2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

A Learning Scheme to Fuzzy C-Means based on a Compromise in Updating Membership Degrees

Shang-Lin Wu[1], *Student, Member*, Yang-Yin Lin[2], *IEEE, Member*, Yu-Ting Liu[1], *Student Member*
[1]Institute of Electrical Control Engineering, [2]Brain Research Center, National Chiao Tung University, Hsinchu 300, Taiwan

Chih-Yu Chen[3] and Chin-Teng Lin[12*], *Fellow, IEEE*
[3]Institute of Biomedical Engineering, National Chiao Tung University, Hsinchu 300, Taiwan
*Email: ctlin@mail.nctu.edu.tw

***Abstract*—Fuzzy C-Means (FCM) clustering is the most well-known clustering method according to fuzzy partition for pattern classification. However, there are some disadvantages of using that clustering method, such as computational complexity and execution time. Therefore, to solve these drawbacks of FCM, the two-phase FCM procedure has been proposed in this study. Compared with the conventional FCM, the usage of a compromised learning scheme makes more adaptive and effective. By performing the proposed approach, the unknown data could be rapidly clustered according to the previous information. A synthetic data set with two dimensional variables is generated to estimate the performance of the proposed method, and to further demonstrate that our method not only reduces computational complexity but economizes execution time compared with the conventional FCM in each example.**

Keywords—Fuzzy C-Means (FCM); Clustering; Data classification; High computational complexity; Long execution time.

I. Introduction

Clustering [1] plays an important role in the given data by showing the fundamental structure and providing some significant information for users [2]. There are several clustering methods (e.g., expectation-maximization (EM) algorithm [3, 4], balanced iterative reducing and clustering using hierarchies (BIRCH) [4, 5] and k-means [6, 7]) applied to a variety of data sets. However, the cluster analysis is limited to the techniques of a statistical classification whether the data is falling into own representative cluster by making quantitative evaluation Cluster method also divides input data into groups or clusters on the basis of some similarity criterion, such that similar data objects belongs to the same group.

Some examples of measures that can be used in clustering (e.g., distance [8], connectivity, pattern recognition [9, 10] and image segmentation [11, 12]). The resulting partition improves the understanding of human beings and helps to be more informed by decision making. There are two kinds of methods in deriving clusters, involving crisp form [13] and fuzzy clustering [14, 15]. In the crisp clustering, data would be divided into a number of distinct clusters, where each data element belongs to exactly one cluster. On the contrary, FCM clustering [16, 17] is based on finding partition function , and then using derived results to assign data elements to one or more clusters or classes in the data sets. The advantage of fuzzy systems is easy to understand and implement the system. FCM is a useful clustering method but it still has some drawbacks. The given solution would not be an optimal solution and the performance of FCM would be lack of efficiency. Therefore, when the data set is corrupted by the noise, the performance might be inadequate.

In this study, we would like to propose the compromised clustering scheme based Fuzzy C-Means. In contrast to the conventional FCM, the proposed approach enables to significantly reduce learning time while the data set is larger. Besides, because of the simple design and less computational time, the proposed method can be implemented simply.

Thus study is organized as follows. In Section II, we both describe the FCM and the proposed method. Simulation results of the Compromised FCM and conventional FCM are shown in Section III. Finally, there are some conclusions are given in Section IV.

II. Materials and Methods

A considerable amount of research for Clustering contains machine learning, data mining and pattern classification. Especially for fuzzy clustering method, it considers each cluster as a fuzzy set and calculates the degree belonging of each cluster with membership function measurement. According to the degree of representative, each data would be assigned to several clusters. Two clustering method of fuzzy clustering presented in this study are FCM and compromised FCM as following:

