International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:4, No:5, 2010

# Parameter Selections of Fuzzy C-Means Based on Robust Analysis

Kuo-Lung Wu

***Abstract*—**The weighting exponent m is called the fuzzifier that can have influence on the clustering performance of fuzzy c-means (FCM) and m∈ [1.5,2.5] is suggested by Pal and Bezdek [13]. In this paper, we will discuss the robust properties of FCM and show that the parameter m will have influence on the robustness of FCM. According to our analysis, we find that a large m value will make FCM more robust to noise and outliers. However, if m is larger than the theoretical upper bound proposed by Yu et al. [14], the sample mean will become the unique optimizer. Here, we suggest to implement the FCM algorithm with m∈ [1.5,4] under the restriction when m is smaller than the theoretical upper bound.

***Keywords*—**Fuzzy c-means, robust, fuzzifier.

## I. Introduction

In fuzzy clustering, the fuzzy c-means (FCM) algorithm is the best-known method [1,2]. Although FCM is a good clustering algorithm, there are some drawbacks when we apply FCM. Therefore, many extensions to the FCM algorithm had been proposed in the literatures. Overall, these extended types of FCM can be divided into two categories. One is to extend the dissimilarity (or distance) measure between the data point and the cluster center in the FCM objective function by replacing the Euclidean distance with the other types of metric measures (see Refs. [3,4,5,6,7]). Another one is to extend the FCM objective function by adding a penalized term (see Refs. [8,9,10,11]).

Another important influence factor to the effectiveness of FCM is the weighting exponent m which had been well investigated by Pal and Bezdek [13] and Yu et al. [14]. Pal and Bezdek [13] suggested to take m∈ [1.5,2.5] and Yu et al. [14] proposed a theoretical upper bound for m that can prevent the sample mean being the unique optimizer of FCM objective function. In this paper, we will analysis the robustness of FCM based on the statistical point of view and show that FCM can be robust to noise and outliers in a large m case. In Section II, we brief review the FCM clustering method. We also discuss the parameter selections of FCM in Section III. In Section IV, we will discuss the robust properties of FCM and show that the parameter m will have influence on the robustness of FCM. We find that a large m value will make FCM more robust to noise and outliers. Conclusions are illustrated in Section V.

K.L. Wu is with the Department of Information Management, Kun Shan University, Yunk-Kang, Tainan 71023, Taiwan. (phone: +886 (6) 2050-149; e-mail: klwu@mail.ksu.edu.tw).

## II. Fuzzy C-Means Clustering Algorithm

Let $X = \{x_1, \cdots, x_n\}$ be a data set in an s-dimensional space $R^s$. Let c be a positive integer greater than one. A partition of X into c parts can be presented by mutually disjoint set $X_1, \cdots X_c$ such that $X_1 \cup \cdots \cup X_c = X$ , or equivalently by the indicator functions $\mu_1, \cdots \mu_c$ such that $\mu_{ij} = \mu_i(x_j) = 1$ if $x_j \in X_i$ and $\mu_{ij} = 0$ if $x_j \notin X_i$ for $i = 1, \cdots, c$ and $j = 1, \cdots, n$ . The set of indicator functions $\mu_1, \cdots \mu_c$ is called a hard c-partition of clustering X into c clusters. Now consider an extension to allow $\mu_{ij} = \mu_i(x_j) \in [0,1]$ to be membership functions of fuzzy sets $\mu_i$ on X such that $\sum_{i=1}^{c} \mu_{ij} = 1$ for all $x_j$. In this section, we will give a brief review of the best-known fuzzy clustering method fuzzy c-means (FCM) and then discuss the parameter selections in FCM. We will also discuss the influence of fuzzifier m on the robustness of FCM.

In unsupervised learning clustering literatures, the fuzzy c-means (FCM) algorithm is the best-known fuzzy clustering method. The FCM is an iterative algorithm using the necessary conditions for a minimizer of the objective function $J_{FCM}$ with

$$J_{FCM}(\mu, a) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^m \left\| x_j - a_i \right\|^2 \tag{1}$$

where the weighting exponent m>1 is a fuzziness index, $\mu = \{\mu_1, \cdots \mu_c\}$ with $\mu_{ij} = \mu_i(x_j)$ is a fuzzy c-partition and $a = \{a_1, \cdots, a_c\}$ is the set of c cluster centers. The necessary conditions for a minimizer $(\mu, a)$ of $J_{FCM}$ are the following update equations:

$$\mu_{ij} = \frac{\left\| x_j - a_i \right\|^{-2/(m-1)}}{\sum_{k=1}^{c} \left\| x_j - a_k \right\|^{-2/(m-1)}} \tag{2}$$

and

$$a_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_j}{\sum_{j=1}^{n} \mu_{ij}^m}, \tag{3}$$

