better adapted to the natural distribution of data.

Once clusters have been identified, we propose to use Takagi-Sugeno (TS) fuzzy models [10] for the set of fuzzy rules identified, which allow to classify and to characterize new data that are presented at the input of the model.

This work is organized as follows. Section two presents the clustering algorithms. Section three the fuzzy and possibilistic rules resulting from the application of the GKPFCM to a dataset. And the fourth and last section presents the main conclusions about this work.

2 Clustering Algorithms

The clustering algorithms can find *c* groups in a set of unlabelled data $Z=\{z_1, z_2, \ldots, z_n\}$ into an M dimensional space, that is, the nearest data z_k to a prototype, or center of a group v_i, belong to this group. The membership of each data z_k to the groups depends on the kind of partition of the space where the data set is defined. With a strict partition, a point belongs only to one group; with a fuzzy partition a point has a membership value to each of the groups, where the maximum membership indicates the group that makes the better description of the point and the sum of all the membership values of this point to all the groups is equal to one; and finally, in a possibilistic partition, where this last condition is relaxed, the compatibility of data with a group is represented through a typicality value [11, 12], that is, it takes into account that some data are more representative or typical than the others.

2.1 Fuzzy c-Means Algorithm

The Fuzzy c-Means (FCM) [5, 6] is an algorithm that calculates the membership value of each point (z_k) to the subgroups A_i. This algorithm generates a matrix **U**, with values in the interval [0, 1], corresponding to the memberships and finds the values of the prototypes for all the subgroups. The membership of each point is relative to the distance among the point and the prototypes of the subgroups, that means, a point has the greater membership to the nearest subgroup while the membership value diminish as the distance with the rest of prototypes increases.

2.2 Possibilistic c-Means Algorithm

The Possibilistic c-Means Algorithm (PCM) [7] has as a main characteristic the relaxation of the constraint of the FCM about the sum of the membership values that must be one. This characteristic gives the opportunity to calculate the proximity of data with respect to a prototype, but independent of the other subgroups, and this proximity is represented through a typicality value in the interval [0, 1]. This is the reason why the nearest data to the prototypes are defined as typical. For this algorithm Krishnapuram and Keller [13] recommend to use the FCM algorithm in order to determine the initial values such that the PCM could have a good initialization.

2.3 Gustafson-Kessel Possibilistic Fuzzy c-Means Algorithm

As it was shown in the previous sections, the fuzzy and the possibilistic clustering algorithm provide a different sight about the internal structure of a data set. That is why Pal et al. [14] have proposed a hybrid algorithm based on the FCM and the PCM in order to get a more robust clustering algorithm; this algorithm was called Fuzzy Possibilistic c-Means (FPCM). Later, Pal et al. [8] have corrected an inconvenience with the typicality values of this algorithm and they rename the algorithm as Possibilistic Fuzzy c-Means (PFCM).

Ojeda et al. [4] have made an improvement to the PFCM. This consists on the kind of distance D_{ikA} used and the method to calculate it. The PFCM uses the Euclidean distance whereas the Gustafson-Kessel Possibilistic Fuzzy c-Means (GKPFCM) uses the Mahalanobis distance. Besides, the calculus of the distance is based on the Gustafson-Kessel method [9] allowing calculating an adaptive distance such that clusters with different geometry can be identified in a data set.

In the Gufstason-Kessel algorithm, the matrix A_i is defined by (1),

$$|A_i| = [\rho_i \det(F_i)]^{1/n} F_i^{-1} \tag{1}$$

where n is the number of characteristics, ρ_i the volume and F_i the covariance matrix of the group i. The calculus of F_i includes a proposal by Babuska et al. [15]. This one is shown in the equations (3) and (4).

$$F_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m (z_k - v_i)(z_k - v_i)^T}{\sum_{k=1}^{N} (\mu_{ik})^m} \tag{2}$$

$$F_i = (1-\gamma)F_i + \gamma\left(F_0\right)^{1/n} I \qquad (3)$$

$$F_i = \left[\phi_{i,1}, \ldots, \phi_{i,n}\right] \cdot diag\left(\lambda_{i,1}, \ldots, \lambda_{i,n}\right) \cdot \left[\phi_{i,1}, \ldots, \phi_{i,n}\right]^{-1}, \quad 1 \le i \le c \qquad (4)$$