A. Fuzzy C-Means

In the conventional FCM [18], all the membership values are based on the distance between data and each cluster center. The formulation of FCM and its updating cluster center can be defined as the following equation:

$$\mathrm{J} = \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m dist_{ij}^2, \sum_{i=1}^{c} u_{ij} = 1,$$
$$\forall j \in 1,2,\ldots,n \text{ and } \forall i \in 1,2,\ldots,c \tag{1}$$

$$u_{ij} = \frac{1}{\left(\frac{\sum_{i=1}^{c}\|x_j - c_i\|}{\sum_{k=1}^{c}\|x_j - c_k\|}\right)^{\frac{2}{m-1}}} \tag{2}$$

$$c_i = \frac{\sum_{j=1}^{n}(u_{ij})^m x_j}{\sum_{j=1}^{n}(u_{ij})^m} \tag{3}$$

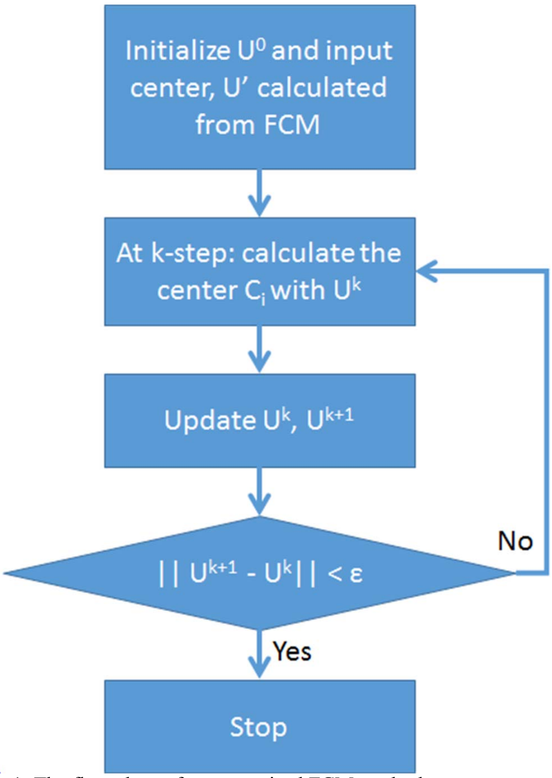

Fig. 1. The flow chart of compromised FCM method.

where u_{ij} is the degree of representative of the j_{th} data to the i_{th} cluster, $dist_{ij}$ is the distance between the j_{th} data and the i_{th} cluster center, m is the degree of fuzziness, c is the number of clusters, and N is the number of the data samples.

When one determines a new location for a cluster center, the current data would be determined according to the memberships in adjusting the location of the cluster centers. As a result, once the location of a cluster center is updated, it may be uncertain whether each sample properly contributes in updating the location of the cluster center. Thus, FCM is modified to compromised clustering based on fuzzy c-means by assigning the membership function and cluster center to the new updating equation.

B. Compromised Clustering Based on Fuzzy C-Means

In order to integrate the proposed method and the FCM algorithm, we introduce the role of the membership values in updating the cluster centers. We give the FCM membership values to compromised FCM by assigning a membership function with β value as shown in Eq. (4). The compromised FCM membership values are obtained by using the previous calculated membership values into the compromised FCM. Fig. 1 shows that membership values and center of the compromised FCM can be computed by the following Eqs (4) and (5). Similar to FCM protocol, the procedure of the compromised FCM utilizes an iterative algorithm in order to minimize the objective function, and it can be denoted as follows:

$$\min \ J = \sum_{i=1}^{c}\sum_{j=1}^{n}(\beta u_{ij} + (1-\beta)u'_{ij})^m dist_{ij}^2,$$
$$\sum_{i=1}^{c} u_{ij} = 1,$$
$$\forall j \in 1,2,\dots,n \text{ and } \forall i \in 1,2,\dots,c \qquad (4)$$

where u_{ij} is the degree of belonging of the j_{th} data to the i_{th} cluster, $dist_{ij}$ is the distance between the j_{th} data and the i_{th} cluster center, m is the degree of fuzziness, c is the number of clusters, and N is the number of the data samples. To minimize the objective function J and substitute for the memberships in the center update equation of the compromised FCM method gives the result as following equation:

$$u_{ij} = \frac{1}{\left(\frac{\sum_{k=1}^{c}\|x_j - c_i\|}{\sum_{k=1}^{c}\|x_j - c_k\|}\right)^{\frac{2}{m-1}}} - \frac{1}{\beta}(1-\beta)u'_{ij} \qquad (5)$$

$$c_i = \frac{\sum_{j=1}^{n}(\beta u_{ij} + (1-\beta)u'_{ij})^m x_j}{\sum_{s=1}^{n}(\beta u_{is} + (1-\beta)u'_{is})^m}$$
$$= \sum_{j=1}^{n}\frac{(\beta u_{ij} + (1-\beta)u'_{ij})^m}{\sum_{s=1}^{n}(\beta u_{is} + (1-\beta)u'_{is})^m} x_j \qquad (6)$$