Note that, $d(x_j, a_i) = \left\| x_j - a_i \right\|^2$ is most used. However, other types of metric may be used to improve the usage and effectiveness of FCM (see Refs. [3,4,5,6,7]). On the other hand, another important factor influencing the effectiveness of FCM is the fuzziness index m, which has previously been thoroughly

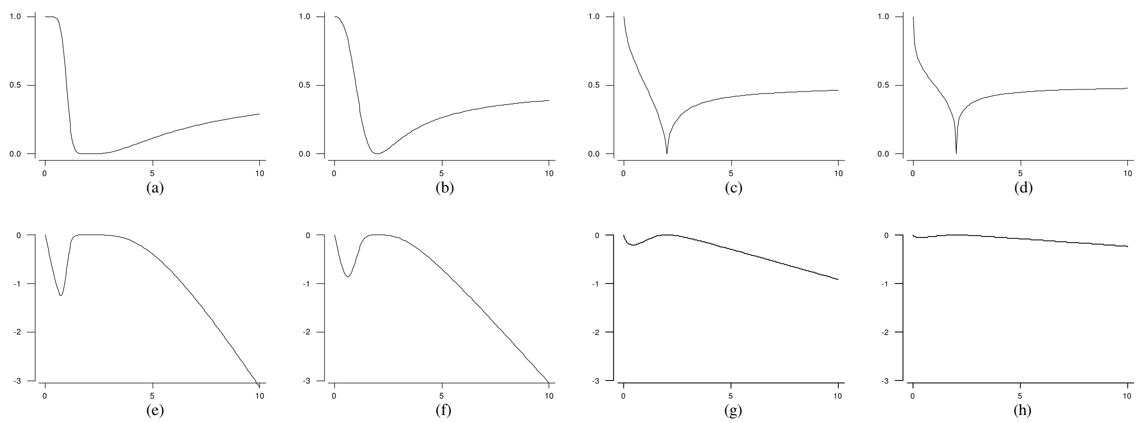


Fig. 1. The membership functions and $\varphi$ functions of FCM. (a), (b), (c) and (d) are membership functions with different fizzifier m=1.5, 2, 4 and 6, respectively. (e), (f), (g) and (h) are $\varphi$ functions with different fizzifier m=1.5, 2, 4 and 6, respectively.

TABLE I
THE UPPER LIMIT OF M FOR THE DATA SETS

| Data set | Number of Samples | Number of Features | Number of Clusters | $\lambda max$ (Cx) | Upper limit of $m$ |
|---|---|---|---|---|---|
| Isolet 1+2+3+4 | 6238 | 617 | 26 | 0.1889 | 1.6072 |
| Isolet 5 | 1559 | 617 | 26 | 0.1926 | 1.6265 |
| Sonar | 208 | 60 | 2 | 0.1949 | 1.6388 |
| Vowel | 990 | 10 | 11 | 0.2189 | 1.7787 |
| PimaIndiansDiabetes | 768 | 8 | 2 | 0.2558 | 2.0475 |
| Waveform | 5000 | 21 | 3 | 0.3272 | 2.8935 |
| Glass | 214 | 9 | 6 | 0.3424 | 3.1726 |
| Iris | 150 | 4 | 3 | 0.6652 | $+\infty$ |

investigated in Pal and Bezdek [13] and Yu et al. [14].

## III. Parameter Selections in FCM

The weighting exponent m is called the fuzzifier that can have influence on the clustering performance of FCM. The influence of the weighting exponent m on the FCM membership function is shown in Fig. 1. This figure is produced by assuming that there are only two clusters with centers 0 and 2. The curves with different m values are the membership functions belonging to the cluster with center 0. When m = 1, the FCM will reduce to the traditional hard c-means. When m tends to infinity, $\mu_{ij} = 1/c$ for all i, j and the sample mean will be a unique optimizer of FCM objective function. In fact, this situation may occur for any specified m values and Yu et al. [14] proposed a theoretical upper bound for m that can prevent the sample mean from being the unique optimizer of FCM objective function. The rule is that $\forall i,\ a_i = \bar{x}$ is stable for FCM if $\lambda_{\max}(C_X) < 0.5$ and $m \geq (1 - 2\lambda_{\max}(C_X))^{-1}$, where $C_X = \sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})^T \Big/ n\|x_j - \bar{x}\|^2$ and $\lambda_{\max}(C_X)$ is the maximum eigenvalue of the matrix $C_X$. Therefore, for FCM, we should set $m < (1 - 2\lambda_{\max}(C_X))^{-1}$ if $\lambda_{\max}(C_X) < 0.5$. If $\lambda_{\max}(C_X) \geq 0.5$, we can take m to be any positive values and $m \in [1.5, 2.5]$ is suggested by Pal and Bezdek [13].

