

Fuzzy C-means based on Automated Variable Feature Weighting

Mousa Nazari, Jamshid Shanbehzadeh, and Abdolhossein Sarrafzadeh

Abstract—Fuzzy C-means (FCM) is a powerful clustering algorithm and has been introduced to overcome the crisp definition of similarity and clusters. FCM ignores the importance of features in the clustering process. This affects its authenticity and accuracy. We can overcome this problem by appropriately assigning weights to features according to their clustering importance. This paper, proposes an improved FCM algorithm based on the method proposed by Huang by automated feature weighting. The simulation results on several UCI databases show that the proposed algorithm exhibits better performance than FCM.

Index Terms—Fuzzy Clustering, Fuzzy C-means, Feature Weighting, Weighted Fuzzy C-means

I. INTRODUCTION

Cluster analysis groups data according to their similarities. Normally, data is described by a vector where each of its components is one of its features. The similarity of two vectors is based on the cumulative sum of the distance of feature vector components. The distance measurement can be based on Euclidian, Chess Board, Mahalanobis and etc. There are two problems associated with these approaches. The first belongs to the definition of similarity; the second belongs to the importance of vector components in distance measurement.

Traditionally, similarity is a crisp concept and, two vectors can be similar or dissimilar. This approach is unable to cover the similarity concept because; there is some degree of similarity rather than a simple yes or no answer. Zadeh [1] first articulated fuzzy set theory which gave rise to the concept of partial membership based on membership function. This approach opens the way to include the partial similarity. Fuzzy clustering is an output of this approach.

This concept produces overlapped rather than fixed cluster partitions based on the fuzzy similarity concept. FCM proposed by Dunn [2] and extended by Bezdek [3] is one of the most well-known algorithms in clustering analysis.

Manuscript received December 20, 2012; accepted Jan 3, 2013.

M. Nazari is a student with the Department of Computer Engineering, Kharazmi University, Tehran, I.R. Iran (e-mail: nazari.mousa@gmail.com).

J. Shanbehzadeh is an Associate Professor with the Department of Computer Engineering, Faculty of Engineering, Kharazmi University (Tarbiat Moallem University of Tehran), Tehran, I.R. Iran (phone: +98 26 34550002; fax: +98 26 34569555; (e-mail: jamshid@tmu.ac.ir).

A. Sarrafzadeh is an Associate Professor and Head of Department of Computing, Unitec Institute of Technology, New Zealand (email: hsarrafzadeh@unitec.ac.nz).

Basically, FCM clustering depends on the measure of distance between samples. In most situations, FCM employs the traditional distance measurement that supposes similar weight to each feature. This assumption seriously affects the performance of FCM, since, in most of the real-world problems, data features have different effects in the overall clustering performance. Wang et al [4] employed Iris data (Fisher, 1936 [5]) to show that the different feature weight affects the performance of FCM. However, improperly choosing the feature weights degrades the FCM performance. So, it is important to select suitable feature weights to guarantee the performance. Wang et al [5] proposed a feature-weight learning approach based on a defined similarity measure and an evaluation function to improve the FCM performance. However, the defined similarity measure and the evaluation function are complicated and difficult to interpret.

This paper, proposes a novel Weighted Fuzzy FCM algorithm (W-FCM) that automatically assigns weights to variables based on their clustering importance. This approach is an extension of WK-means clustering proposed by Huang[6]. Similar to WK-means, W-FCM adds a new step to the basic FCM algorithm to update the weight of features based on their current data partition. The rest of this paper is organized as follows. Section II, reviews FCM and K-means clustering algorithms and lists some cluster validity functions to measure the performance. Section III demonstrates the WK-means algorithm and introduces the novel W-FCM algorithm based on it. Section IV presents the experimental results based on some UCI databases. The final section is our conclusion.