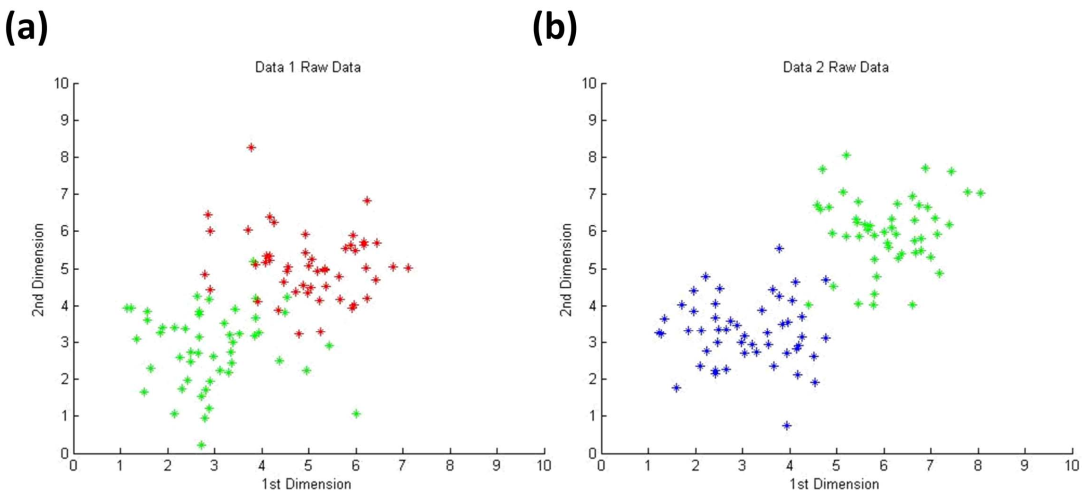

Fig. 2. These two two-dimensional data sets with Gaussian distribution with (a) 3 and 5 mean value and standard deviation with 1and (b) 3 and 6 mean value and standard deviation with 1. There are 100 data points are contained in each data set.

In spite of referring information from FCM, the procedure of compromised FCM is similar to the conventional FCM. After the final locations of the cluster centers are determined, the iteration would be finished.

III. Results and Discussions

The simulation results of the proposed clustering method would be explained through two classes with two-dimensional data sets, as shown in Fig. 2. Each data set is comprised of 100 data points. These data sets are generated by using a Gaussian distribution with exploiting 3, 5 and 6 for mean value and standard deviation that is set to be one. The performance of the compromised FCM is expressed through the values of previous calculated partition matrix and then uses them to cluster the new data sets.

A. Performance of this study

In comparison with FCM and compromised FCM, simulation results are completely displayed in Fig. 3. Left column in Fig. 3 shows that the data set 1 clustering result after FCM in the number of clusters to be equal to 2 and has 12, 12 and 11 times iterations. The center and partition matrix are all recorded and would be used in the compromised FCM. As can be seen in Fig. 3, the simulation result of compromised FCM is shown in the middle column. As mentioned in the previous section, we have less computational time with 3 times iterations of proposed clustering scheme compared to that of FCM. The right column in Fig. 3 is just showing the result of FCM with 11, 9 and 11 iterations from up to down. Running the compromising clustering for unknown data, we have noticed the proposed strategy could achieve the fewer numbers of iterations and enhance its effectiveness.

IV. Conclusions

The purpose of this study for clustering strategy is to present a compromised FCM, which is able to effectively economize computational load compared to the conventional Fuzzy C-Mean (FCM) method. The results of the proposed method are compared with FCM using two-dimensional synthetic data sets as input data. The results show that proposed method using a compromise concept can provide less computational time because of obtaining the previous information, and further to calculate the new partition matrix. Therefore, execution time of the proposed clustering method is faster than that of conventional FCMs. In addition, the proposed clustering method can handle uncertainties data sets well, which would prevent the CFCM clustering from being re-computed the newly incoming data set.

