A New Proposal for a Granular Fuzzy C-Means Algorithm

Elid Rubio and Oscar Castillo

Abstract Fuzzy clustering algorithms are able to find the centroids and partition matrices, but are predominantly numerical, although each cluster prototype can be considered as a granule of information it continues to be a numeric value, in order to give a similar representation structure data. Granular theory and clustering algorithms can be combined to achieve this goal, resulting in granular prototypes and granular matrices of belonging and a more reflective data structure.

1 Introduction

Fuzzy Clustering algorithms [2, 5, 10–12, 14] are popular and widely used in different areas of research like pattern recognition [2], data mining [6], classification [8], image segmentation [16, 18], data analysis and modeling [3] among others, obtaining good results in these implementations. The popularity of this kind of algorithms is due to the fact that allow a datum to belong to different data clusters into a given data set, the main objective of the fuzzy clustering algorithms are find interesting patterns or group of data that share similar characteristics into a given data set.

In a general point of view the process of clustering is considered as a granular information process [4, 15], but the information granule is represented by prototypes (centers of clusters) and partition matrices (matrices of belonging) which are represented by numerical values. Due to this the data structure is not too reflective. In order to make the data structure more reflective, improvements of clustering algorithms like the FCM [2] and PCM [10, 11] in combination with Interval Type-2 Fuzzy logic techniques [9, 13] and of this combination arise the IT2FCM [7, 17]

E. Rubio · O. Castillo (✉)
Tijuana Institute of Technology, Tijuana, Mexico
e-mail: ocastillo@tectijuana.mx

E. Rubio
e-mail: elid.rubio@hotmail.com

P. Melin et al. (eds.), *Design of Intelligent Systems Based on Fuzzy Logic, Neural Networks and Nature-Inspired Optimization*,
Studies in Computational Intelligence 601, DOI 10.1007/978-3-319-17747-2_4

and IT2PCM [19] algorithms, that are capable to create high order granules, due to that algorithms found upper and lower bounds of the interval to prototypes and partition matrices into a given data set.

In [1, 4, 15] a proposal of how to make Granular Clustering using the FCM algorithm is presented, and a way to create granular prototypes and granular partition matrices. Based on this work we present a new proposal of a granular fuzzy C-means algorithm.

This work is organized as follows. In Sect. 2 we show a brief overview of the Fuzzy C-Means algorithm, Sect. 3 described a new proposal of a Granular Fuzzy C-Means Algorithm presented in this paper, Sect. 4 shows the plots of the granular prototype and results with benchmark datasets, Sect. 5 contains the conclusions obtained during the elaboration of this work.

2 Fuzzy C-Means Algorithm

The FCM algorithm is a clustering unsupervised method widely used in data clustering, image segmentation, pattern recognition, etc.; this algorithm creates soft partitions where a datum can belong to different clusters with a different membership degree to each cluster of the dataset. This clustering method is an iterative algorithm, which uses the necessary condition to achieve the minimization of the objective function J_m represented by the following equation [2, 12]:

$$J_m(U,V,X) = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_i(x_k)^m \cdot d_{ik}^2 \tag{1}$$

where the variables in (1) represent the following:

n	the total number of patterns in a given data set
c	is the number of clusters, which can be found from 2 to n − 1
X	are data characteristics, where $X = \{x_1, x_2, \ldots, x_n\} \subset R^s$
V	are the centers of the clusters, where $V = \{v_1, v_2, \ldots, v_n\} \subset R^s$
$U = \mu_{ij}$	is a fuzzy partition matrix, which contains the membership degree of each dataset x_j to each cluster v_i
d_{ik}^2	is the Euclidean distance between each data x_k of the dataset and the centers v_i of clusters
m	is the weighting exponent

The corresponding centers of the clusters and membership degrees for each respective data to solve the optimization problem with the constraints in (1) are given by Eqs. (2) and (3), which provide an iterative procedure. The aim is to improve a sequence of fuzzy clusters until no further improvement in (1) can be performed [2, 12].

$$v_i = \frac{\sum_{k=1}^{n} \mu_i(x_k)^m \cdot x_k}{\sum_{k=1}^{n} \mu_i(x_k)^m} \tag{2}$$

$$\mu_{\mathrm{i}}(x_k) = \left(\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}} \right)^{-1} \tag{3}$$

