Better Interpretation of Numerical Data Sets by Relative and Absolute Typicality of Fuzzy Clustering Algorithms

B. Ojeda-Magaña[1], R. Ruelas[1], M.A Corona-Nakamura[2], and D. Andina[3]

[1] Departamento de Ingeniería de Proyectos-CUCEI, Universidad de Guadalajara.
José Guadalupe Zuno No. 48, C.P. 45101 Zapopan, Jalisco, México.
benojed@hotmail.com, rruelas@newton.dip.udg.mx.
[2] Departamento de Ciencias Computacionales-CUCEI, Universidad de Guadalajara.
Revolución 1500. C.P. 44840 Guadalajara, Jalisco, México.
[3] Group for Automation in Signals and Communications (GASC)
Universidad Politécnica de Madrid.

Abstract. In this work we take the concept of typicality from the cognitive and psychological point of view, and we apply their meaning to the interpretation of numerical data through fuzzy clustering algorithms, looking for a better profit of the typicality in the clusters found, and a better interpretability of the processed data. With the Fuzzy c-Means clustering algorithm (FCM) we get a relative typicality (membership degree), and with the Possibilitistic c-Means (PCM) an absolute typicality (typicality value). Nowadays, some hybrid clustering algorithms have been developed that combine both algorithms and where both typicalities are available; one such algorithm is the Possibilistic Fuzzy c-Means (PFCM), which is the basis of the algorithm used in this work. Thus, it is important to make better use of both typicalities, for learning and analysis of data. To get some results we use a synthetic data set. The results clearly show the advantages of the information obtained about the data set, taking into account the different meaning of typicalities and the availability of both values with the clustering algorithm used.

Keywords: Clustering, Relative Typicality, Absolute Typicality.

1 Introduction

The objective of clustering algorithms is to find an internal structure in a numerical data set into n different subgroups, where the members of each subgroup have a high similarity with its prototype (centroid, cluster center, signature, template, code vector) and a high dissimilarity with the prototypes of the other subgroups. This justifies the existence of each one of the subgroups.

The clustering algorithms help us to get a simplified representation of a numerical data set into n subgroups, as a way to get a better comprehension and knowledge about data. Also, the clustering algorithms that partition in a hard, fuzzy, probabilistic or possibilistic way a given space, according to a data set and

© J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad and H. Sossa (Eds.)
Special Issue in Advances in Pattern Recognition
Research in Computing Science 44, 2009, pp. 157-166

after a learning process, provide a set of prototypes as the most representative elements of each subgroups.

In this work we propose to take the interpretation of the typicality concept according to the cognitive and psychological point of view, such that it is possible a better interpretation of data. Rosch *et al* [1] proposed a theory of prototypes for the classification of objects that belong to a semantic characteristic, taking into account their proximity to a prototype, according to a given criterion. So, in each category there must be an internal resemblance among category members, and an external dissimilarity meaning that the similarity with the members of the other categories is low.

A prototype is based on the typicality notion; all members belonging to the same category do not represent it in the same way, that is, some members are more typical than others. Berlin *et al* [2] were among those who first recognized this variation inside the categories, when they discovered the focal color which is a psychological definition for color identification. They show a similarity among different languages in color categorization, the so-called "basic colors", that is, each color is represented by the best example (some nuances of colors are more representative than others), and inside a category of color there is no other color. For example, the scarlet color is not a basic color because it is in the red color category. Thus, the colors are not uniform and they are represented by a center or better example and peripheral members. However, if we are interested knowing the edge of each category of color, such as the threshold where the color stops being red but begins to be orange, for example, we can use fuzzy sets. Zadeh [3] has proposed the theory of fuzzy sets in order to solve the problem of boundaries. With fuzzy sets, the members of a particular category have a membership degree in the [0, 1] interval, where 1 is assigned to the most representative members, and 0 to the elements that are not members of the category.

In this work we use the concepts of typicality and membership degrees in order to categorize linguistic concepts, looking for a better understanding of the information extracted from a numerical data set through the Possibilistic Fuzzy c-Means (PFCM) clustering algorithm.