Table I (refer to Yu et al. [14]) shows the upper limit of m for some data sets obtained from the UCI Repository of Machine Learning Databases. For the data sets with $\lambda_{\max}(C_X) < 0.5$, sample mean will be the unique optimizer of FCM when the fizzifier $m \geq (1 - 2\lambda_{\max}(C_X))^{-1}$ (the upper limit). The traditional specified m for FCM is 2 and we find that there are four data sets with the upper limit of m smaller than 2. In IRIS data set [15,16], since the maximum eigenvalue of the matrix $C_X$ is larger than 0.5, the upper limit of m for the Iris data set is positive infinity. Yang and Wu [11] also confirmed a part of these results. We now give a simple example to demonstrate above properties.

We implement the Normal-4 data set to test the inferences of m on the performances of FCM. Pal and Bezdek [13] proposed the Normal-4 data set, which is a 4-dimensional data set with the sample size n=800 and each of four clusters contains 200 points. The population mean vectors are $c_1 = (3,0,0,0)$, $c_2 = (0,3,0,0)$, $c_3 = (0,0,3,0)$ and $c_4 = (0,0,0,3)$. The covariance matrix for each population is the identity matrix $I_4$. We randomly generate 100 Normal-4 data sets and implement the FCM algorithm for each one of the data set with the parameters m=1.5, 2, 2.5, 3 and 3.5. We then calculate the average MSE and average number of iterations for these 100

TABLE II
MEAN SQUARE ERROR (MSE) AND NUMBER OF ITERATIONS OF 100 RANDOMLY GENERATED NORMAL-4 DATA SETS

| | MSE | | | Number of iterations | | |
|---|---|---|---|---|---|---|
| | best case | average | worst case | best case | average | worst case |
| m=1.5 | 0.05791 | 0.14554 | 0.33201 | 10 | 14.93 | 23 |
| m=2 | 0.16153 | 0.45912 | 0.91060 | 13 | 20.20 | 35 |
| m=2.5 | 3.87754 | 5.95604 | 6.52174 | 50 | 50.00 | 50 |
| m=3 | 6.67529 | 6.74813 | 6.79683 | 36 | 42.24 | 50 |
| m=3.5 | 6.73916 | 6.75495 | 6.78513 | 27 | 30.80 | 36 |