Equations (2) and (3) are an iterative optimization procedure. The aim is to improve a sequence of fuzzy clusters until no further improvement in $J_{m,\eta}(U, V, X)$ can be made. The Fuzzy C-Means algorithm consists of the following steps [2, 5, 12]:

1. Given a pre-selected number of clusters c and a chosen value for m, initialize the fuzzy partition matrix μ_{ij} of x_j belonging to cluster i such that:

$$\sum_{i=1}^{c} \mu_{ij} = 1 \tag{4}$$

2. Calculate the center of the fuzzy clusters, v_j for i = 1, 2,…, c using Eq. (2).
3. Use Eq. (3) to update the fuzzy membership μ_{ij}.
4. If the improvement in $J_m(U, V, X)$ is less than a certain threshold (ε), then stop, otherwise go to step 2.

The FCM clustering is completed through a sequence of iterations, where we can start from a certain randomly initiated centroids of clusters or a certain randomly partition matrix, and iterate over the formulas (2) and (3) given above.

3 Proposal of Granular Fuzzy C-Means Algorithms

From a general point of view, fuzzy clustering is about forming information granules and revealing the structure in data. Fuzzy clusters are information granules capturing the data. Observing fuzzy clustering from a different point of view this kind of algorithms are providers of a certain granulation-degranulation mechanism.

Considering that clustering has been realized previously, we can express it in terms of cluster leading to a granular description of x to any datum x into a given dataset, the granular description mentioned above typically in fuzzy clustering algorithms is done by computing the membership degrees of x to the cluster prototypes found by the fuzzy clustering algorithm, and this process is considering a granulation phase. The degranulation is about data reconstruction on basis of the granular or internal representation (Fig. 1).

In the FCM algorithm the granulation-degranulation mechanisms can be described in the two phases mentioned below:

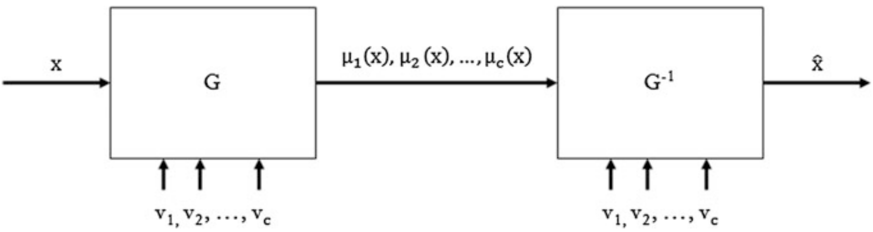


Fig. 1 Representation of the granulation-degranulation mechanisms

1. Granulation of the data x is made in terms of membership grades of the constructed information granules, the membership grades are computed by Eq. (3).
2. Degranulation provides a reconstruction of data in terms of the prototypes and membership grades computed by (2) and (3) respectively. Formally, the reconstruction of data is determined by solving the following optimization problem, with the minimization of the reconstruction error

$$\sum_{i=1}^{c} \mu_i(x_j) \cdot (v_i - \hat{x}_j)^2 \tag{5}$$

As a result, the equation of minimization error shown below is obtained:

$$\hat{x}_j = \frac{\sum_{i=1}^{c} \mu_i(x_j)^m \cdot v_i}{\sum_{i=1}^{c} \mu_i(x_j)^m} \tag{6}$$

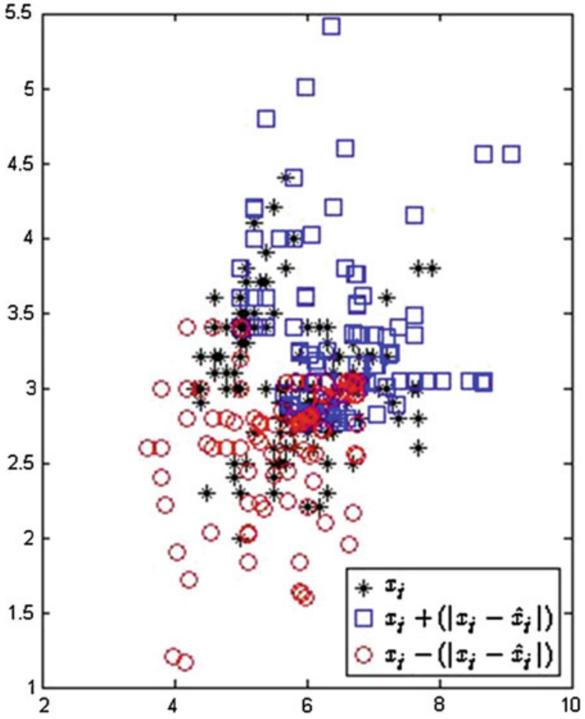
The reconstruction error becomes a function that takes into account the prototypes of the clusters and the matrices of belonging; Eq. (6) is very similar to Eq. (2) for prototype computing, observing this we can say that the reconstruction error is by finding x in Eq. (2) to compute the reconstruction error.