II. BACKGROUND

A. K-Means

K-means algorithm is classified as a partitional or nonhierarchical clustering algorithm [7], where it assumes a fixed number of clusters, and generates the clusters by the use of an error function. This algorithm proceeds, for a given initial k clusters, by assigning data to the nearest clusters and then repeatedly changing the membership of the clusters according to the error function, until the function does not change more than a specified threshold, or the membership of the clusters no longer changes. The conventional k-means algorithm is briefly described as below:

Let X be the dataset of N samples with D dimensions,

$X = (x_1, x_2, \dots, x_n), x_i \in R^d$. Our goal is to assign data points into K partitions. Assume that the K centers are v_1, v_2, \dots, v_k and in cluster i there exists N_i instances. So we can calculate its center by averaging its members:

$$v_i = \frac{1}{N} \sum_{i=1}^{N_i} x_i, \text{ for } i = 1..K \quad (1)$$

Based on Euclidean distance and the criteria of the within-groups sum of squared error, (2) shows the objective function of k-means clustering.

$$e^2(U, V) = \sum_{k=1}^K \sum_{i=1}^n \sum_{j=1}^d u_{k,i}^2 (x_{i,k} - v_{k,j})^2 \quad (2)$$

Where U is a $K \times n$ partition matrix, $u_{k,i}$ is a binary variable, and $u_{k,i} = 1$ indicates that object i belongs to cluster k . K-mean algorithm as follows:

-
- Randomly initialize the position of the c cluster centers.*
- i. Calculate the distance between all of the samples and each center.
 - ii. Each sample is assigned to a cluster based on the minimum distance.
 - iii. Recalculate the center position using (1)
 - iv. Recalculate the distance between each sample and each center.
 - v. Reassign each sample to a cluster.
 - vi. If no data was reassigned, then stop, otherwise repeat from step (iii).
-

B. Fuzzy C-Means

FCM clustering algorithm [3, 8] allows one piece of data to belong to more than one cluster according to a membership function. Let $X = (x_1, x_2, \dots, x_n)$ be a set of numerical data in R^d and c to be an integer between 1 and n . Given X , we say that c fuzzy subsets $\{u_k : X \rightarrow [0, 1]\}$ are a c -partition of X if the following conditions are satisfied:

$$0 \leq u_{k,j} \leq 1, \quad \forall k, j \quad (3)$$

$$\sum_{k=1}^c u_{k,j} = 1 \quad \forall j \quad (4)$$

$$0 < \sum_{i=1}^n u_{k,i} < n \quad \forall k \quad (5)$$

Where $u_{k,i} = u_k(x_i)$, $1 \leq k \leq c$ and $1 \leq i \leq n$. Let $u_{k,i}$ satisfy the above conditions represented by a $c \times n$ matrix $U = [u_{k,i}]$. FCM aims to determine cluster centers $v_k (k = 1, \dots, c)$ and the fuzzy partition matrix U by minimizing the objective function J defined as follows:

$$J(U, V, X) = \sum_{k=1}^c \sum_{i=1}^n u_{k,i}^m d_{k,i}^2 \quad (6)$$

Where $d_{k,i}$ is the Euclidean distance from sample x_j to the cluster center v_i defined as:

$$d_{k,i} = \sqrt{\sum_{j=1}^d (v_{k,j} - x_{i,j})^2} \quad (7)$$

The exponent m in (6) is the degree of fuzziness associated with the partition matrix ($m > 1$). If we consider m to be one, the soft clustering will be changed into hard one. Usually, we set m to 2. FCM algorithm as follows:

i. Choose an integer c and a threshold value. Let $m=2$. Fix and initialize the fuzzy partition matrix U with a random value such that it satisfies conditions (3), (4) and (5).

ii. Calculate the fuzzy centers v_k using

$$v_k = \frac{\sum_{i=1}^n (u_{k,i})^m x_i}{\sum_{i=1}^n (u_{k,i})^m}, \quad \forall k = 1, \dots, c \quad (8)$$

iii. Update the fuzzy partition matrix U with

$$u_{k,i} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{k,i}}{d_{k,\bar{k}}} \right)^{\frac{2}{m-1}}} \quad (9)$$

Where $d_{k,i}$ Compute according to (7)

Compute the objective function J by using (6). If it converges or the difference between two adjacent computed values of objective function J is less than the given threshold e then stops. Otherwise go to step (ii).

