Journal of Computers Vol. 30 No. 3, 2019, pp. 162-175
doi:10.3966/199115992019063003012

A Novel Fuzzy Clustering Approach Based on Breadth-first Search Algorithm

Ying Zhang[1*], Li-li Zhai[1], Muhammad Shahbaz[1,2]

[1] School of Economics and Management Information Systems Laboratory, Harbin University of Science and technology, Harbin 150040, China
littlezhangyi@163.com; zhailili@126.com

[2] Lyallpur Business School, Government college university Faisalabad, Punjab 38000, Pakistan
Shahbaz755@yahoo.com

Received 14 September 2017; Revised 13 January 2018; Accepted 1 March 2018

Abstract. Fuzzy C-means (FCM) clustering algorithm is one of the most popular fuzzy clustering techniques because it is efficient, straightforward, and easy to implement. However, FCM is sensitive to initialization resulting in local minimization and noise points. In this paper, a novel Fuzzy C- means clustering algorithm based on breadth-first search algorithm (BFS) and coefficient of variation weighting is proposed. Breadth-first search algorithm is a global optimization tool and it is employed to determine the appropriate initial clustering centers and eliminate the noisy data. Moreover, the objective function of FCM is improved by introducing coefficient of variation weighting for reducing noise contributions. The experimental results show that our proposed method is efficient and can overcome the defects of the traditional FCM. In addition, compared to other clustering algorithms, the new one makes convergence faster, clustering accuracy better and noise immunity higher.

Keywords: breadth-first search, coefficient of variation weighting, fuzzy C-means

1 Introduction

Cluster analysis is an important part of data mining technique, which extracts valuable information from massive data sets. It has been widely applied in many application areas such as image processing, pattern recognition, information retrieval research, bioinformatics and social network analyses, etc. [1-2]. Clustering is used to separate data into meaningful groups of similar elements. The groups are referred to as clusters, which comprise data objects that are similar to each other. The important feature of clustering is to vary as far as possible between clusters and differ as little as possible within clusters. This demarcation is absolutely strict, in that the membership belonging to each cluster is either one or zero. However, in the real world, there are many practical problems with no strict attributes, especially for data resources under big data and cloud computing background, which are normally fuzzy and uncertain in terms of behavior and attribute. In order to describe the characteristics of data resources more accurately, fuzzy clustering becomes the first choice in dealing with data resources with uncertainty and noise.

Among approachable clustering methods, the method of fuzzy C-means, introduced by Bezek [3-4], has become one of the most widely utilized methods of data analysis in recent years. The FCM approach uses a fuzzy membership which assigns a degree of membership for every class and is efficient, straightforward, and easy to implement. However, many researchers have pointed some drawbacks those are actively associated with FCM like: (1) Traditional fuzzy clustering algorithm is able to utilize only point based membership**,** which may lead to inaccurate description of data vagueness, (2) Slower Convergence speed, (3) Highly sensitive to initialization, noise and outliers, (4) Testing for fuzziness. Many improvements have been developed for the basic FCM algorithm. Krishnapuram and Keller [5]

[*] Corresponding Author

relaxed membership constraints of FCM and proposed the Possibilistic C-Means (PCM) algorithm for the noise and outliers problems of FCM. However, it can only get the global optimal solution when all clustering centers coincide. To overcome the weaknesses of FCM and PCM, Pal and Bezdek [6] proposed a Possibilistic Fuzzy C-Means (PFCM) algorithm. In addition, in order to deal with special circumstances of data and cluster centers coincide, Li and Mukaidon [7] introduced the principle of maximum entropy to FCM and proposed a clustering algorithm Maximum Entropy Clustering (MEC). M. Gong et al. introduced an improved fuzzy c-means algorithm by applying a kernel distance measure to the objective function [8]. FLICM was proposed by Krinidis and Chatzis taking advantage of a fuzzy local similarity measure, which achieves the goal of ensuring noise insensitiveness [9]. Y. Dong et al. [10] investigated an algorithm for hierarchical clustering based on fuzzy graph connectedness algorithm (FHC) which performs in high dimensional datasets and finds the clusters of arbitrary shapes such as the spherical, linear, elongated or concave ones. A modified suppressed fuzzy c-means (MS-FCM) algorithm used for both the clustering and parameter selection was proposed by Hung et al [11]. Kühne et al. developed a novel fuzzy clustering algorithm by using the observation weighting and context information for the separation of reverberant blind speech [12]. Lin et al. developed a size-insensitive integrity-based fuzzy c-means method to deal with the cluster size sensitivity problem [13].