This paper is organized as follows: Section 2 presents the typicality and vagueness concepts. Section 3 contains the relative and absolute typicalities, as well as the PFCM clustering algorithm. Section 4 presents the concepts of typicality applied to a synthetic numerical data set. Finally, Section 5 presents the main conclusions.

2 Typicality and vagueness

In the search of prototypes Hampton [4] mentions the phenomena of vagueness, typicality, genericity and opacity. In this work we use vagueness, and typicality, both different but of interest as they help us better define the characteristics of the concepts. The typicality alludes to a measure in which the objects under study are considered good examples of the concept [4]. For example, in the category of birds, dove is a typical case as it has the following characteristics: it can fly, it has feathers, it puts eggs, and it has a nest in a tree. On the other side, an atypical case is the penguin because it satisfies only some characteristics, but not all. For this example, the prototype has the most common characteristics or the mean values of these characteristics.

The categories of some concepts can be vague or fuzzy, i.e. there exist objects whose membership to the category is uncertain, and this is not due to a lack of knowledge, but to the lack of a clear rule defining the edges of the categories [5]. Some classical examples are the adjectives: *high*, or *red*; or the nouns: *vegetable*, or *chair*. Vagueness is mostly a question of truth (yes or not), and it represents a measure of correspondence of an object with a conceptual category. For some categories the edges are defined in an easier way, such as the category of birds. However, for other categories, as for the adjective *high*, the edges are not so easy to define, and therefore a membership degree in [0, 1] is used.

For a better understanding of the differences between typicality and vagueness, take an example of Osherson *et al* [4], in order to appreciate the difference between typicality and vagueness, they note that *...many people believe that penguins are an atypical case of birds, only a few doubt that they are birds in reality; in this case the typicality is involved, but not the vagueness.*

The membership degree of the penguin and the dove to the category of birds is 1. However, the dove is more typical than the penguin. For the concept of *high* there is no example with maximum typicality, due to the fact that in some contexts height can be increased infinitely. On the other hand, the membership degrees of the concept *high* represents the variation on the certainty degrees using values in 0 and 1, both included, and they provide the edge to determine at which height something is "*high*" and at which height it is not.

3 Relative and Absolute Typicality in the Clustering Algorithms

The partition clustering algorithms have a great similarity to the theory of prototypes, although the latter is related to building categories about concepts and the clustering algorithms focus on the classification of numerical data; however both approaches have the same objective.

In this section we give a better interpretation of the typicality of the fuzzy clustering algorithms, based on a psychological and cognitive interpretation, as presented in the previous section, as a way to gain a greater knowledge than usual from numerical data sets.

3.1 Fuzzy c-Means Clustering Algorithm (Relative Typicality)

The first to use fuzzy sets for clustering was Ruspini [6]. After that, Dunn [7] proposed the first fuzzy clustering algorithm named Fuzzy c-Means (FCM) with the parameter of fuzziness m equal to 2. Later on Bezdek [8] generalized this algorithm. The FCM is an algorithm where the membership degree of each point to each fuzzy set A_i is calculated according to its prototype. The sum of all the membership degrees of each individual point must be equal to one. Therefore the degree of membership to a particular fuzzy set is influenced by the position of all the prototypes of the fuzzy sets, and that is the reason why Pal *et al* [9] interpret the membership as a relative typicality.

With the FCM, the calculus of the membership degree of a point z_k to the fuzzy sets A_i is inversely proportional to the relative distance of this point to the prototypes (centers) of the fuzzy sets. Pal *et al* [10] show a deficiency of the algorithm when there are several equidistant points from two prototypes, as the membership degrees to both fuzzy sets are the same, but the distance to the prototypes is different; one point is further than the other. These data must be handled with care as they do not represent both prototypes in the same way. Another disadvantage of the FCM algorithm is its sensitivity to noise, or points far away from a concentration of prototypes.

3.2 Possibilistic c-Means Clustering Algorithm (Absolute Typicality)

The Possibilistic c-Means (PCM) clustering algorithm was proposed by Krishnapuram and Keller [11], and its principal characteristic is the relaxation of the restriction that gives the relative typicality property of the FCM. As a consequence, the PCM help us to calculate a similarity degree between data points and each one of the prototypes; value known as absolute typicality or simply typicality [9]. The nearest points to a prototype are identified as typical, whereas the furthest points as atypical, and noise if their typicality is zero or almost zero [12]. The PCM is very sensitive to the initial value of its parameters. Also, to avoid the coincidence of several prototypes, it is convenient to use the modified objective function proposed by Tim *et al* [13–15], which contains a restriction resulting in a repulsion of the prototypes and avoiding prototypes located at the same place.