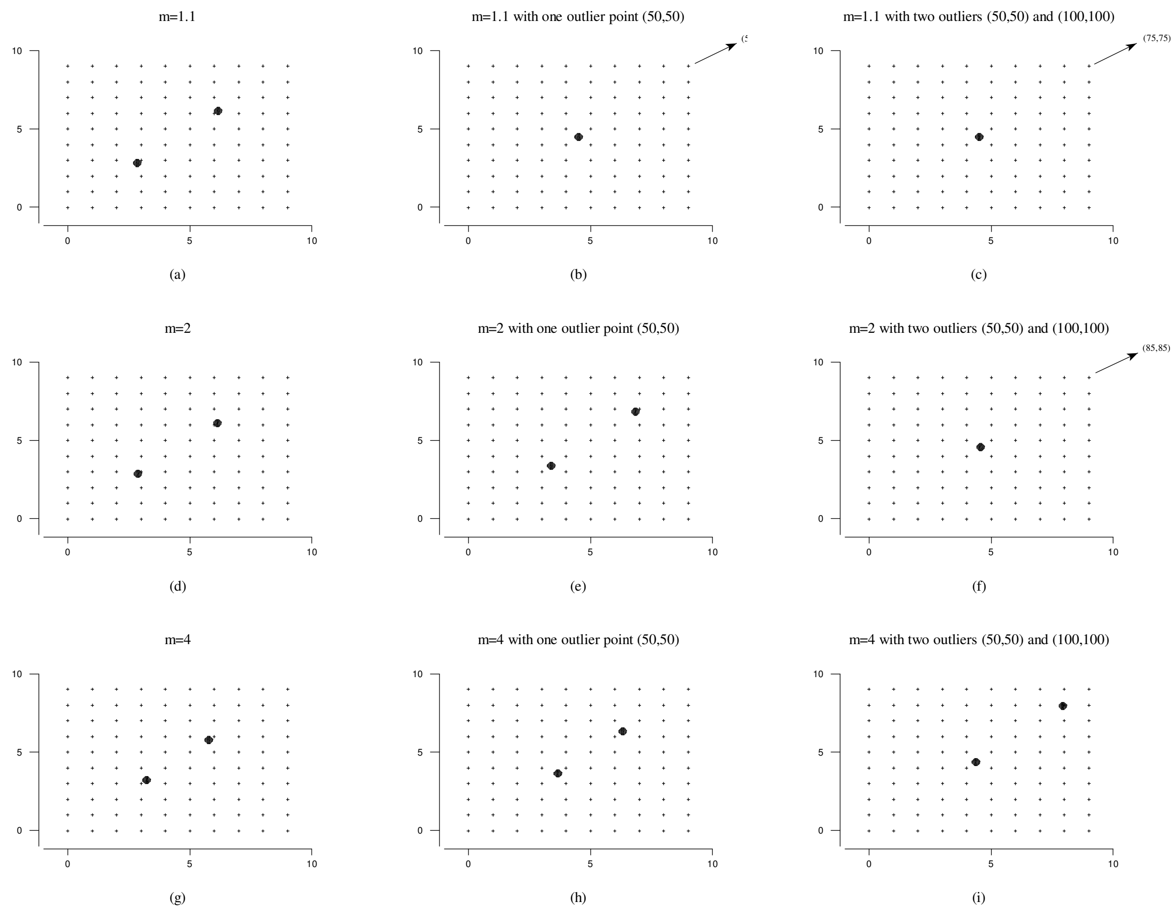


Fig. 2. Clustering results of FCM with grid points.

randomly generated Normal-4 data sets. The results are shown in Table II. Yu et al. [14] concluded that the theoretical valid m for a random Normal-4 data set should be not greater than 2.6 which sample mean is an optimizer with approximately 50% probability. In Table II, when m=2.5, the worst case is the sample mean being the unique optimizer with MSE=6.52. When m=3 and 3.5, the FCM always produce the unique optimizer $\bar{x}$ . These results are coincident to Yu et al. [14]. Although a too large m may cough trouble in FCM, a suitable large m value can make FCM more robust to noise and outlier. We will discuss this property in next subsection.

## IV. Robust Analysis of Fuzzy C-Means

The influence curve (IC) can help us to assess the relative influence of an individual observation toward the value of an estimate. The influence function of an M-estimator is proportional to its $\varphi$ function [17]. If the influence function of an estimator is unbounded, an outlier might cause trouble where the $\varphi$ function is used to denote the degree of the influence. Let the loss between the data point $x_j$ and cluster center $a_i$ be

$$\rho(x_j - a_i) = \mu_{ij}^m \left\| x_j - a_i \right\|^2 \tag{4}$$

and

$$\varphi(x_j - a_i) = \frac{d}{da_i}\rho(x_j - a_i) = -2\mu_{ij}^m \left( x_j - a_i \right) \tag{5}$$

By solving the equation $\sum_{j=1}^{n} \varphi(x_j - a_i) = 0$ , we have the result shown in Equation (3). Thus, the FCM cluster center estimate is an M-estimator with the loss function (4) and $\varphi$ function (5). Refer to Figs. 1(a)~1(d), the corresponding $\varphi$ functions with different m are illustrated in Figs. 1(e)~1(h). The influences of adding a point on the cluster center 0 will become very small when m is large. That is, FCM can be robust to noise and outliers with a large m value. Figure 2 is a simple example to illustrate this phenomenon.

This is an artificial data set with grid points. We implement FCM with different m values and the results are shown in Figs. 2(a), 2(d) and 2(g). We then add an outlier point in the coordinate (50,50) and the results are shown in Figs. 2(b), 2(e) and 2(h). Moreover, we add one more outlier point in the coordinate (100,100) and the results are shown in Figs. 2(c), 2(f) and 2(i). These results are coincident to the phenomenon illustrated in Fig. 1. We also implement the IRIS data set [15,16] to test the influences of parameter m on the results of FCM. The IRIS data set has n = 150 points in an s = 4 dimensional space. It consists of three clusters wih Iris Setosa, Iris Versicolor and

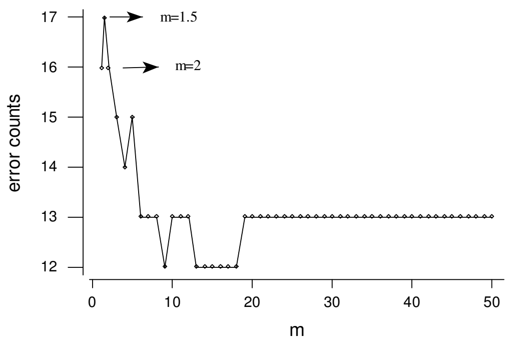


Fig. 3. The error counts of IRIS data set.