C. Cluster Validity Functions

Unlike hard clustering, its fuzzy version allows each data point to be in every cluster with a different degree of membership. Similar to hard clustering, we can define a validity index for the fuzzy case. The objective is to seek clustering schemes where most of the data points in the data set exhibit a high degree of membership in one cluster. Generally, there are two categories of fuzzy validity index. The first one considers the membership values and the second considers both the membership values and the data values mentioned in [9]. Various fuzzy validity indices are defined and discussed as follows.

The Partition Coefficient Index is the first validity index associated with FCM and it is [3], [10] defined by:

$$PC(c) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^2 \quad (10)$$

Where $\frac{1}{c} \leq PC(c) \leq 1$. In general, we find an optimal

cluster number c^* by solving $\max_{2 \leq c \leq n-1} PC(c)$ to produce the best clustering performance for the data set X .

The partition entropy (PE) index is another fuzzy validity index that involves only the membership values. It is defined as [4], [11]:

$$PE(c) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c (u_{ij} \log_2 u_{ij}) \quad (11)$$

Where $0 \leq PE(c) \leq \log_2 c$. In general, we find an optimal c^* by solving $\min_{2 \leq c \leq n-1} PE(c)$ to produce the best clustering performance for the data set X . Both PC and PE possess a monotonic evolution tendency in respect to c . Modification of the PC index [12] presented by (12) can reduce the monotonic tendency.

$$MPC(c) = 1 - \frac{c}{c-1} (1 - V_{PC}) \quad (12)$$

Where $0 \leq MPC(c) \leq 1$. Note that the MPC index is equivalent to the non-fuzziness index (NFI) [13]. In general, an optimal cluster number c^* is found by solving $\max_{2 \leq c \leq n-1} MPC(c)$ to produce the best clustering performance for the data set X .

The above indices only use fuzzy memberships and there is no connection to the geometrical structure of data. Fukuyama and Sugeno [14] proposed a fuzzy validity index based on both the membership values and the data values. Let $X = (x_1, x_2, \dots, x_n)$ be a data set, and $U = (u_{k,i})$ be the membership matrix of a fuzzy c -partition of X . Then the Fukuyama-Sugeno (FS) index is defined as:

$$FS(c) = \sum_{k=1}^c \sum_{i=1}^n u_{k,i}^m \|x_i - v_k\|^2 - \sum_{k=1}^c \sum_{i=1}^n u_{k,i}^m \|v_k - \bar{v}\|^2 \quad (13)$$

Where $\bar{v} = \frac{1}{c} \sum_{k=1}^c v_k$. In general, an optimal c^* is found

by solving $\max_{2 \leq c \leq n-1} FS(c)$ to produce the best clustering performance for the data set X .

III. EXTENSION

A. Weighted K-Means

Weighted K-Means algorithm (abbreviated as WK-Means) described by Chan, Huang and their collaborators [6], [15], [16] is a modification of K-Means (2) to consider weights to the features. Their approach relates feature weights to a set of patterns during the process of clustering, aligned to the wrapper approach idea for feature selection.

$$e^2(U, V, W) = \sum_{k=1}^K \sum_{i=1}^n \sum_{j=1}^d u_{k,i} w_j^s (x_{i,j} - v_{k,j})^2 \quad (14)$$

Subject to:

$$u_{k,i} \in [0,1], \quad \sum_{k=1}^K u_{k,i} = 1, \quad \sum_{j=1}^d w_j = 1$$

According to the criterion, each feature weight should be non-negative and their Addison should be equal to one. The criterion has a user-defined parameter s , which expresses the rate of impact weights on their contribution to the distance.