One of the important parameters is the fuzziness index m which influences the performance of the FCM algorithm when clusters in the data set have different densities. When $m=1$, the FCM algorithm degenerates into the HCM algorithm. A good choice of m should take the data distribution of the given data set into account. Zhu *et al*. [14] presented a generalized algorithm called GIFP-FCM, which allows m not to be fixed at the usual value $m=2$ and improves the robustness and convergence. The other way to deal with the parameter m is realizing the management of uncertainty on the basis of the fuzziness index. Ozkan and Turksen [15] introduced an approach that evaluates m according to entropies after removing uncertainties from all other parameters. Hwang *et al*. [16] incorporated the interval type-2 fuzzy set into the FCM algorithm to manage the uncertainty for fuzziness index m.

To solve the problem of easy fall into local minimization caused by the random selection in center points, recently evolutionary algorithms such as simulated annealing (SA) [17], group search optimizer (GSO) [18], ant colony optimization (ACO) [19] have been successfully applied. M. S. Yang proposed penalized FCM which tries to maximize the similarity of clusters [20]. A. Siraj and R. B. Vaughn [21] used Fuzzy Cognitive Maps to create the initial clusters for preventing high impact on the final results. Izakian and Abraham [22] developed a hybrid fuzzy clustering method based FCM and fuzzy PSO to overcome the shortcomings of FCM. Another interesting contribution was collaborative FCM proposed by Pedrcycz and Rai [23].

In recent years, with the increase of requirement for classification in data mining, weighted distance for FCM has attracted research interest of many scholars [24-28]. A sample weighted FCM algorithm with affinity was developed by Gou *et al* [29]. Jiang *et al* introduced a new weight vectors calculation based on entropy to measure distance accurately and proposed a fuzzy c-means based on gene expression programming (GEP) [30]. Askari *et al* combined generalizing Entropy C-means (ECM) with Possibilistic Fuzzy C-means (PCM) for clustering noisy data [31]. Siminski proposed a fuzzy weighted C-ordered means clustering algorithm (FWCOM) to handle both various importance of attributes and outliers [32].

In their proposed method, the clusters interact with each other to improve the final results. Most clustering methods, even those with feature weighting extensions, consider all samples to have an equal weight during the clustering process. However, it is not prudent to assume that every sample in a dataset have the same weight in cluster analysis. In practice, different features may have different contributions to the cluster analysis, and thus, the contribution of noisy data will be enhanced, which leads to lack of certainty in the FCM clustering output. In order to solve the above problems, the paper proposes a novel fuzzy C-means clustering algorithm based on Breadth-First Search (BFS) and coefficient of Variation Weighting, namely BFS-VMFCM. Combined the global searching ability of BFS with effective clustering evaluation function, the proposed algorithm not only determines the appropriate initial clustering centers but also eliminate the noisy data. To further improve the performance of BFS-FCM in restraining noise and outliers, another novelty in this study is introducing coefficient of variant weighting to its objective function, and get a new membership and cluster centers update formula by theoretical proofs.

The rest of the paper is organized as follows. Section 2 describes introduce fuzzy c-means clustering algorithm, in Section 3 we present our proposed clustering algorithm based on BFS and coefficient of variation weighting. In Section4experiments, results on UCI database and artificial large data will be shown. Section 5 concludes the paper.