3.3 Categories, Typicality and Clustering Algorithms

The prototypes are selected as the best examples, according to a given criterion, to represent groups, and they have the most important characteristics. In the case of birds, for example, the dove is more typical than the ostrich and the penguin, because it has more characteristics of a bird. However, ostriches and penguins are members of the category of birds. Therefore, there is an internal resemblance among the members of a group, and an external dissimilarity to the members of other categories, even when several categories share some characteristics, as it happens with birds and reptiles, as both kinds of animals reproduce by eggs. However, each group has its own characteristics that define them as members of a particular group, and different to the others.

A similar situation happens with a numerical data set, that is, it is possible to take into account an external dissimilarity and an internal resemblance.

- *External dissimilarity*: it results from fuzzy clustering algorithms, because the membership degree of a data point to a group depends on the membership degrees to the rest of groups. Finally, the data point is considered a member of the group to which it has the maximum membership degree. In other words, it belongs to the nearest group.

- *Internal resemblance*: it results from algorithms such as the PCM, and represents the resemblance between a data point and a prototype. For this reason it is possible to establish thresholds in order to identify typical, atypical and noisy data.

Thus, the FCM and the PCM algorithms provide information about a numerical data set, and in this way they help us to better understand its structure. Pal *et al* [16] have proposed a hybrid algorithm, the PFCM, that provides of the advantages of the FCM and PCM algorithms, and it avoids some problems of these algorithms used separately. This algorithm is described in the next Section.

3.4 Hybrid prototypes

Pal *et al.* [9] have proposed a hybrid algorithm that uses membership degrees (relative typicality), as well as typicality values (absolute typicality), trying to take profit of both algorithms in a single one. However, due to a problem with a restriction on the typicality values, they modified the algorithm and proposed the Possibilistic Fuzzy c-Means (PFCM) algorithm [16].

The PFCM algorithm is based on the Euclidian distance and the identified clusters are constrained to spherical shapes. Attempting to find a more general result, we decided to use the Gustafson Kessel Possibilistic Fuzzy c-Means (GKPFCM) algorithm proposed by the authors [17], which is based on the Mahalanobis distance according to the Gustafson and Kessel method [18], such that the identified groups are better adapted to the distribution of the data set.

`GKPFCM clustering algorithm`

I Initialize the prototypes (centers) $v_i, i = 1, .., c$. These are regularly obtained on a random basis.

II Find the value of the parameter δ_i. To do this we would run the FCM clustering algorithm [19] and then use Equation (1) proposed by Krishnapuram and Keller [11] [20]:

$$\delta_i = K \frac{\sum_{k=1}^{N} \mu_{ik}^{m} \|z_k - v_i\|_A^2}{\sum_{k=1}^{N} \mu_{ik}^{m}} \tag{1}$$

$K > 0$, the most common option is $K = 1$.

III Propose new values for parameters that are added to the algorithm PFCM, they are (ρ_i, β, γ), regularly the first two parameters remain constant $\rho_i = 1$, $i = 1, .., c$, and $\beta = 10^{15}$, the only parameter that is modified is the value of γ, and it can take values between 0 and 1.

IV Propose values of the parameters a, b, m, and η. These parameters play an important role in the calculation of the prototypes and the membership and typicality values.

V Calculate the covariance matrices for each group:

$$F_i = \frac{\sum_{k=1}^{N}(\mu_{ik})^m (z_k - v_i)(z_k - v_i)^T}{\sum_{k=1}^{N}(\mu_{ik})^m}, \quad 1 \le i \le c. \tag{2}$$

and estimate the covariance as proposed by Babuska:

$$F_i = (1-\gamma)F_i + \gamma \det(F_0)^{\frac{1}{n}} I \tag{3}$$

Extract the eigenvalues λ_{ij} and eigenvectors ϕ_{ij}, matrix F_i, and find all $\lambda_{i,max} = max_j \lambda_{ij}$ and $\lambda_{i,max} = \lambda_{ij}/\beta$, $\forall_j$ which satisfies $\lambda_{i,max}/\lambda_{i,j} \ge \beta$.