WK-means presents an extra step in relation to K-means, as it updates the feature weights by:

$$W_j = \begin{cases} 0 & \text{if } D_j = 0 \\ \frac{1}{\sum_{t=1}^h \left[\frac{D_j}{D_t} \right]^{\frac{1}{s-1}}} & \text{if } D_j \neq 0 \end{cases} \quad (15)$$

Where D_j is the sum of within-cluster variances of feature J weighted by clusters cardinalities:

$$D_j = \sum_{k=1}^K \sum_{i=1}^n u_{k,i} (x_{i,j} - v_{k,j})^2 \quad (16)$$

and h is the number of variables where $D_j \neq 0$. The WK-means can be presented in detail as follow:

i. Initial setting:

Similar to k -means, this method uses an external parameter to define the number K of clusters. WK-means then randomly initializes the centroids and the feature weights, ensuring that they add up to unity.

ii. Cluster update:

The cluster update assign each entity to their closest centroid, using the adjusted distance measure that takes into account the feature weights;

iii. Stop condition:

This is as per the K -means algorithm.

iv. Centroids update:

Update centroids to the center of gravity of its cluster. Again, as the method is based on Euclidean distance, this is achieved by moving the centroids to the mean of their clusters.

v. Weights update:

Update the weights according to (15), constrained by the condition that the sum of weights should be one.

B. Weighted-FCM

The presented method in this study (W-FCM) the FCM method is modified by considering feature importance in respect to (6) which can be expressed by (17).

$$J(W, U, V) = \sum_{k=1}^c \sum_{i=1}^n u_{k,i}^2 (d_{k,i}^{(w)})^2 \quad (17)$$

Where $d_{k,i}^{(w)}$ is computed by:

$$d_{k,i}^{(w)} = \sqrt{\sum_{j=1}^d w_j^s (x_{i,j} - v_{k,j})^2} \quad (18)$$

Minimizing (17) subject to (2), (3) and (4), then $u_{k,i}$ and v_k are calculated as:

$$v_k = \frac{\sum_{i=1}^n (u_{k,i})^m x_i}{\sum_{i=1}^n (u_{k,i})^m}, \quad \forall k = 1, \dots, c \quad (19)$$

$$u_{k,i} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{k,i}^{(w)}}{d_{k,k}^{(w)}}\right)^{\frac{2}{m-1}}} \quad (20)$$

The other parts of the algorithm are the same as FCM given in Section 2.

Also, similar to WK-means, W-FCM adds an extra step to FCM, as it updates the feature weights by (15) and (16). The main steps of W-FCM are shown as follows:

- i. Fix the number of clusters, c , where $2 \leq c \leq n$, and initialize the fuzzy partition matrix U with a random value such that it satisfies conditions (4) and (5), and initialize the weighting vector W with a random value such that it satisfies conditions (17) and (18)
- ii. Calculate the fuzzy centers v_j using (19)
- iii. update the fuzzy partition matrix U with (20)
- iv. Update the weighting vector W with (15) and (16)
Repeat step (ii) to (iv) until one of the termination criterion is satisfied.

The W-FCM algorithm follows an iterative optimization similar to FCM, and consequently it is affected by some of its strengths, such as its convergence in a finite number of iterations, and its weaknesses, such as the algorithm yet it initializes the centroids randomly, not guaranteeing an optimal solution. Also, the same applies to the weights, creating the possibility that these could be far from representing the relevancy of features.

The computational complexity of the algorithm is $O(tcnd)$, where t is the total number of iterations required for performing Steps ii, iii and iv, c is the number of clusters, d is the number of attributes, and n is the number of objects.

IV. SIMULATE RESULTS AND DISCUSSION

This section compares the performance of W-FCM against FCM on several UCI databases. The performance of clustering is measured by the validity functions which are descriptions in pervious section. In all the following experiments we consider $m=2$, $=2$ and threshold equal to 10^{-5} . We performed experiments on four UCI databases. Table I presents the databases and their attributes. Also, in the follow Table II presents the clustering results of FCM and WFCM algorithm based on the databases in Table I.