2 FCM

Fuzzy C-means accepts the fact that real-world data cannot be effectively divided into hard classes or clusters, rather, each datum may belong to more than one cluster with a non-unique degree of membership of each of the clusters [33]. Fuzzy C-means clustering algorithm can be described as minimizing the objective function:

$$\min J_m(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{m} d^2(x_i, v_j) \tag{1}$$

Where $J_m(U,V)$ is the loss function. $X = \{x_i | i = 1,2,...,n\}$ is the data set that contains n unlabeled data. $V = \{v_i | i = 1,2,...,c\}$ is a set containing cluster centers, c is the number of clusters. $d(x_i, v_j)$ is a measure of the dissimilarity between the datum x_i and the center v_j of a specific cluster j. $U = [u_{ij}]$ is the $n \times c$ fuzzy membership matrix that represents the degree to which x_i belongs to the j thcluster, m is a fuzzification parameter, $m \in (1, \infty)$.

The membership matrix satisfies the following constraints.

$$\sum_{j=1}^{c} u_{ij} = 1, u_{ij} \in [0,1], i = 1,2,...,n; \mathrm{j} = 1,2,...,\mathrm{c} . \tag{2}$$

The necessary conditions for minimizing (1) with the constraint (2) result in the following iterative update formulas for the prototypes and the partition matrix:

$$u_{ij} = \frac{1}{\sum_{k=1}^{n} \left[\frac{d_{ij}}{d_{ik}} \right]^{\frac{2}{m-1}}} . \tag{3}$$

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}} . \tag{4}$$

FCM is a simple iterative process and the iterations are carried out until the changes in the values of the partition matrix reported in consecutive iterations are lower than a certain preset threshold. The FCM algorithm is sensitive to initial values and noisy data and it is likely to fall into local optima.

3 Proposed Fuzzy C-means

An overall process of our proposed BFS-VMFCM method is depicted in Fig. 1. It begins with the BFS clustering and then determines the number of cluster centers by clustering evaluation method. Finally, the VMFCM cluster is carried out.

3.1 Breadth-first search algorithm

Breadth-first search is a graph traversal path and a hierarchical search process. It has global search capability and can access all nodes in the graph. The idea of the breadth-first search is applied in the improved clustering algorithm. The nodes x in the weighted network represent the sample data while the edges S represent the similarities, or the weights, between the adjacent nodes. Its process does not end until the similarity is smaller than a preset threshold value S or all nodes are labeled with a category

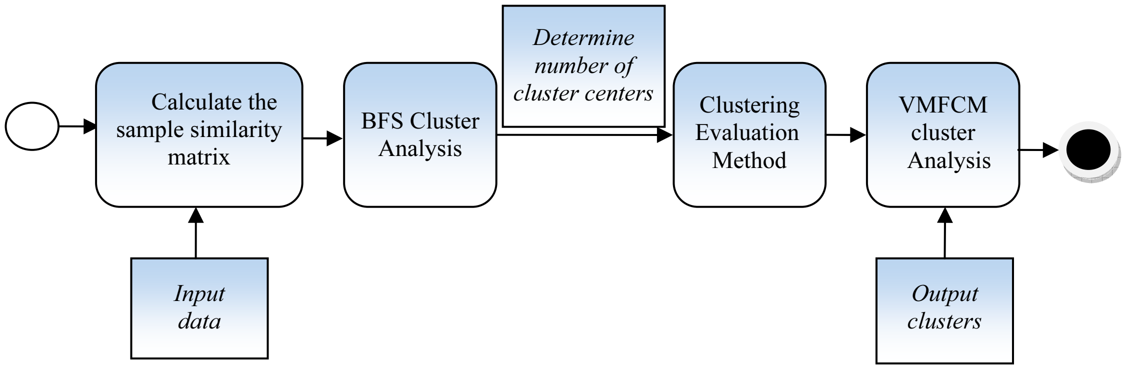


Fig. 1. The BFS-VMFCM process

group. A node is considered noise or isolated point if its similarity with all category groups is less than S so that it can't cluster. Therefore, Breadth-first search is a global search algorithm and employed to get the optimal initial clustering centers. The procedure of the BFS clustering can be described as follows.