Finally F_i is rebuilt with the equation (4)

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

VI Calculation of distances:

$$D_{ikA_i}^2 = (z_k - v_i)^T [\rho_i \det(F_i)^{-1/n} F_i^{-1}](z_k - v_i) \tag{5}$$

VII Determine the membership matrix $U = [\mu_{ik}]$ using equation (6)

$$\mu_{ik} = \left(\sum_{j=1}^{c} \left(\frac{D_{ikA_i}}{D_{jkA_i}} \right)^{2/(m-1)} \right)^{-1}, \quad 1 \le i \le c; \ 1 \le k \le n \tag{6}$$

VIII Determine the typicality matrix $T = [t_{ik}]$ using equation (7)

$$t_{ik} = \frac{1}{1 + \left(\frac{b}{\gamma_i} D_{ik_A}^2 \right)^{1/(\eta-1)}}, \quad 1 \le i \le c; \ 1 \le k \le n \tag{7}$$

IX Modify the v_i prototypes according to equation (8).

$$v_i = \sum_{k=1}^{N} (a\mu_{ik}^m + bt_{ik}^\eta) z_k \Big/ \sum_{k=1}^{N} (a\mu_{ik}^m + bt_{ik}^\eta), \quad 1 \le i \le c. \tag{8}$$

X Verify that this error is within the proposal tolerance ε.

$$\|V_{nuevo} - V_{viejo}\|_{error} \le \varepsilon$$

XI If the error is greater than ε return to step **V**.

4 Application of the Typicality Concept to Numerical Data Sets

For this work we use a synthetic numerical data set [21]. For that, two Gaussian clouds with 400 points were generated. The covariance matrix is the same for both subsets. Forty noisy points were added at the corners of the subsets (see Fig. 1a).

$$\sum_1 = \sum_2 = \begin{bmatrix} 4.47 & 0 \\ 0 & 0.22 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 2.5 \end{bmatrix}$$

In order to get a better knowledge from a data set, the following algorithm was proposed:

I Propose the parameters a, b, m, and η, and the number of c clusters before the execution of the GKPFCM algorithm.
II Run the GKPFCM algorithm to estimate the relative typicality provided by the U matrix, and the absolute typicality from the T matrix.
III From the U values, providing the external dissimilarity, find the boundaries among the clusters.
IV From the T values providing an internal dissimilarity, and using thresholds, it is possible to differentiate among typical, atypical, or noise.

To reduce the effects of noise in the prototype it is necessary to make a good choice of parameters a, b, m, and η for the GKPFCM algorithm. The parameters a and b have a great influence on the calculation of the prototype. Pal et al [16] recommended a value of b greater than the value of a, such that the prototypes are more influenced by the membership values. On the other hand, it is recommended a small value for η and a value greater than 1 for m. Nevertheless, a too high value of m reduces the effect of membership of data to the clusters, and the algorithm behaves as a simple PCM. For this work, we use the values $a = 1, b = 5, m = 2$, and $\eta = 2$ for the GKPFCM algorithm. The estimated values for the prototypes are: $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0.1659 & -0.0112 \\ 0.2273 & 2.5721 \end{bmatrix}$ (see Fig. 1b).

The U matrix provides the external dissimilarity, and it helps to discriminate the clusters, that is, the maximum value of each vector indicates the cluster to which each data point belongs. In Fig. 2a we can see that the membership degree of noisy data is very high, as a consequence of the relation of data with the prototypes. In Fig. 3a we use the external dissimilarity to separate clusters with a straight line.