Iris Virginica. Two clusters have substantial overlapping. The clustering error counts of this data are approximate 16. Figure 3 shows the clustering error counts of the IRIS data obtained by FCM with different m values (m=1.1, 1.5, 2, 3, 4, …, 50). The error counts are approximate 13 when m is large. These examples reveal the robust properties of FCM when m becomes large.

## V. Conclusions

Note that, when m tends to infinity, the $\varphi$ function of a finite point $x_j$ will tend to 0. However, in real application, we never observe an infinity data point. We then have

$$\lim_{m\to\infty} \varphi(x_j - a_i) = 0 \tag{6}$$

for a real data point. That is, a very large m value will make FCM very robust. However, this is not a good guideline for selecting m in FCM. Although FCM becomes very robust in a large m case, the membership value for each data point will very close to 1/c in this case and the sample mean will become the unique optimizer. Figure 1(d) also shows that the membership values for the data points becomes closed to 0.5 when m=6. Here, we suggest implementing FCM with $m\in[1.5,4]$ under the restriction that m is smaller than the theoretical upper bound proposed by Yu et al. [14].

## Acknowledgment

This work was supported in part by the National Science Council of Taiwan, under Kuo-Lung Wu's Grant: NSC-98-2118-M-168-001.

## References

[1] J.C. Bezdek, *Pattern Reccognition with Fuzzy Objective Function Algorithm*, Plenum Press, 1981.
[2] M.S. Yang, "A survey of fuzzy clustering," *Math. Comput. Modelling*, vol.18, pp. 1-16, 1993.
[3] J. Yu, "General c-means clustering model," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, pp. 1197-1211, 2005.
[4] K.L. Wu and M.S. Yang, "Alternative c-means clustering algorithms," *Pattern Recognition*, vol. 35, pp.2267-2278, 2002.
[5] J. Leski, "Towards a robust fuzzy clustering," *Fuzzy Sets and Systems*, vol. 137, pp.215-233, 2003.
[6] H. Frigui and R. Krishnapuram, "A robust competitive clustering algorithm with applications in computer vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, 450-465, 1999.
[7] J. Yu and M.S. Yang, "Optimality test for generalized FCM and its application to parameter selection," *IEEE Trans. Fuzzy Systems*, vol. 13, pp. 164-176, 2005.
[8] D. Özdemir and L. Akarun, "A fuzzy algorithm for color quantization of images," *Pattern Recognition*, vol. 35, pp.1785-1791, 2002.
[9] D. Özdemir and L. Akarun, "Fuzzy algorithms for combined quantization and dithering," *IEEE Trans. Image Processing*, vol. 10, pp. 923-931, 2001.
[10] R. Krishnapuram and J.M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Systems*, vol. 1, pp. 98-110, 1993.
[11] M.S. Yang and K.L. Wu, "Unsupervised possibilistic clustering," *Pattern Recognition*, vol. 39, pp. 5-21, 2006.
[12] J.C. Bezdek, "Cluster validity with fuzzy sets," *J. Cybernet.*, vol. 3, pp. 58-73, 1974.
[13] N. R. Pal and J. C. Bezdek, "On cluster validity for fuzzy c-means model," *IEEE Trans. Fuzzy Systems*, vol. 1, pp. 370-379, 1995.
[14] J. Yu, Q. Cheng and H. Huang, "Analysis of the weighting exponent in the FCM," *IEEE Trans. Systems, Man, Cybern., Part B*, vol. 34, pp. 634-639, 2004.
[15] E. Anderson, "The Irises of the gaspe peninsula," Bull. Am. IRIS Soc. 59, pp. 2–5, 1935.
[16] J.C. Bezdek, J.M. Keller, R. Krishnapuram, L.I. Kuncheva, N.R. Pal, "Will the Iris data please stand up?," *IEEE Trans. Fuzzy Syst*. Vol. 7, pp. 368–369, 1999.
[17] P.J. Huber, *Robust Statistics*, Wiley, 1981.

**Kuo-Lung Wu** received the BS degree in mathematics in 1997, the MS and PhD degrees in applied mathematics in 2000 and 2003, all from the Chung Yuan Christian University, Chungli, Taiwan. In 2003-2006, he was an Assistant Professor in the Department of Information Management at Kun Shan University of Technology, Tainan, Taiwan. He has been an Associate Professor at the same university. He is a member of the Phi Tau Phi Scholastic Honor Society of Taiwan. His research interests include fuzzy theorem, cluster analysis, pattern recognition and neural networks.