Improvement of Fuzzy C-Means by using variance-based weighted Feature

Maryam Kamjou
Shiraz University of Technology, Shiraz, IRAN

Marzieh Ahmadzadeh
Shiraz University of Technology, Shiraz, IRAN

Abstract – Fuzzy algorithm is one of the most powerful clustering algorithm, but in analysis maybe some features importance than other. Therefore to solve this problem, weighted features C-mean clustering provided. Nevertheless, in weighted features clustering algorithm they are some problem. For example, during the training phase the weights are not automatically weighted. This paper shows that the feature-weighted clustering can improve the performance of fuzzy c-means clustering. We propose a new weighted feature fuzzy c-means clustering algorithm in which way that this algorithm be able to obtain the importance of each features. An Experimental results on two datasets, show that our proposed method have better performance than FCM algorithm.

Index Terms – Fuzzy clustering, fuzzy C-means, feature weighted, weighted Fuzzy C-means.

1. INTRODUCTION

Clustering Finding groups of objects that be similar to another and different from the objects in other groups. [1] In our proposed method we offer weighted features C-mean clustering methods with variances [2,3]. These experiments show that when we use variances for weighting, which can improve the clustering performance. Among the fuzzy clustering method, the fuzzy c-means algorithm [4] that studied in many researches area, for example: in pattern recognition, machine learning, image segmentation and image clustering [5], wireless sensor network [6] data mining [7]. Clustering can be classified into 5 categories: Hierarchical methods, Partitioning methods, Grid-based methods, Density-based methods, Model-based methods that c-mean is one of the Partitioning methods [8]. In our proposed method, weighted feature and clustering phase dynamically updates and weighted Fuzzy c-mean algorithm using variance [9].

This paper is organized as follows: In section 2 we introduce the related works that have been done by other researchers. In Section 3, first we explain about Fuzzy C-means clustering, then propose our method and update formula. In section 4 we present our experimental results and in Section 5 we conclusions from our discussion.

2. RELATED WORK

FCM algorithm was first introduced in 1974 by Dunn. [10] And then developed by Jim Bezdek in 1980[4], fuzzy c-means clustering studied in many cases and it's one of the most important method in clustering. In recent years many improved versions of it have been proposed [11, 12]. In most fuzzy clustering algorithms, the features be considered the same. While in real, some features are more important than other. So for make cluster, those features that significantly more effective, should be considered more important than others. In recent years, a variety of fuzzy clustering algorithm is presented by weighting features. Wang et al [8] proposed a feature-weighted based on a defined similarity measure and an evaluation function to improve the performance of FCM.

3. PORPOSED MODELLING

In this section we explain:

- Fuzzy C-means clustering
- Proposed method
- Update formula

3.1. Fuzzy C-means clustering

Fuzzy C-means is one of the most commonly clustering algorithms. The samples are divided into C clusters. It [8, 9] allows one sample of data belongs to more than one clusters according to a membership function.

$D = \left\{X_j\right\}_{j=1}^{N}$ is the dataset that $X_j=(x_1,x_2,\ldots,x_n)$ be a set of numerical data in $\mathcal{R}^d$. "m" is the degree of fuzziness associated with the partition matrix (m>1).

The fuzzy c-means algorithm, minimizing the objective function J defined as follows [4]:

$$(1)\ \min\{\ J(U,\ V;D)\ = \sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}\ \|X_j - V_i\|^2\}$$

Where $\mathrm{U} = \left(\mu_{\mathrm{ij}}\right)_{\mathrm{C}\times\mathrm{N}}$ is a fuzzy partition matrix that 1<i<c and 1<j<n.

$\mathrm{V} = \left(\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_\mathrm{C}\right)^\mathrm{T} = \left(\mathrm{v}_{\mathrm{iq}}\right)_{\mathrm{C}\times\mathrm{d}}$ is the cluster center.

The update formula of membership μ_{ij} is as follows:

(2) $\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{d_{ij}^2}{d_{kj}^2}\right)^{\frac{1}{m-1}}}$ $\mu_{ij} \grave{o} \left[0,1\right]$

The update formula of $\mathrm{V_i}$ is as follows:

(3) $V_i = \frac{\sum_{j=1}^{N}\mu_{ij}^m X_j}{\sum_{j=1}^{N}\mu_{ij}^m}$

Step 1 Choose c and a threshold value ε and Initial the fuzzy partition matrix U
Step 2 Compute v according to
Step 3 *Compute d_{ij}* and Thus update U
Step 4 Compute the objective function J by using (2.1). if $\lvert \mu_{\mathrm{ij}}^{\mathrm{t}} - \mu_{\mathrm{ij}}^{\mathrm{t-1}} \rvert < \varepsilon$ otherwise go to step 2

Table 1 The FCM clustering

3.2. Proposed method

The initialization clusters centers and weighted feature in this algorithm has great influence on the final result. So instead of the initializing random weighting it is better to use feature selection criteria such as the distribution index data. The weighting of features makes them closer to the actual value of the features and improve performance clustering method. On the other hand it makes the weights real than first and increases the performance and accuracy. Variance is the simplest measure of unsupervised for evaluating the characteristics. So it is defined by:

(4) $\sigma_N^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$

That x_i is the number of features and $\lvert N \rvert$ is the total number of samples.

