An Insensitivity Fuzzy C-means Clustering Algorithm Based on Penalty Factor

Jiashun Chen 1,2
1Huaihai Institute of Technology /College of Computer Engineering Lianyungang China
2Nanjing University of Aeronautics and Astronautics /College of Information Science and Technology, Nanjing China
Email:1976139@163.com
Dechang Pi 2, Zhipeng Liu 2,3
2Nanjing University of Aeronautics and Astronautics /College of Information Science and Technology, Nanjing China
3Nanjing University of Posts and Communications /College of Computer Science and Technology, Nanjing, China
Email: dc.pi@nuaa.edu.cn, Email: liuzhipengcs@139.com

Abstract—This paper analyzes sensitivity of Fuzzy C-means to noisy which generates unreasonable clustering results. We also find that Fuzzy C-means possess monotonicity, which may generate meaningless clustering results. Aiming at these weak points, we present an improved Fuzzy C-means named IFCM (Improved Fuzzy C-means). Firstly, we research the reason of sensitivity and find that constraint leads to sensitivity of algorithm, we propose abolish constraint; secondly, we replace membership with typicality for acquiring more reasonable clustering results; finally, we add penalty factor to objective function to avoid monotonicity and coincident clustering results. On the basis of these, we improve objective function and provide step of algorithm. Experiments on various datasets show that new algorithm recognizes noisy data effectively and makes cluster effect improve furthermore.

Index Terms—IIFCM, membership, constraint, noisy data

I. INTRODUCTION

Cluster analysis as a branch of data mining has been applied in many fields. The target of cluster is partition sample data set into finite class, and makes intra-class maximal similarity and inter-class maximal difference. More and more clustering algorithms are coming out. Current clustering algorithms are thought to be of two kinds: hard cluster and soft cluster. Hard cluster is an absolute partition. Many algorithms are belong to this type, such as k-means [1], and improved k-means [2-4], and STING [5-6] and so on. K-Means is hard partition, which partition sample into some cluster strictly, that is to say sample only belongs to one cluster. This method hardly copes with some fuzzy problems. After fuzzy set theory was introduced, Bezdek [7] proposed fuzzy cluster method which partitioned data set softly and abolished the idea of either 0 or 1, and used membership to denote relation between sample and class. In FCM, there are some constraints such as \( \sum_{i=1}^{c} u_{ij} = 1, (j = 1, 2, \ldots, n) \), \( 0 < \sum_{i=1}^{c} u_{ij} < n \) which make clustering results become hard to interpret and understand. Clustering results are sensitive to noisy data and do not recognize outlier. Fig.1 is the Dataset \( X_{12} \) distribution on coordinate axis, which shows that sample data \( x_i \) and \( x_{12} \) are noisy. The distances from two points to two centroids are different, however, the membership of two points are the same. In fact, membership of \( x_i \) should far less than \( x_{12} \), therefore, clustering results are difficult to interpret and understand. FCM is sensitive to noisy and could not recognize outlier.

<table>
<thead>
<tr>
<th>dataset</th>
<th>( U_i )</th>
<th>( U_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>0.0636</td>
<td>0.9364</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>0.0327</td>
<td>0.9673</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>0.0103</td>
<td>0.9897</td>
</tr>
<tr>
<td>( x_4 )</td>
<td>0.1006</td>
<td>0.8994</td>
</tr>
<tr>
<td>( x_5 )</td>
<td>0.0844</td>
<td>0.9156</td>
</tr>
<tr>
<td>( x_6 )</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>( x_7 )</td>
<td>0.9156</td>
<td>0.0844</td>
</tr>
<tr>
<td>( x_8 )</td>
<td>0.9673</td>
<td>0.0327</td>
</tr>
<tr>
<td>( x_9 )</td>
<td>0.9897</td>
<td>0.0103</td>
</tr>
<tr>
<td>( x_{10} )</td>
<td>0.8994</td>
<td>0.1006</td>
</tr>
<tr>
<td>( x_{11} )</td>
<td>0.9364</td>
<td>0.0636</td>
</tr>
<tr>
<td>( x_{12} )</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
</tbody>
</table>

Figure 1. Dataset \( X_{12} \) distribution on coordinate axis.