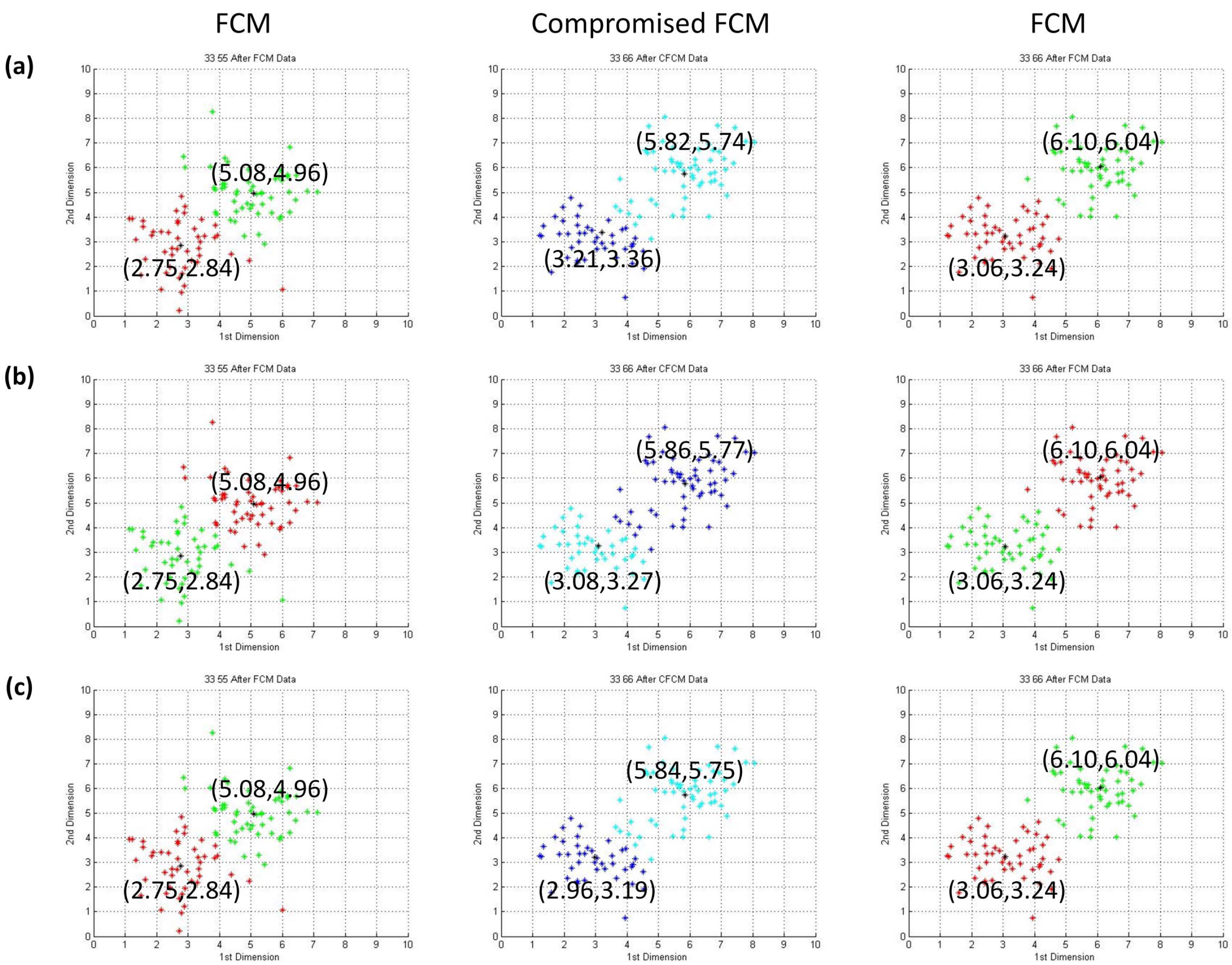

Fig. 3. It shows that the experimental results compared with FCM and compromise FCM in β = (a) 0.2 (b) 0.5 (c) 0.8, respectively.