Based in the granulation-degranulation mechanism of the FCM algorithm, we propose a new Granular Fuzzy C-Means algorithm, which is different to that mentioned in [1, 4, 15]. This new proposal, like all algorithms that use granular computing is performed in two phases. The first phase is the granulation is a regular process of the FCM algorithm to obtain the membership degrees and prototypes of data, which are numerical values, to make this to granular membership degrees and granular prototypes. Now we proceed with the second phase, where the reconstruction error is computed with Eq. (6) and used to create two new data sets using the following equations:

$$x_j^+ = x_j + \left(\left| x_j - \hat{x}_j \right| \right) \tag{7}$$

$$x_j^- = x_j - \left(\left| x_j - \hat{x}_j \right| \right) \tag{8}$$

Fig. 2 Data distribution of the Iris flower data set and data distribution of the data set created by Eqs. (7) and (8)



Equations (7) and (8) basically compute the upper and lower bound of the data in agreement with the reconstruction error, with the new data set created using addition and subtraction to original data, the difference between original data and reconstructed data. In Fig. 2 we can observe the distribution of the data set created with Eqs. (7) and (8).

We can obtain a granular prototype and granular membership degrees by making the grouping of the data set created with Eqs. (7) and (8) using the FCM algorithm, in Fig. 3 we can observe the process of the granular fuzzy C-means algorithm proposed and Fig. 4 show the block diagram of the FCM algorithm.

As we may observe in Fig. 3 the process of creating granular prototypes and granular membership degrees is performed by the execution of FCM algorithms over the data set created by Eqs. (7) and (8), this is due to the fact that the data created are affected by reconstruction error and make a upper and lower bound of data. In the next section we show the granular prototype found by the proposed Granular Fuzzy C-Means algorithm over some benchmark data sets.

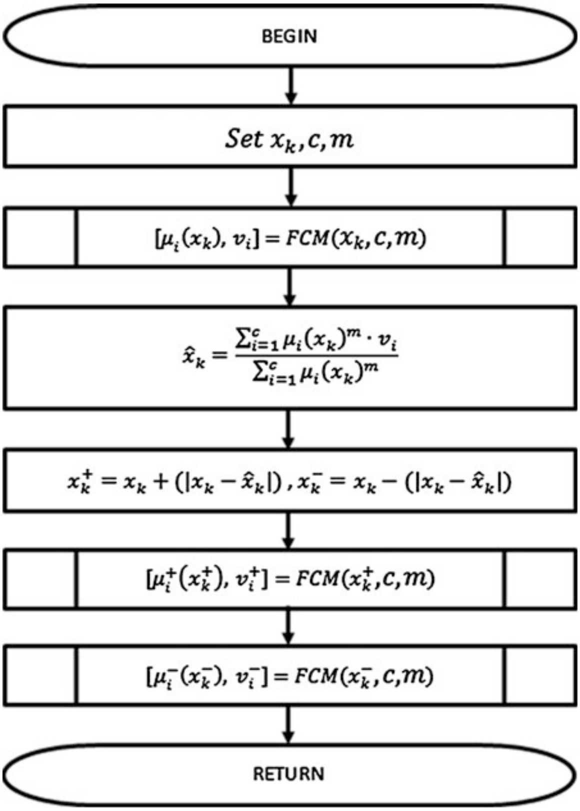

BEGIN
$Set\ x_k, c, m$
$[\mu_i(x_k), v_i] = FCM(x_k, c, m)$
$\hat{x}_k = \frac{\sum_{i=1}^{c} \mu_i(x_k)^m \cdot v_i}{\sum_{i=1}^{c} \mu_i(x_k)^m}$
$x_k^+ = x_k + (|x_k - \hat{x}_k|), x_k^- = x_k - (|x_k - \hat{x}_k|)$
$[\mu_i^+(x_k^+), v_i^+] = FCM(x_k^+, c, m)$
$[\mu_i^-(x_k^-), v_i^-] = FCM(x_k^-, c, m)$
RETURN