Above analysis shows that especial sample data are required for FCM algorithm. FCM could not cope with abnormal data in noisy environment, which show FCM is
sensitive to outlier. In order to overcome deficiency, we present a improved FCM named IFCM. Constraint of FCM has been abolished and we replace membership with typicality. At same time, penalty factor is added to objective function to avoid coincident clustering results and monotonicity. The rest of this paper is organized as follows: in section 2, relate works about FCM are introduced, and section 3 present improved algorithm. After discussion of penalty factor, we modify objective function on the basis of FCM. According to minimize objective function, we acquired iterative equation of typicality and cluster center. On the basis of these work, we provide steps of IFCM. Section 4, some experiments are conducted in different data sets, and we compare and analyze experimental results. Some conclusions are made in section 5.

II. RELATE WORK

After Bezdek [7] proposed FCM in 1981, as the research moves along, FCM has been applied in many fields. For example, FCM is applied in area of medicine to realize medical image cluster, which contribute to research of classification of diseases [9-12]. Combining FCM with self-organizing network [13] could be used in web log mining. Park et.al [14] applied FCM to deploy and optimize sensor node which improved energy efficiency and reduced cost. Power system applied FCM to forecast long-term load [15].

With increase use of FCM, improved FCM appeared a lot. The improved methods can be divided into two kinds and one is that FCM is improved by changing weight factor or fuzzy factor [16-18]. Literature [16] added weight to improve FCM. Literature [17] discussed that different fuzzy factors influence on membership. Liu [18] proposed an improved method named FCM-SM. They replace Euclidean with Mahalanobis distance to realize different type data set clustering. Unfortunately, in that paper we do not find discussion about sensitivity with noisy data, therefore, this drawback still exists. Former several improved methods change factors of FCM. The other method [19-24] is to combine FCM with other algorithms to realize their property of advantage. Literature [19] proposed an improved AFSA with adaptive Visual and adaptive step which combined artificial fish swarm algorithm with FCM to enhance outperformance of FCM. They used the ability for searching the global optimum to overcome local optimum and got better experimental results. Hesam [20] had similar target with literature [19] but they combined FCM with PSO to cope with local optimum. Prabhjot[21] et.al identified outliers and separated them from the data-set into one cluster before applying any clustering algorithm. This method scanned neighborhood and compared with threshold to determine outliers, which were similar to DBSCAN [22]. How to determine threshold is needed to further discussion. Executive time and efficiency of algorithm are hardly to ensure. Literature [23-24] combined FCM with fuzzy system to improve algorithm. Tamalika [24] improved FCM based on intuitionistic fuzzy set theory. They construct intuitionistic fuzzy set by using Yagertype[25] intuitionistic fuzzy generator, and replaced membership with intuitionistic membership and acquired intuitionistic membership iterative function. This algorithm was used to cluster medicine image and had good performance. All these algorithms cannot recognize noisy data effectively [26-28].

FCM used membership to denote relation between sample and class. FCM had good performance for some indistinct samples. FCM introduced membership; therefore objective function was modified on the basis of hard partition. Objective function was defined as follows:

\[ J(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij} D_{ij}^{2} \]  

Where \( D_{ij} = \| x_{j} - v_{i} \| \) was Euclidean distance which denoted distance from \( j \)th sample to the \( i \)th clustering center. The letter \( m \) was fuzzy factor which was used to change membership and benefited to interpret clustering results. FCM was subject to three constraints as follows:

- \( u_{ij} \in [0, 1] \) \( (i = 1, 2, \ldots, c; j = 1, 2, \ldots, n) \)
- \( \sum_{j=1}^{n} u_{ij} = 1 \) \( (i = 1, 2, \ldots, c) \)
- \( 0 < \sum_{j=1}^{n} u_{ij} < n \) \( (i = 1, 2, \ldots, c) \)

Membership and clustering center iterative function were acquired by minimizing objective function.

\[ u_{ij} = \frac{1}{\sum_{j=1}^{n} \left( \frac{D_{ij}^{2}}{D_{ij}^{2}} \right)^{\frac{1}{m-1}}} \]  

\( (i = 1, 2, \ldots, c; j = 1, 2, \ldots, n) \)

\[ v_{i} = \frac{\sum_{j=1}^{n} u_{ij} x_{j}}{\sum_{j=1}^{n} u_{ij}} \]  

\( (i = 1, 2, \ldots, c) \)

The main steps of FCM are given as follows

Step1: initializing parameters \( m \) and \( c \), and the maximal iterative times ITER_TIME, and setting the number of class \( c \), generating randomly initial membership matrix \( U^{0} \);

Step2: computing clustering center by using formula (3);

Step3: amending membership according to formula (2);

Step4: if the number of iterative times is great than ITER_TIME or \( J^{l} - J^{l-1} \) < \( \epsilon \), then stop, or go to step2, and the number of iterative times add one.

Experimental results show that FCM could not recognize noisy data, and clustering results are not consistent to the fact, and hard to understand and interpret. In order to overcome these drawbacks, we present an improved FCM based on penalty factor.

III. IMPROVED FUZZY C-MEANS(IFCM)
A. Definition Objective Function

From above analysis, we know the constraints are the main reason that FCM is sensitive to noisy and has bad performance. If we abolish constraints, according to the requirement of convergence, objective function

\[ J(U) = \sum_{i=1}^{m} \sum_{j=1}^{n} u_{ij} D_{ij} \]

has minimal value when \( u_{ij} = 0 \).

Obviously, it is meaningless. We define a new objective by replacing membership with typicality and adding penalty factor to avoid monotonicity and coincident clustering results.

**Definition 1:** Supposing sample \( X = \{X_1, X_2, \ldots, X_p\} \subset R^p \), \( R^p \) denote \( p \)-dimensional real vector space, if a fuzzy \( c \)-partition is the optimum partition, and then exiting \( T = [t_{ij}] \) and \( V = (v_1, v_2, \ldots, v_c) \) and \( c(1 \leq c \leq n-1) \) satisfy the below (formula 4) function for any \( T, V \).

\[ J(T, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} (t_{ij}^m D_{ij} + \lambda t_{ij}^m \log t_{ij}^m) \]  

(4)

Where \( m \) is fuzzy factor, and \( \lambda \) is a constant, and \( 1 \leq i \leq c, 1 \leq j \leq n \) we replace \( u_{ij} \) in FCM with \( t_{ij} \) which is the typical value that denotes the \( j \)th sample belonging to the \( i \)th class. \( T = [t_{ij}] \) is typicality matrix.

We abolish constraint \( \sum_{i=1}^{c} u_{ij} = 1 \), which makes clustering results be insensitive to noisy and be easier to interpret and understand. With the number of class increase, the first part \( \sum_{i=1}^{c} \sum_{j=1}^{n} (t_{ij}^m D_{ij}) \) in formula 4 is monotonous. The second part \( \lambda t_{ij}^m \log t_{ij}^m \) is introduced to avoid this monotonicity. According to the requirement of minimize objective function, we derive two sides of formula 4, and make \( \frac{\partial J(T, V)}{\partial t} = 0 \), and then get some formulas as follows:

\[ \frac{\partial J(T, V)}{\partial t} = mt_{ij}^{m-1}D_{ij}^m + \lambda m (mt_{ij}^{m-1} \log t_{ij}^m + t_{ij}^{m-1} \frac{1}{t_{ij}}) = 0 \]  