ACKNOWLEDGMENT

This work was supported in part by the UST-UCSD International Center of Excellence in Advanced Bio-engineering sponsored by the Taiwan National Science Council I-RiCE Program under Grant Number: NSC 102-2911-I-009-101, in part by NSC-NSC102-2220-E-009-041, in part by the Aiming for the Top University Plan, under Contract 103W963.

REFERENCES

[1] R. Xu and D. C. Wunsch, 2nd, "Clustering algorithms in biomedical research: a review," *IEEE Reviews in Biomedical Engineering,* vol. 3, pp. 120-54, 2010.

[2] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*: Kluwer Academic Publishers, 1981.

[3] R. D. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Transactions on Signal Processing,* vol. 51, pp. 2245-2253, 2003.

[4] P. R. Kersten, L. Jong-Sen, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture Radar images using fuzzy clustering and EM clustering," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 43, pp. 519-527, 2005.

[5] L. Kovacs and L. Bednarik, "Parameter optimization for BIRCH pre-clustering algorithm," in *2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2011, pp. 475-480.

[6] Y. H. Liu, C. T. Wu, Y. H. Kao, and Y. T. Chen, "Single-trial EEG-based emotion recognition using kernel Eigen-emotion pattern and adaptive support vector machine," *Conference of the IEEE Engineering in Medicine and Biology Society,* vol. 2013, pp. 4306-9, 2013.

[7] M. M. Hossain and M. N. Norazmi, "Pattern Recognition Receptors and Cytokines in Infection-The Double-Edged Sword?," *BioMed Research International,* vol. 2013, p. 179174, 2013.

[8] M. J. O'Malley, M. F. Abel, D. L. Damiano, and C. L. Vaughan, "Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters," *IEEE Trans Rehabil Eng,* vol. 5, pp. 300-9, Dec 1997.

[9] H. Zhang, Y. Zhao, F. Yao, L. Xu, P. Shang, and G. Li, "An adaptation strategy of using LDA classifier for EMG pattern recognition," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2013, pp. 4267-70, 2013.

[10] A. Fernandez-Caballero, M. V. Sokolova, and J. Serrano-Cuerda, "Lateral inhibition in accumulative computation and fuzzy sets for human fall pattern recognition in colour and infrared imagery," *ScientificWorldJournal,* vol. 2013, p. 935026, 2013.

[11] F. Salzenstein and C. Collet, "Fuzzy Markov random fields versus chains for multispectral image segmentation," *IEEE Trans Pattern Anal Mach Intell,* vol. 28, pp. 1753-67, Nov 2006.

[12] H. Zhou, G. Schaefer, and C. Shi, "A mean shift based fuzzy c-means algorithm for image segmentation," *Conference of the IEEE Engineering in Medicine and Biology Society* vol. 2008, pp. 3091-4, 2008.

[13] S. Karamchandani, S. N. Merchant, U. B. Desai, and G. D. Jindal, "Application of crisp and fuzzy clustering algorithms for identification of hidden patterns from plethysmographic observations on the radial pulse," *Conference of the IEEE Engineering in Medicine and Biology Society,* vol. 2010, pp. 3978-81, 2010.

[14] W. Wieclawek and E. Pietka, "Fuzzy clustering in Intelligent Scissors," *Computerized Medical Imaging and Graphics,* vol. 36, pp. 396-409, Jul 2012.

[15] L. Han, X. Zeng, and H. Yan, "Fuzzy clustering analysis of microarray data," *Proc Inst Mech Eng H,* vol. 222, pp. 1143-8, Oct 2008.

[16] M. P. Windham, "Cluster Validity for the Fuzzy c-Means Clustering Algorithrm," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 4, pp. 357-63, Apr 1982.

[17] X. Y. Wang, X. J. Zhang, H. Y. Yang, and J. Bu, "A pixel-based color image segmentation using support vector machine and fuzzy C-means," *Neural Networks,* vol. 33, pp. 148-59, Sep 2012.

[18] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences,* vol. 10, pp. 191-203, 1984.