Therefore, the calculus of distance in the GKPFCM algorithm is based on the equations (1) to (4). These equations with (5) to (7) provide the complete GKPFCM algorithm.

$$\mu_{ik} = \left(\sum_{j=1}^{c}\left(\frac{D_{ikA}}{D_{jkA}}\right)^{2/(m-1)}\right)^{-1}, \quad 1 \le i \le c; \quad 1 \le k \le n \qquad (5)$$

$$t_{ik} = \frac{1}{1 + \left(\frac{b}{\delta_i} D_{ikA}^2\right)^{1/(\eta-1)}}, \quad 1 \le i \le c; \quad 1 \le k \le n \qquad (6)$$

$$v_{ik} = \sum_{k=1}^{N}\left(a\mu_{ik}^m + bt_{ik}^\eta\right) z_k \Big/ \sum_{k=1}^{N}\left(a\mu_{ik}^m + bt_{ik}^\eta\right), \quad 1 \le i \le c; \quad 1 \le k \le n \qquad (7)$$

2.3 Example

In order to show the performance of the GKPFCM algorithm, in this example the algorithm is applied to a data set generated with two ellipsoidal groups and an additive noise. The data set $\mathbf{Z}_{1200}$ contains 1200 data in total; 400 data for each group and 400 noisy data. The mean value of one group is $(0, 2.5)^T$, and the other one is in $(0, 0)^T$. The noise is distributed in a uniform way in the space [-12, 12]X[-7, 10].

The groups of this example have a covariance matrix identical and it is given by (8). The Fig. 1a shows the distribution of data Z_{1200}.

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 4.47 & 0 \\ 0 & 0.22 \end{bmatrix} \qquad (8)$$

For comparison, three clustering algorithms have been applied to the data set: the GKFCM, the PFCM and the GKPFCM. The results are in the Fig. 1b, Fig. 1c, and Fig. 1d. Table 1 shows the prototypes generated with these algorithms.

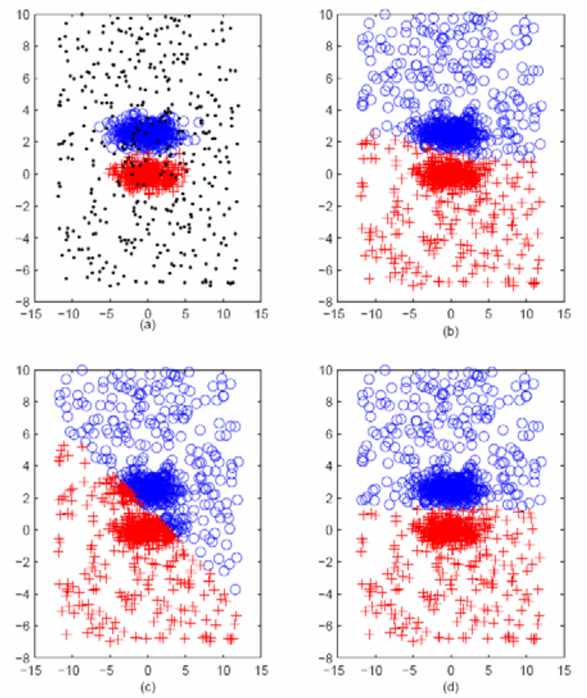

Figure 1: (a) Dataset $\mathbf{Z}_{1200}$. Partition of the space with the algorithm: (b) GK, (c) PFCM, and (d) GKPFCM.

Table 1: Prototypes produced by GKFCM, PFCM AND GKPFCM.

Results for $\mathbf{Z}_{1200}$	
GKFCM (ρ=1, M=2, γ=2)	
v_1	v_2
0.2126	-0.1688
3.0565	-0.3061
PFCM (a=1, b=5, m=2, η=2)	
v_1	v_2
0.2005	-0.2276
1.7506	0.7647
GKPFCM (a=1, b=5, m=2, η=2)	
v_1	v_2
0.0561	0.0027
2.8887	-0.2421

3 Fuzzy Rules

The GKPFCM algorithm provides the membership and the typicality values which can be the base to build a model consisting of fuzzy rules derived from the relation among variables for each cluster, and the membership functions of the fuzzy sets approximated from the shape of the clusters.