3.3. Update formula

3.4. $D = \left\{X_j\right\}_{j=1}^{N}$ is a dataset that X_j belong to R^d.

The fuzzy c-means algorithm, minimizing the objective function J defined as follows:

(5) $J_m\left(U, V, w : D\right) = \sum_{i=1}^{C}\sum_{j=1}^{N}\left(\mu_{ij}\right)^m \left[d_{ij}^{(w)}\right]^2$

where d_{ij} is the Euclidean distance with $w = \left(w_1, w_2, \ldots, w_d\right)^T$

for all features is computed by:

(6) $d_{ij}^{(w)} = \lvert diag\left(w\right)\left(X_j - V_i\right)\rvert$

"d_{ij}" is the Euclidean distance from sample $\mathrm{X_j}$ to the cluster center $\mathrm{V_i}$.

And

$$(7)\, diag(w) = \begin{pmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & w_d \end{pmatrix}.$$

And

(8) $\sum_{q=i}^{d} w_w = 1.$

Therefore V_i is centers cluster calculated as:

(9) $v_i = \frac{\sum_{j=1}^{n}\mu_{ij}^m X_j}{\sum_{j=1}^{n}\mu_{ij}^m}$

Membership function defined as follows:

(10) $\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}^2}{d_{kj}^2}\right)^{1/(m-1)}}$

W_k the feature-weight that defined as follows:

(11) $W_k = \frac{1}{\sum_{l=1}^{d}\left[\frac{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^m\left(x_{jk} - v_{ik}\right)^2}{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^m\left(x_{jl} - v_{il}\right)^2}\right]}$

The procedure of the FWCM algorithm is summarized in table 2.

```
Input: Data set $D=\{X_j\}_{j=1}^{N}$

Output: Terminal fuzzy partition matrix U and terminal fuzzy feature-weight vector w.
Initialization: Initialization Feature-weight using variance and C, m, ε
    $U=(\mu_{ij})_{C\times N}$

Step 1: Compute the cluster centers:
    $v_i=\frac{\sum_{j=1}^{n}\mu_{ij}^{m}X_j}{\sum_{j=1}^{n}\mu_{ij}^{m}}$

Step 2: Calculate the distances as $d_{ij}^{(w)}$
    $d_{ij}^{(w)}=|\,diag(w)(X_j-V_i)\,|$

Step 3: Update the fuzzy partition matrix:
    $\mu_{ij}=\frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}^{2}}{d_{kj}^{2}}\right)^{1/(m-1)}}$

Step 4: Update the elements in the feature-weight vector:
    $W_q=\frac{1}{\sum_{l=1}^{d}\left[\frac{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jq}-v_{iq})^{2}}{\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{m}(x_{jl}-v_{il})^{2}}\right]}$

End for
Until $|\mu_{ij}^{t}-\mu_{ij}^{t-1}|<\varepsilon$
```

Table 2 The FWCM algorithm by using variance

In this algorithm we have D dataset in input and in the output is two matrix, U and terminal fuzzy feature-weight and cluster center "V". First W and V calculate and then M and C define by user. After initialization, first Compute center cluster (V) then calculate d_{ij} and step 3 updates the fuzzy partition matrix and step 4 updates feature-weighted until $|\mu_{ij}^{t}-\mu_{ij}^{t-1}|<\varepsilon$.

4. RESULTS AND DISCUSSIONS

In this experiment, two data sets are utilized. The description of the two data sets is as follows:

Iris data set: This data set consists of 150 samples with three classes [13].

Bupa data set: This data set consists of 345 samples with two classes (Table 3).

Number classes	Number Features	Number samples	Dataset
3	4	150	Iris
2	6	345	Bupa

Table 3 Two benchmark data sets

Moreover, the fuzzification exponent m is set to be 1.5, while the termination Tolerance ε is still 10^{-2}.

Figure 1 shows the graphical data clustering for each data collection including Iris. In this figures different clusters shows the different colors.

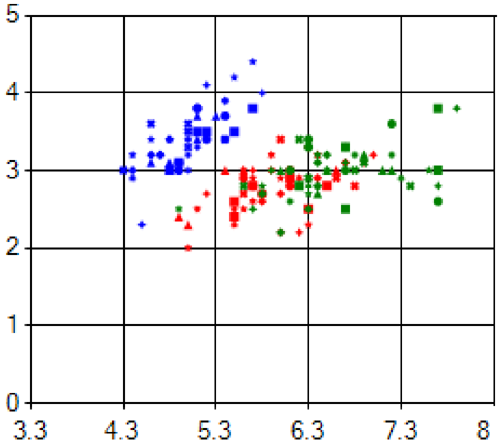

Figure 1 Graphical clustering data set Iris

Figure 2 shows the graphical data clustering for two numbers of classes of Iris. In this figures different clusters is shows the different colors.