(5)

\[ \Rightarrow mt_{ij}^{m-1}D_{ij}^m + \lambda mt_{ij}^{m-1} (m \log t_{ij}^m + 1) = 0 \]

\[ \Rightarrow \log t_{ij}^m = \frac{-D_{ij}^m}{m \lambda} \frac{1}{m} \]

\[ \Rightarrow t_{ij} = \exp \left( -\frac{D_{ij}^m}{m \lambda} \frac{1}{m} \right) \]

(6)

\[ V_i = \sum_{j=1}^{n} \frac{t_{ij}^m x_j}{\sum_{j=1}^{n} t_{ij}^m} \]

(7)

**Theorem 1:** penalty factor that is introduced to objective function can effectively penalize monotonicity of objective function.

**Proof:** Because with the number of class increase the first part \( \sum_{i=1}^{c} \sum_{j=1}^{n} (t_{ij}^m D_{ij}) \) of objective function decreases monotonously. \( \therefore t_{ij}^m \log t_{ij}^m \leq 0 \). \( \therefore \lambda t_{ij}^m \log t_{ij}^m \geq 0 \), the second part \( \lambda t_{ij}^m \log t_{ij}^m \) can restrain to decrease while \( \lambda < 0 \). Conversely, \( \lambda > 0 \) can restrain to increase. In conclusion, \( \lambda t_{ij}^m \log t_{ij}^m \geq 0 \) penalize monotonicity of objective function.

B. Improved FCM

The main steps of improved FCM as follows:

**Algorithm 1:**

**Input:** initializing parameters \( m, \varepsilon, \lambda \) and typicality matrix \( T^0 \);

**Output:** typicality matrix \( T \) and cluster center \( V \).

**Figure 2. Flow chart of algorithm**

IV. RESULTS AND ANALYSIS

**Experiment 1:**

**Dataset:** Artificial experimental dataset

**Algorithm:** FCM, IFCM

Dataset in experiment is an artificial data which is generated randomly by using Gaussian distribution. Dataset can be divided two classes, the first class is that mean is 120 and variance is 100; the second class is that mean is 90 and variance is 100. In each class, there are 15 data points, and each point includes two attributes.

Clustering results generated by FCM and IFCM are shown in Figure 3. From Figure 3 there are two clear classes, therefore we know that FCM and IFCM have good performance on dataset. In order to test sensitivity of two algorithms to noisy, we add some noisy data to original dataset. Figure 4 shows the clustering results by FCM.
after adding noisy. Clustering centers of two classes have obviously deviated from original position in Figure 4. Clustering results generate a lot of resubstitution errors by FCM. FCM is sensitive to noisy. By contrast, clustering results in Figure 5 are better than in Figure 4. We also recognize two classes but clustering center do not deviate. Although there are some incorrect data points in Figure 5 but the number is far less than in Figure 4. IFCM evidently lowered insensitivity, and has better performance.

<table>
<thead>
<tr>
<th>algorithms</th>
<th>FCM</th>
<th>IFCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>resubstitution error</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Rate error</td>
<td>66.67%</td>
<td>13.33%</td>
</tr>
</tbody>
</table>

**Table 2.**

The number and rate of resubstitution error produced by FCM, IFCM

Experiment 2:

**Dataset:** $X_{12}$

**Algorithm:** FCM, IFCM

Dataset $X_{12}$ includes twelve data points and each data point has two attributes. Table 1 shows coordinate value of each point. Except for data point $x_6$ and $x_{12}$, the other ten points can be divided two classes which distribute two sides of $y$ coordinate. Fig. 1 shows distribution on the axis. We easily identify that data point $x_6$ and $x_{12}$ are noisy. Distances from $x_6$ to two cluster centers are equal but less than distance from $x_{12}$ to two cluster centers. Setting $\varepsilon=0.000001$ and iter_time=100, Cluster_n=2.