Step 1. Calculate the weights between all the nodes connected to each other in the weighted network, that is, similarity. Denote s_{ij} as the similarity between the two data objects i and j, where $i=1,2,...,n; j=1,2,...n$, w_k as the weight of attribute factor K, d as the number of attribute factors, and x_{ik} as the corresponding attribute value corresponding to attribute factor K. Thus, s_{ij} is calculated as

$$s_{ij}=\frac{\sum_{k=1}^{d}(x_{ik}w_k)(x_{jk}w_k)}{\sqrt{\sum_{k=1}^{d}(x_{ik}w_k)^2\sum_{k=1}^{d}(x_{jk}w_k)^2}}. \tag{5}$$

Step 2. Establish an undirected weighted graph, that is, build the similarity matrix. Equation (5) shows that $s_{ij}=s_{ji}$ and $s_{ii}=1$. That is, the similarity matrix is symmetric.

$$\begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{pmatrix}. \tag{6}$$

Step 3. Identify the largest value in the similarity matrix, say S_{cr}. If $S_{cr}>S$, then x_c and x_r are classified as one category and denoted as $b_1=\{x_c,x_r\}$. Continue to search the adjacent nodes of x_c and x_r. If x_q satisfied $S_{cq}\geq S$ and $S_{rq}\geq S$, then thenode x_q is included in b_1 such as $b_1=\{x_r,x_c,x_q\}$. Similarly, the other accessible nodes are classified into the category group b_1.

Step 4. Repeat step (3). Among all points outside category group b_1 until all the sample points have been labeled, then the search ends. A node is considered as noise or an isolated point if its similarity with all category groups is less than a threshold *S* so that it cannot cluster.

Step 5. Update threshold $S_{l+1}=S_l+\lambda$, $l=1,2,...,n$, where λ is a constant representing the step size and generally takes the value 0.02. Repeat step (3) to (5) until $S_{l+1}\geq 1$. Then different dynamic clustering results can be achieved.

3.2 Clustering Evaluation Model

The breadth-first-search clustering algorithm overcomes the defects of FCM on sensitivity to initial values as well as the easy trap into local optima. However, the clustering results and algorithm's performance will depend heavily on the threshold *S*. In real problems, the data sample is unknown, and the selection of threshold *S* is also set according to the experience of the prophet. However, it is lower objectivity and can't guarantee the accuracy of the clustering results. As a result, in order to get more precise cluster centers and cluster numbers, an effective clustering evaluation model is established to select an optimal threshold *S* from the perspective of within variance and between variance.

Let n_i be the sample size of the class w_i. $x_k^{(\mathrm{i})}(k=1,2...n_i, i=1,2,...,\mathrm{C}_s)$ are the sample in class w_i, C_s is the class number of classes with respect to a threshold *S*. $x^{(i)}$ is considered as the center of class w_i and *x* is the center of the entire sample.

The class center formula is:

$$x^{(\mathrm{i})} = \frac{\sum_{i=1}^{n_i} x_k^{(\mathrm{i})}}{n_i} \tag{7}$$

The formula for the center of the entire sample is:

$$x = \frac{\sum_{i=1}^{n} x_j}{n} \tag{8}$$

Within variance is calculated as:

$$Intra\_dis = \frac{1}{n_i}\sum_{i=1}^{n_i}\left\| x_j^{(\mathrm{i})} - x^{(\mathrm{i})} \right\|^2 \tag{9}$$