Fig. 3 Blocks diagram of the proposed granular fuzzy c-means algorithm

4 Simulation

In order to observe, if the proposal is capable to create a granular prototype, we realized the clustering of benchmark datasets, and the benchmark dataset used to perform these tests are the following:

- Wine
- WDBC
- Iris Flower

Wine data set is composed by 3 classes, with 178 instances in total and 13 attributes the number of instances per class show below:

- Class 1: 59 instances
- Class 2: 71 instances
- Class 3: 48 instances

WDBC data set is composed by 2 classes, with 569 instances in total and 32 attributes, the number of instances per class show below:

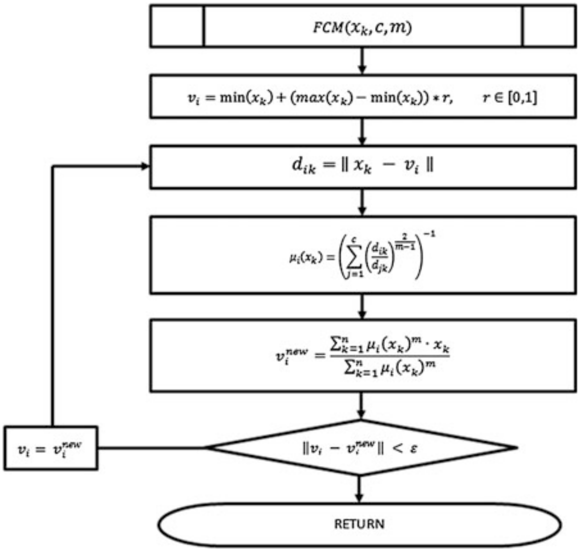

$FCM(x_k, c, m)$
$v_i = \min(x_k) + (max(x_k) - \min(x_k)) * r, \quad r \in [0,1]$
$d_{ik} = \| x_k - v_i \|$
$\mu_i(x_k) = \left(\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}\right)^{-1}$
$v_i^{new} = \frac{\sum_{k=1}^{n}\mu_i(x_k)^m \cdot x_k}{\sum_{k=1}^{n}\mu_i(x_k)^m}$
$v_i = v_i^{new}$
$\|v_i - v_i^{new}\| < \varepsilon$
RETURN

Fig. 4 Blocks diagram of the original fuzzy c-means algorithm

- Class 1 (benign): 357 instances
- Class 2 (malignant): 212 instances

Iris Flower data set is composed by 3 classes, with 150 instances in total and 4 attributes the number of instances per class show below:

- Class 1 (Setosa): 50 instances
- Class 2 (Versicolour): 50 instances
- Class 3 (Virginica): 50 instances

Figures 5, 7 and 9 show the prototypes found by the FCM algorithm over the Wine, WDBC, and Iris Flower data sets respectively, and these prototypes are