The T matrix provides the internal dissimilarity which is used for the partition of data, and to know the data within each cluster. In this case we are interested how near or far the data points are from the prototypes. Thus, in this work we propose a threshold such that typical and noise data can be differentiated. In Fig. 3b we use a threshold of 0.01 for dividing the typicality T matrix into two submatrices (T_1, T_2), T_1 for data with typicality values greater or equal

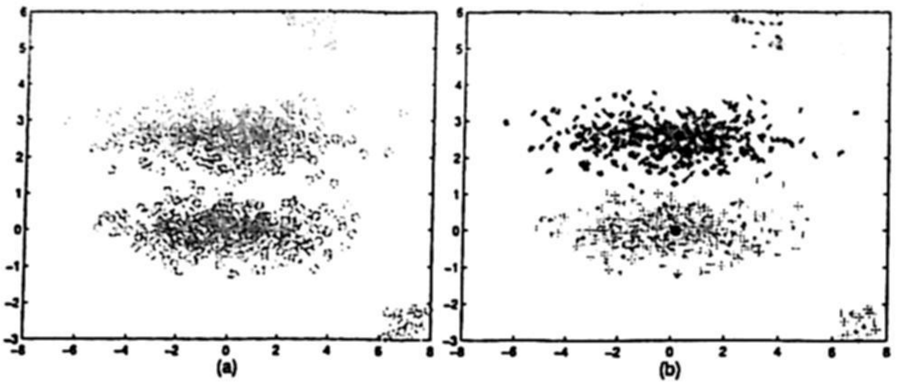

Fig. 1. (a) Synthetic data set. (b) Partition of the data set with the GKPFCM algorithm.

than the threshold, whereas T_2 for the other data, or data considered as noise as their typicality values are below the proposed threshold. Data of T_2 have a very low typicality values as they are far away from the prototype. In this paper, the threshold value has been proposed empirically. However, this must be a very low value as we try to identify data that really represent the corresponding cluster, and to eliminate those data considered as noise.

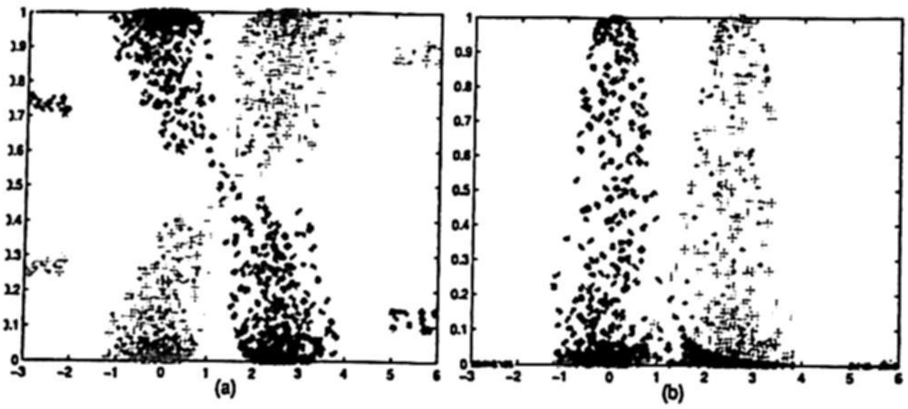

Fig. 2. ((a) Projection of the membership degrees of the U matrix, and (b) Projection of the typicality values of the T matrix.

5 Conclusions

Categorizing data into concepts based on the theory of prototypes, allows us to better understand the problem of classification and its application to a numerical data set. The external dissimilarity among data points of a numerical data set is evaluated with algorithms that use the relative typicality, and the internal

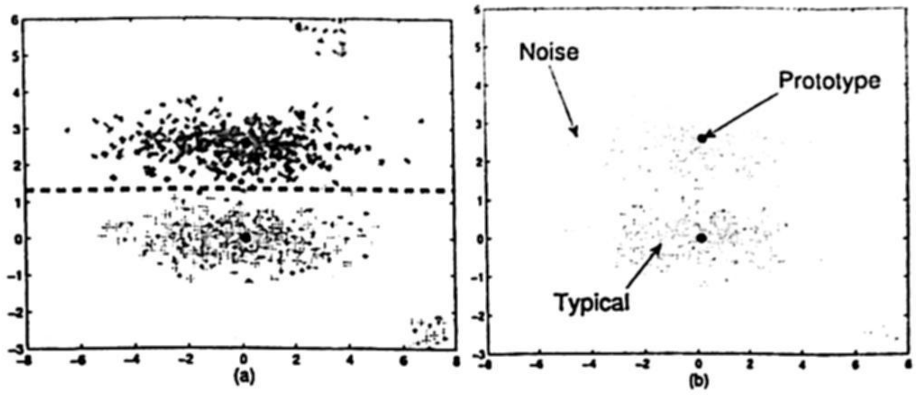

Fig. 3. Boundary generated from the external dissimilarity. (b) Separation of clusters using the internal dissimilarity (see prototypes, typical data, and noise).

resemblance is evaluated with algorithms that use absolute typicality. Using both typicalities through a single algorithm, as the GKPFCM, we can get more information about the internal structure of a numerical data set. In this work we have tried to relate the classification made by human beings and the automatic algorithms. That is the reason we have tried to meld the theory of prototypes and the partitional clustering algorithms.