Between variance is calculated as:

$$Inter\_dis = \left\| x^{(\mathrm{i})} - x \right\|^2 . \tag{10}$$

The clustering validity evaluation model is:

$$F = \frac{\sum_{i=1}^{C_s} Inter\_dis \Big/ C_s - 1}{\sum_{i=1}^{C_s} Intra\_dis \Big/ n - C_s} \tag{11}$$

The initial cluster centers can be expressed as:

$$v_i = (x^{(1)}, x^{(2)} .....x^{(\mathrm{i})})\ i = 1,2,...,\mathrm{C}_s , \tag{12}$$

The value of threshold *F* is determined by within variance and between variance. The larger the value of *F* shows the larger distance between classes and closer points within a class, the better the clustering is. The optimal threshold *S* is achieved when *F* is the maximum and the corresponding clustering results are optimal. The results can be used as initial cluster centers and the number of clusters in FCM, which can avoid local minimization due to random of initial values.

3.3 Clustering Fuzzy C-means Algorithm Based on Variation Weighted

The traditional FCM algorithm minimizes the weighted sum of squares of the distances from all data points to their corresponding cluster centers. However, FCM assumes all properties have the same effects on clustering, regardless of the effects on different data properties. Yet the sample noise attributes and clustering unrelated attributes exist, leading to the clustering result not ideal. Some of the data objects in a large sample may have great influences on clustering results due to separation from other objects. However, most clustering methods, even those with feature weighting extensions, consider all samples to have an equal weight during the clustering process. As a result, the algorithm is sensitive to noise. The data objects in a large sample may belong to different attributes. Some of the attribute objects may have a large impact on clustering results and are thus called data with strong reparability. On the contrary, others may have low contributions to clustering results and are thus named isolated data or noise data. This paper applies the coefficient of variation to reduce the contributions of noise attributes.

In the proposed objective function, the similarity between objects is determined by introducing an additional weight of the variation coefficient method. The proposed objective function reads as:

$$\min J_m(\mathrm{U},\mathrm{V}) = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{m} d_{ij}{}^{2} w . \quad \mathbf{(13)}$$

Where $w = (w_1, w_2, ..., w_p)^T, w_k = \frac{v_k}{\sum_{k=1}^{p} v_k}$. w_k is the weight of each attribute factor and $v_x = \frac{S_x}{|\overline{X}|}$ is the coefficient of variation, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean and $S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$ is standard deviation, J is the loss function, n is the number of data; C is the number of clusters and m is a fuzzy parameter.

The Lagrange multiplier method can be used to find the solution. The proposed FCM along the Lagrange function is

$$J(\mathrm{U},\mathrm{V},\lambda) = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{m} d_{ij}^{m} w - \sum_{i=1}^{n} \lambda_i (\sum_{k=1}^{c} u_{ik} - 1) . \quad \mathbf{(14)}$$

Where, $\begin{matrix} \frac{\partial J(\mathrm{U},\mathrm{V},\lambda)}{\partial u_{ik}} = \frac{\partial J(\mathrm{U},\mathrm{V})}{\partial u_{ik}} - \sum_{i=1}^{n} \lambda_i \frac{\partial}{\partial u_{ik}} (\sum_{k=1}^{c} u_{ik} - 1) \\ \frac{\partial J}{\partial \lambda} = 0 \end{matrix}$.