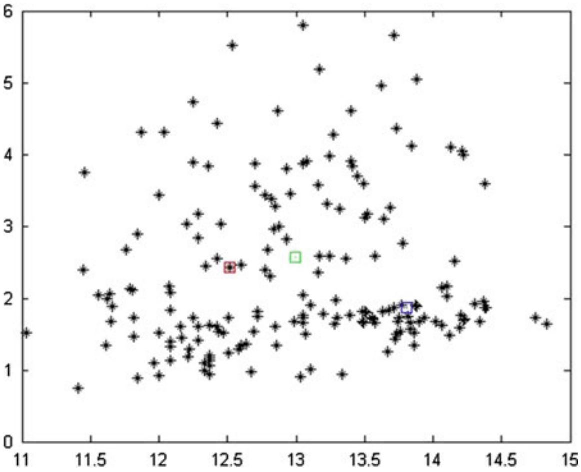


Fig. 5 Prototype found by the FCM algorithm in the Wine dataset

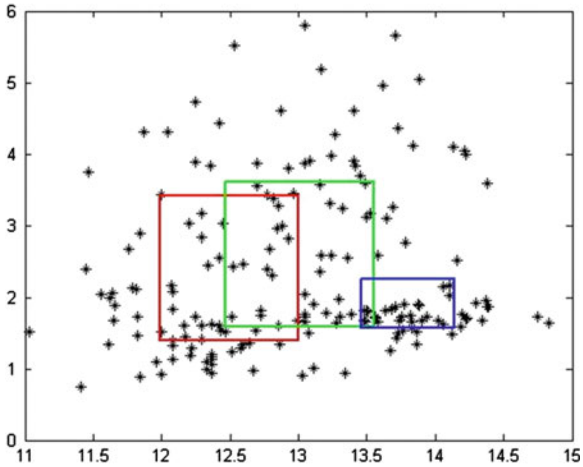


Fig. 6 Granular prototype found by the GFCM algorithm in the Wine data set

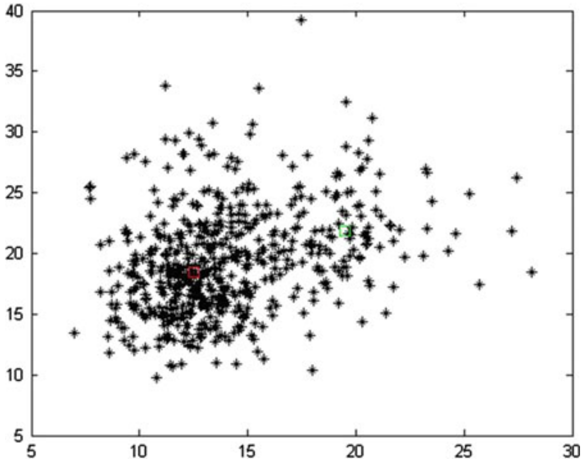


Fig. 7 Prototype found by the FCM algorithm in the WDBC dataset

represented by a numerical value and as can we see that prototype found represent the data structure but this representation is not reflective. Figures 6, 8 and 10 show the granular prototypes found by a GFCM (Granular Fuzzy C-Means) algorithm proposed over Wine, WDBC, and Iris Flower data sets respectively, in these figures we can observe that the representation of the data structure is more reflective.

5 Conclusions

Granular theory is very general and there is no a specific way of implementing this theory with clustering algorithms.

In this work a method is presented to apply the granular theory to the Fuzzy c-Means algorithm, making a granular fuzzy c-means algorithm capable of create granular prototypes and granular matrices of belonging instead of numerical prototypes and numerical matrices of belonging making more reflective the data structure.

Figures 6, 8 and 10 show the granular prototypes found by the proposed GFCM (Granular Fuzzy C-Means) algorithm applied to the Wine, WDBC, and Iris Flower data sets respectively that represent the data structure of the each dataset mentioned