References

1. Rosch E; and Mervis C. Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.
2. Berlin B; and Kay P. *Basic Color Terms: Their Universality and Evolution.* University of California Press, Berkley CA, 1968.
3. Zadeh L. A. Fuzzy sets. *Information and Control*, pages 338–353, 1965.
4. Osherson D; and Smith E. E. On typicality and vagueness. *Cognition*, 64:189–206, 1997.
5. Hampton J. A. Typicality, graded membership, and vaguess. *Cognition*, 31:355–386, 2007.
6. Ruspini E. Numerical method for fuzzz clustering. *Inf. Sci. 2*, pages pp. 319–350, 1970.
7. Dunn J. C. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, Vol 3:pp. 32–57, 1973.
8. Bezdek J. C. *Pattern Recognition With Fuzzy Objective Function Algorithms.* Plenum Press, New York, 1981.
9. Pal N. R; Pal S. K; and Bezdek J. C. A mixed c-means clustering model. In *IEEE International Conference on Fuzzy Systems, Spain*, 11-21, 1997.
10. Pal N. R; Pal S. K; Keller J. M; and Bezdek J. C. A new hybrid c-means clustering model. In *Proc. of the IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE-04, I. Press, Ed.*, 2004.
11. Krishnapuram R; and Keller J. A possibilistic approach to clustering. *International Conference on Fuzzy Systems*, Vol. 1, No. 2:pp. 98–110, 1993.

12. Ojeda-Magaña B, Ruelas R, Buendía F.S, and Andina D. A greater knowledge extraction coded as fuzzy rules and based on the fuzzy and typicality degrees of the gkpfcm clustering algorithm. *Intelligent Automation and Soft Computing*, Vol 15, no. 4:555–557, 2009.
13. Timm H; Borgelt C; Döring C; and Kruse R. Fuzzy cluster analysis with cluster repulsion. In *presented at the Euro. Symp. Intelligent Technologies (EUNITE), Tenerife, Spain*, 2001.
14. Timm H and Kruse. A modification to improve possibilistic fuzzy cluster analysis. In *Conference Fuzzy Systems, FUZZ-IEEE, Honolulu, HI, USA*, 2002.
15. Timm H; Borgelt C; Döring C; and Kruse R. An extension to possibilistic fuzzy cluster analysis. *Fuzzy Sets and systems*, Vol 147, No 1:pp 3–16, 2004.
16. Pal N. R; Pal S. K; Keller J. M; and J. C Bezdek. A possibilitic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, 13, no. 4:517–530, 2005.
17. Ojeda Magaña B; Ruelas R; M. A. Corona Nakamura; and Andina D. An inprovement to the possibilistic fuzzy c-means clustering algorithm. In *Image Procesing and Biomedicine*, TSI Press Series on Intelligent Automation and Soft Computing. Vol 20, pp 585-592, 2006.
18. Gustafson E. E and Kessel W. C. Fuzzy clustering with a fuzzy matriz de covarianza. *Proceedings of the IEEE CDS, San Diego CA*, pages 761–766, 1979.
19. Krishnapuram R; Bezdek J. C, Keller J and Pal N.R. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing.* Boston, London, 1999.
20. Krishnapuram R. and Keller J. The possibilistic c-means algorithm: Insights and recommendations. *International Conference on Fuzzy Systems*, Vol 4, No 3:pp 385–393, 1996.
21. Marie-Jeanne Lesot and Rudolf Kruse. Gustafson-kessel-like clustering algorithm based on typicality degrees. In *Information Processing and Management of Uncertainty in Knowledge-based Systems*, 2006.