TABLE I.
THE CHARACTERS OF FOUR UCI DATABASES[17]

Database Name	Number of Samples	Number of Feature	Category of Feature	Number of Clusters
Iris	150	4	Numerical	3
Pima	768	8	Numerical	2
Statlog (Heart)	270	13	Categorical, Real	2
SPECT	267	22	Binary	2

Heart

TABLE II.
THE CLUSTERING RESULTS OF UCI DATABASES BY FCM AND W-FCM

		V_{PC}	V_{PE}	V_{MPC}	V_{FS}
Iris	FCM	0.783	0.571	0.674	-449.7148
				8	
	W-FCM	0.920	0.206	0.881	-518.1711
		8	6	2	
Pima	FCM	0.824	0.428	0.648	-
		2	2	5	2.7072e+6
	W-FCM	0.929	0.174	0.858	-
		3	2	4	5.1006e+6
Statlog (Heart)	FCM	0.712	0.649	0.425	1.2394e+5
		6	3	2	
	W-FCM	0.862	0.649	0.725	1.7200e+4
		8	3	7	
SPECT	FCM	0.585	0.868	0.170	4.9240e+5
		3	9	6	
	W-FCM	0.867	0.324	0.735	1.1689e+5
		9	9	8	

V. CONCLUSION

This paper investigated the effect of feature weight on FCM. W-FCM automatically calculated weights based on the current partition in an iterative FCM clustering process. The calculated weights of each iteration were based on the variation of the within cluster distances. It revealed that an appropriate assignment for feature weights could improve the performance of FCM clustering. Experiments on some UCI databases illustrated the outperformance of W-FCM over FCM.

REFERENCES

- [1] L. A. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338–353, 1965.
- [2] R. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics* 7, pp. 179–188, 1936.
- [3] J. C. Dunn, "Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems," *J. Cybernetics*, vol. 4, pp. 1–15., 1974.
- [4] X. Z. Wang, Y. D. Wang, and L. J. Wang, "Improving fuzzy c-means clustering based on feature-weight learning," *Pattern Recognition Lett.*, vol. 25, pp. 1123–1132., 2004.
- [5] Fisher, R.. "The use of multiple measurements in taxonomic problems". *Ann. Eugenics* 7, pp. 179–188, 1936.
- [6] J. Huang, M. K. Ng, H. Rong, and Z. Li, "Automated Variable Weighting in k-Means Type Clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 657-668, 2005.
- [7] A. Jain and R. Dubes, "Algorithms for Clustering Data," *Englewood Cliffs, NJ: Prentice-Hall*, 1988.
- [8] L. Bobrowski and J. Bezdek, "c-means clustering with the L_1 and L norms.," *IEEE Transactions on Systems, Man and Cybernetics*, pp. 545–554, 1991.

- [9] M. Halkidi and M. Vazirgiannis, "Clustering validity assessment: Finding the optimal partitioning of a data set," *Proceedings of the IEEE international conference on data mining*, pp. 187-194, 2001.
- [10] E. Trauwert, "On the meaning of Dunn's partition coefficient for fuzzy clusters," *Fuzzy Sets and Systems*, pp. 217-242., 1988.
- [11] J. Bezdek, "Cluster validity with fuzzy sets," *Journal of Cybernetics*, pp. 58-72., 1974.
- [12] R. N. Dave, "Validating fuzzy partition obtained through c-shells clustering.," *Pattern Recognition Lett.*, pp. 613-623, 1996.
- [13] M. Robubens, "Pattern classification problems and fuzzy sets.," *Fuzzy Sets Systems*, pp. 239-253, 1978.
- [14] [Y. Fukuyama and M. Sugeno, "A new method of choosing the number of clusters for the fuzzy c-means method.," *Proceeding of fifth Fuzzy Syst.*, pp. 247-250, 1989.
- [15] Y. Chan, W. K. Ching, N. M. K., and H. J. Z, "An optimization algorithm for clustering using weighted dissimilarity measures," *Pattern Recognition.*, pp. 943-952, 2004.
- [16] J. Z. Huang, J. Xu, M. Ng, and Y. Ye. (2008). *Weighting Method for Feature Selection in K-Means*. H. Liu, H. Motoda (Eds.), *Computational Methods of Feature Selection*, Chapman and Hall/CRC, pp 193-209, 2008.
- [17] UCL repository of machine learning databases and domain theories. FTP address: www.ics.uci.edu/~mllearn