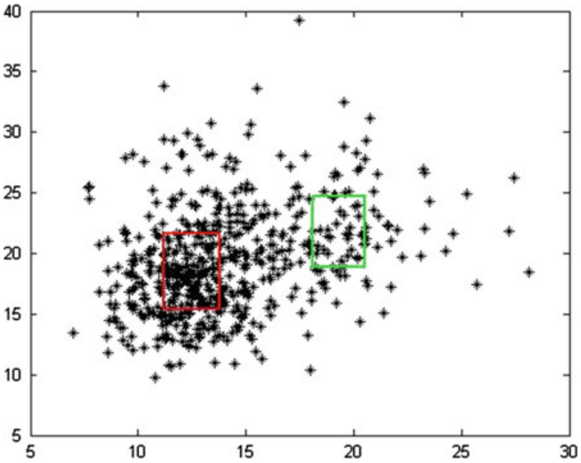


Fig. 8 Granular prototype found by the GFCM algorithm in the WDBC data set

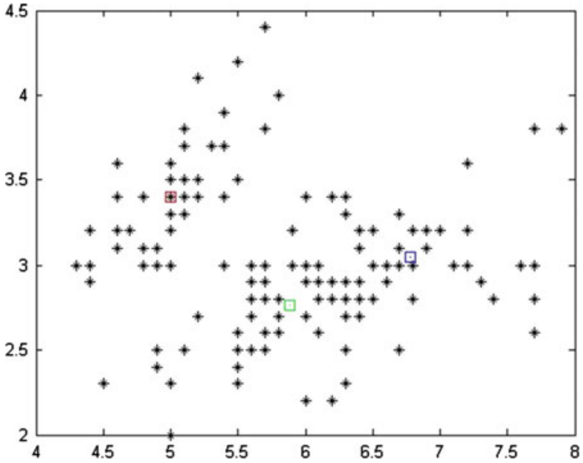


Fig. 9 Prototype found by the FCM algorithm in the Iris flower dataset

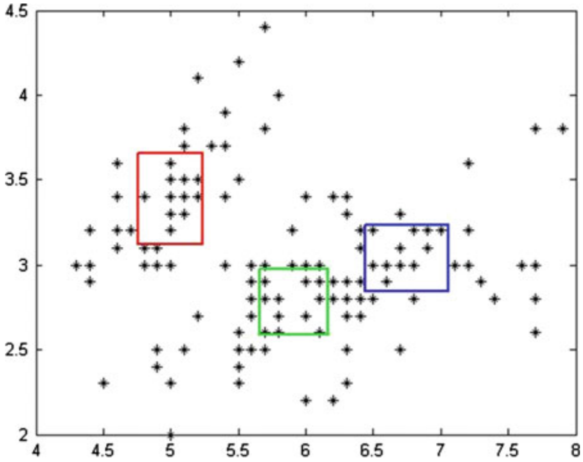


Fig. 10 Granular prototype found by the GFCM algorithm in the Iris flower data set

above. In these figures we can observe that the representation of the data structure is more reflective in comparison with the data structure found by the FCM algorithm for the same datasets.

References

1. Bargiela, A., Pedrycz, W., Hirota, K.: Granular prototyping in fuzzy clustering. IEEE Trans. Fuzzy Syst. **12**(5), 697–709 (2004)
2. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, Berlin (1981)
3. Chang, X., Li, W., Farrell, J: A C-means clustering based fuzzy modeling method. In: The Ninth IEEE International Conference on Fuzzy Systems, 2000. FUZZ IEEE 2000, vol. 2, pp. 937–940 (2000)
4. Gacek, A.: From clustering to granular clustering: a granular representation of data in pattern recognition and system modeling. In: IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint, pp. 502–506, 24–28 June 2013
5. Gustafson, D.E., Kessel, W.C.: Fuzzy clustering with a fuzzy covariance matrix. In: Proceedings of IEEE Conference on Decision and Control, San Diego, CA, pp. 761–766 (1979)
6. Hirota, K., Pedrycz, W.: Fuzzy computing for data mining. Proc. IEEE **87**(9), 1575–1600 (1999)
7. Hwang, C., Rhee, F.: Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means. IEEE Trans. Fuzzy Syst. **15**(1), 107–120 (2007)
8. Iyer, N.S., Kendel, A., Schneider, M.: Feature-based fuzzy classification for interpretation of mamograms. Fuzzy Sets Syst. **114**, 271–280 (2000)
9. Karnik, N., Mendel, J.: Operations on type-2 set. Fuzzy Set Syst. **122**, 327–348 (2001)
10. Krishnapuram, R., Keller, J.: A possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. **1** (2), 98–110 (1993)
11. Krishnapuram, R., Keller, J.: The possibilistic c-Means algorithm: Insights and recommendations. IEEE Trans. Fuzzy Sys. **4**(3), 385–393 (1996)

12. Kruse, R., Döring, C., Lesot, M.J.: Fundamentals of fuzzy clustering. In: Advances in Fuzzy Clustering and its Applications. Wiley, Chichester, pp. 3–30 (2007)
13. Mendel, J.: Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, pp. 213–231. Prentice-Hall, Englewood Cliffs (2001)
14. Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.: A possibilistic fuzzy c-means clustering algorithm. IEEE Trans. Fuzzy Syst. **13**(4), 517–530 (2005)
15. Pedrycz, W., Bargiela, A.: An Optimization of allocation of information granularity in the interpretation of data structures: toward granular fuzzy clustering. IEEE Trans. Syst. Man Cybern. B Cybern **42**(3), 582–590 (2012)
16. Philips, W.E., Velthuinzen, R.P., Phuphanich, S., Hall, L.O., Clark, L.P., Sibiger, M.L.: Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of hemorrhagic glioblastoma multiforme. Magn. Reson. Imaging **13**(2), 277–290 (1995)
17. Rubio, E., Castillo, O.: Interval type-2 fuzzy clustering for membership function generation. In: IEEE Workshop on Hybrid Intelligent Models and Applications (HIMA), pp. 13–18, 16–19 Apr 2013
18. Yang, M.-S., Hu, Y.-J., Lin, K.C.-R., Lin, C.C.-L.: Segmentation techniques for tissue differentiation in MRI of Ophthalmology using fuzzy clustering algorithms. Magn. Reson. Imaging **20**, 173–179 (2002)
19. Zarandi, M.H.F., Zarinbal, M., Türksen, I.B.: Type-II fuzzy possibilistic C-mean clustering. In: IFSA/EUSFLAT Conference, pp. 30–35 (2009)