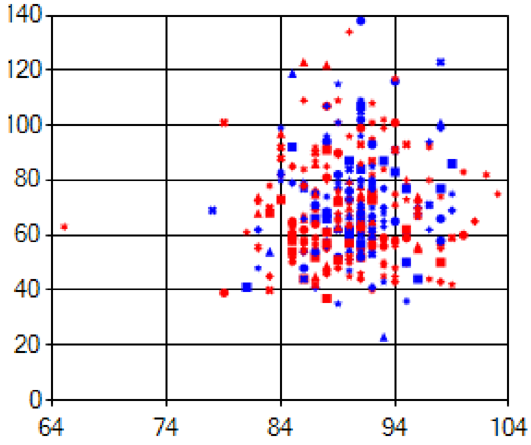

Figure 2 Graphical clustering data set Bupa

In this section we compare our method from FCM method. In table 2 and table3 indicate the error rates and run times. It can be easily observed that the proposed method is the best of all.

For the Iris data set and Bupa dataset, our proposed method is better than other method because their error rates are all less than that of FCM. We had repeated this algorithm 10 times. Our result defied as follows:

Time (ms)	Error	Method
13.1	8.97	FCM
96.8	2.49	Proposed method

Table 4 Comparing clustering results from Iris data set

Table 4 shows the error and run time of the FCM method and our proposed method in Iris dataset. It shows that our proposed method has longer computing than FCM method but in these algorithms, error rate is more important than run time. For example, the error rate of FCM method is 8.97 but in our proposed method the error rate is 2.49. So it shows that our proposed method has been functioning properly.

Time (ms)	Error	Method
77	44018.12	FCM
212.2	761.67	Proposed method

Table 5 Compare clustering results from Bupa data set

Table 5 shows the error and run time of FCM method and our proposed method in data collection Bupa. It shows that our proposed method has longer computing than FCM method but in these algorithms, error rate is more important than run time. For example the error rate of FCM method is 44018.12 but in our proposed method the error rate is 761.67. So our proposed method has the best performance.

5. CONCLUSION

FCM is one of the most important clustering algorithms. In this paper we propose Weighted FCM algorithm which is initializing Feature-weighted using term variance and reformulates the objective function and the update Formula from the existing FCM. It shows that our method can be improves the performance of FCM clustering.

REFERENCES

[1] P. Tan, M. Steinbach, V. Kumar, “Introduction to Data Mining”, Pearson Addison-Wesley, 2006, pp. 488-555.

[2] C. Borgelt and A. N¨urnberger, “Experiments in Document Clustering using Cluster Specific Term Weights” 2004

[3] C. Borgelt and A. Nurnberger, “Fast Fuzzy Clustering of Web Page Collections” Proc. PKDD Workshop on Statistical Approaches for Web Mining SAWM, Pisa, Italy. 2004.

[4] J.C. Bezdek, “A convergence theorem for the fuzzy ISODATA clustering algorithms”, IEEE Trans. Patt. Anal. Mach. Intell. 2 (1) 1980,pp 1–8.

[5] M. Gong Y, Y Liang,W Ma and J.Ma,J. IEEE Transactions on Image Processing. Vol. 22 NO. 12, 2013.

[6] D. C. Hoang, R. Kumar and S. K. Panda, J. “IET Wireless Sensor Systems”, vol. 3, no. 3, 2012.

[7] X. Yang, J. Lu and J. Ma, J. IEEE “Transactions on Fuzzy Systems”, vol. 19, no. 1, 2011.

[8] Wang, X.; Wang, Y.; Wang, L. “Improving fuzzy c-means clustering based on feature-weight learning” Pattern Recognit. Lett. 2004, 25, pp.1123–1132.

[9] H.Jie Xing , M.Hu Ha,” Further improvements in Feature-Weighted Fuzzy C-Means” Information Sciences 267,2014,pp 1–15.

[10] J.C. Dunn, “Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems”, J. Cybernet. 4,1974.

[11] B.K. Brouwer, “A method of relational fuzzy clustering based on producing feature vectors using FastMap”, Inform. Sci. 179 2009, pp. 3561–3582.

[12] C. Borgelt, “Accelerating fuzzy clustering”, Inform. Sci. 179 2009 pp. 3985–3997.

[13] C.L. Blake, C.J. Merz, “UCI Repository of Machine Learning Databases”, Department of Information and Computer Science, University of California, Irvine, CA, USA, 1998 <http://www.ics.uci.edu/mlearn/MLRepository.htm>.