Due to the flexibility of the possibilistic algorithms to take into account or to ignore the noise, it is possible to make the adjustment of the GKPFCM parameters such that the membership functions identified from the typicality values are centered in the most representative values of each cluster, ignoring or giving little importance to the furthest data. For example, thresholds can be used in order to identify the class of each new data, but also to describe it about its compatibility with the class assigned. From the highest threshold, the first group corresponds to data with the greatest typicality, the second group is composed of data with a smaller typicality, but greater than the second threshold, the third group contains data with typicality even smaller and this last group can be divided in subgroups. The thresholds depend on the knowledge and intuition about the range of values that can be considered as typical for a given problem.

From here the interest to use the typicality values as a mean to build a fuzzy model as a way to get more information about new data that go into the system. If thresholds are given for both models, the total space is partitioned as shown in Fig. 2. As a result of the bounded space, each new data entering to the system will be identified as an element of a class and it will be also characterized as typical, atypical or noise. The fuzzy rules representing the knowledge are similar to the rules defined later.

In this work it is proposed to generate fuzzy rules with hybrid algorithms, such as the GKPFCM, because the membership values of the **U** matrix and the typicality values of the **T** matrix are available. However, the fuzzy rules are separated into two models as a way to have a better interpretation of the data processed with both models. In this case one model is built with the memberships and the other one with the typicalities. The inference method elected for the rules is the one proposed by Takagi-Sugeno [10] where the conclusions are constant values or zero order functions.

Fuzzy rules have one or several antecedents and one consequent. The antecedents depend on the input variables and the space where each rule is

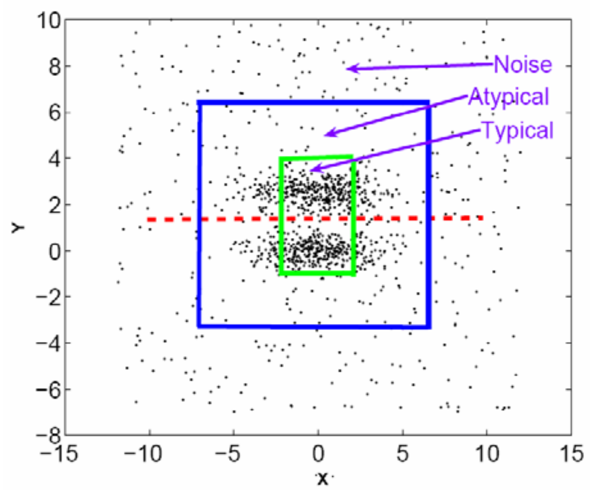

Figure 2: Decision frontier for the fuzzy and the possibilistic models.

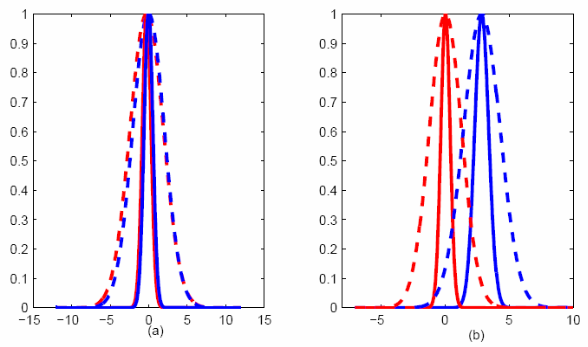

Figure 3: Adjusted gaussian functions from the orthogonal projections; solid lines for the $\mathbf{U}_{Z1200}$ values and dashed lines for the $\mathbf{T}_{Z1200}$ values. (a) Axe x. (b) Axe y.

defined. This space is totally defined by the membership functions of the antecedent fuzzy sets. So, when learning fuzzy rules by the GKPFCM clustering algorithms, the fuzzy values of the antecedents result from the orthogonal projection of the **U** or **T** matrix over the universe of discourse of the variables. For two groups, as in the example, the corresponding fuzzy rules are:

R_{11}: **If x** is (*Very*) **Small** and
y is (*Very*) **Small Then** z =c_1
R_{12}: **If x** is (*Very*) **Small** and
y is (*Very*) **Big Then** z =c_2

Once the orthogonal projections are available, the next step is the calculus of the membership functions for all the rules. These are obtained by a parametric adjustment of the kind of mathematical function selected for the membership functions such as the triangular, the trapezoidal, the gaussian or any other monotone function. In the particular case of this example the functions are approximated by gaussians. The parameters of this function are the mean (μ) and the standard deviation (σ). The mean value is directly determined from the prototypes, so it suffices to approximate the (σ) value.