Table 3 shows the membership produced by FCM and IFCM. As a whole, IFCM is better than FCM about dataset partition. From table3, we know that typicality of data points $x_6$ and $x_{12}$ are equal to one, because in real data set these two points are cluster center, which shows that IFCM is easier to find cluster center. Data points $x_6$ and $x_{12}$ are noisy. Membership of two data points are the same, and do not recognize them as noisy. However, in table 3, typicality of these two data points are different, and $T(x_6) = 0.7383, T(x_{12}) = 0.3730$. Clustering results produced by IFCM are easier to understand and interpret. In conclusion, IFCM is insensitive to noisy and gets better clustering results.

**Table 3.**

Membership and Typicality produced by FCM and IFCM on $X_{12}$

<table>
<thead>
<tr>
<th>Data set</th>
<th>FCM</th>
<th>IFCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_6$</td>
<td>0.0636</td>
<td>0.9364</td>
</tr>
<tr>
<td>$x_7$</td>
<td>0.0327</td>
<td>0.9673</td>
</tr>
<tr>
<td>$x_8$</td>
<td>0.0103</td>
<td>0.9897</td>
</tr>
<tr>
<td>$x_9$</td>
<td>0.1006</td>
<td>0.8994</td>
</tr>
<tr>
<td>$x_{10}$</td>
<td>0.0844</td>
<td>0.9156</td>
</tr>
<tr>
<td>$x_{11}$</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>$x_{12}$</td>
<td>0.9156</td>
<td>0.0844</td>
</tr>
<tr>
<td>$x_{13}$</td>
<td>0.9673</td>
<td>0.0327</td>
</tr>
<tr>
<td>$x_{14}$</td>
<td>0.9897</td>
<td>0.0103</td>
</tr>
<tr>
<td>$x_{15}$</td>
<td>0.8994</td>
<td>0.1006</td>
</tr>
<tr>
<td>$x_{16}$</td>
<td>0.9364</td>
<td>0.0636</td>
</tr>
<tr>
<td>$x_{17}$</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
</tbody>
</table>

**V. CONCLUSION**

This paper discuss sensitive problem about FCM algorithm, and propose an improved Fuzzy C-means named IFCM. We firstly analyze the main reason of sensitivity in FCM and find that constraints make clustering results be unreasonable. Therefore, improved algorithm abolished constraints, and replace membership
with typicality. In order to overcome monotonicity and coincident clustering results, we add penalty factor to objective function. On the basis of these, we construct objective function. At last, we provide the main step of IFCM. In the last part, some experiments are on different data sets. Experiment one on data set generated by Gaussian distribution show that IFCM recognizes the number of class effectively and acquires better clustering result and has less resubstitution errors. Experiment two shows that IFCM distinguishes noisy from data set and is easier to find cluster center. New algorithm has good performance in noisy environment, is apt to find outlier. This work is just the first step, and there are many challenging issues discussed above. We are currently investigating into detailed issues as a further study.

ACKNOWLEDGEMENTS

This research is sponsored by Qing Lan Project and National High Technology Research and Development Program of China (863 Program, No. 2007AA01Z404).

REFERENCES


Chen Jiashun was born in 1976. He received M.S. degree in computer science and technology from China University of geosciences, in 2006. Now, he is a Ph.D. candidate in Nanjing University of Aeronautics and Astronautics. His research interests are data mining and distributed
systems.

**Pi Dechang** was born in 1971. He was a Ph.D. of Nanjing University of Aeronautics and Astronautics (NUAA) of China and now he is a professor and Ph.D. supervisor in NUAA. His research interests are data mining and database systems etc.

**Liu Zhipeng** was born in 1980. He received the B.S. and M.S. degree in computer science and technology from Nanjing University of Posts and Telecommunications, in 2002, 2005, respectively. Now, he is a Ph.D. candidate in Nanjing University of Aeronautics and Astronautics. His research interests are data mining and distributed systems.