When all the σ have been calculated, the result are gaussian functions as those of the Fig. 3a and Fig. 3b. In the same figures there are the functions resulting from the $\mathbf{U}_{Z1200}$ matrix, and the $\mathbf{T}_{Z1200}$ matrix.

As can be seen in Fig. 3a and Fig. 3b, the functions generated with the typicality values are narrower than those generated with the membership values. For this reason the linguistic term of the functions derived from the membership values is *small* and this is *very small* for the functions adjusted to the typicality values. The linguistic modifier *very* acts in the same way as proposed by Zadeh even if it does not correspond exactly to his definition, that is, *very small* = $(small)^2$. In the same way, the linguistic terms for the fuzzy sets of the Fig. 3b are *Small* and *Big* when the functions have been adjusted with the membership values and *Very Small* and *Very Big* on the other case.

4 Conclusions

The goal of this work has been the generation of fuzzy rules from the membership and the typicality values of the GKPFCM, as a way to classify and to characterize new data. Two models were generated, which can be considered as redundant knowledge that provides two different interpretations of data, useful for their characterization. This conducts us to obtain a bigger and better knowledge from a data set. Besides, the selected method to get the clusters has been the GKPFCM which, as proved in this work, carries out a good identification of the clusters. This work proposes an improvement to the knowledge represented by means of fuzzy rules, since there are general rules generated through the membership values, but also more specific rules resulting from using typicality values.

References

[1] M. Sugeno and T. Yasukawa, A Fuzzy-Logic-Based Approach to Qualitative Modeling, *IEEE Trans. on Fuzzy Systems*, Vol 1, No. 1, 1993, pp. 7-31.

[2] R. R. Yager and D. P. Filev, Generation of fuzzy rules by mountain clustering, *Journal of Intel. and Fuzzy Systems*, Vol 2, 1994, pp. 209-219.

[3] S.L Chui, Fuzzy model identification based on cluster estimation, *Journal of Intel. and Fuzzy Systems*, Vol 2, No. 3, 1994, pp. 7-31.

[4] B. Ojeda Magaña, R. Ruelas, MA. A. Corona Nakamura and D. Andina, An improvement to the possibilistic fuzzy c-means clustering algorithm, *World Automation Congress (WAC) 2006*, Budapest, Hungría, July 24-26, 2006.

[5] J.C. Dunn, A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well separated Clusters, *Journal of Cybern*, Vol. 3, 1974, pp. 95-104.

[6] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum, 1981.

[7] R. Krishnapuram and J. Keller, A possibilistic approach to clustering, *Int. Conf. on Fuzzy Systems*, Vol. 1, No. 2, 1993, pp. 98-110.

[8] Nikhil R. Pal, Kuhu Pal, James M. Keller, and James C. Bezdek, A Possibilistic Fuzzy c-Means Clustering Algorithm, *IEEE Trans. on Fuzzy System*, Vol. 13, No. 4, 2005, pp. 517-530.

[9] E. E. Gustafson and W. C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, *Proc. of the IEEE CDC*, San Diego, CA, 1979, pp. 761-766.

[10] T. Takagi and M. Sugeno, Fuzzy identification of systems and its application to modelling and control, *IEEE Trans. on Systems, Man, and Cybernetics*, Vol.15, No. 1, 1985, pp. 116-132.

[11] M.-J. Lesot, L. Mouillet, and B. Bouchon-Meunier, Fuzzy prototypes based on typicality degrees, *Advances in Soft Computing*, Springer, 2005.

[12] M.-J. Lesot, Similarity, typicality and fuzzy prototypes for numerical data, *6th European Congress on Systems Science*, Workshop Similarity and resemblance, 2005.

[13] R. Krishnapuram and J. Keller, The possibilistic c-Means algorithm: Insights and recommendations, *Int. Conf. on Fuzzy Systems*, Vol. 4, No. 3, 1996, pp. 385-393.

[14] N. R. Pal, K. Pal and J. C. Bezdek, A mixed c-means clustering model, *IEEE Int. Conf. on Fuzzy Systems*, Spain, 1997, pp.11-21.

[15] R. Babuska, P. J. van der Veen, and U. Kaymak, Improved covariance estimation for Gustafson-Kessel clustering, *Int. Conf. on Fuzzy Systems*, 2002, pp. 1081-1085.