Expanding Equation (14) yields,

$$\begin{aligned} J(\mathrm{U},\mathrm{V},\lambda) = {} & (\mathrm{u}_{11}^{m}\mathrm{d}_{11}^{2} w + \mathrm{u}_{21}^{m}\mathrm{d}_{21}^{2} w + \cdots + \mathrm{u}_{n1}^{m}\mathrm{d}_{n1}^{2}\mathrm{w}) + (\mathrm{u}_{12}^{m}\mathrm{d}_{12}^{2} w + \mathrm{u}_{22}^{m}\mathrm{d}_{22}^{2} w + \cdots + \mathrm{u}_{n2}^{m}\mathrm{d}_{n2}^{2}\mathrm{w}) + \cdots \\ & + (\mathrm{u}_{1c}^{m}\mathrm{d}_{1c}^{2} w + \mathrm{u}_{2c}^{m}\mathrm{d}_{2c}^{2} w + \cdots + \mathrm{u}_{nc}^{m}\mathrm{d}_{nc}^{2}\mathrm{w}) \cdots - \begin{bmatrix} (\lambda_1 u_{11} + \lambda_2 u_{21} + \cdots + \lambda_n u_{n1}) \\ +(\lambda_1 u_{12} + \lambda_2 u_{22} + \cdots + \lambda_n u_{n2}) \\ + \cdots\cdots + (\lambda_1 u_{1c} \quad \lambda_2 u_{2c} \quad \lambda_n u_{nc}) \\ -(\lambda_1 + \lambda_2 + \cdots + \lambda_n) \end{bmatrix} . \end{aligned}$$

In the initial stage of optimization, the following is obtained.

$$\frac{\partial J}{\partial u_{11}} = m u_{11}^{m-1} d_{11}^{2} w - \lambda_1 = 0$$

$$\Rightarrow u_{11} = (\frac{\lambda_1}{m d_{11}^{2} w})^{\frac{1}{m-1}} .$$

Similarly, u_{12} is optimized, as follows.

$$\Rightarrow u_{12} = (\frac{\lambda_1}{m d_{12}^{2} w})^{\frac{1}{m-1}} \cdots\cdots u_{1c} = (\frac{\lambda_1}{m d_{1c}^{2} w})^{\frac{1}{m-1}} .$$

Applying the condition, $\sum_{k=1}^{c} u_{ik} = 1$,

We get,

$$\left(\frac{\lambda_1}{m}\right)^{\frac{1}{m-1}}\left[\sum_{k=1}^{c}\left(\frac{1}{d_{1k}^2 w}\right)^{\frac{1}{m-1}}\right]=1 . \quad \textbf{(15)}$$

From Equation (12), $u_{11}=\frac{\left(\frac{1}{d_{11}^2 w}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{c}\left(\frac{1}{d_{1k}^2 w}\right)^{\frac{1}{m-1}}},\cdots\cdots,u_{1c}=\frac{\left(\frac{1}{d_{1c}^2 w}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{c}\left(\frac{1}{d_{1k}^2 w}\right)^{\frac{1}{m-1}}}$,

Similarly, $u_{2c}=\frac{\left(\frac{1}{d_{2c}^2 w}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{c}\left(\frac{1}{d_{2k}^2 w}\right)^{\frac{1}{m-1}}},\cdots,u_{nc}=\frac{\left(\frac{1}{d_{nc}^2 w}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{c}\left(\frac{1}{d_{nk}^2 w}\right)^{\frac{1}{m-1}}}$.

From the above, the general form of the center updating equation is,

$$u_{ij}=\frac{\left(\frac{1}{d_{ij}^2 w}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{c}\left(\frac{1}{d_{ik}^2 w}\right)^{\frac{1}{m-1}}},\cdots\cdots d_{ij}=\left\|x_i-v_j\right\| . \quad \textbf{(16)}$$

In general, to obtain v_j, the objective function in Equation (11) is expanded to

$$J(\mathrm{U},\mathrm{V})=\sum_{i=1}^{n}u_{ij}^m w\left[\left\|x_i-v_1\right\|^2+\left\|x_i-v_2\right\|^2+\cdots\left\|x_i-v_c\right\|^2\right].$$

Optimizing the above objective function yields,

$$\frac{J(\mathrm{U},\mathrm{V})}{\partial v_1}=-2\sum_{i=1}^{n}u_{ij}w\left\|x_i-v_1\right\|=0$$

$$\Rightarrow v_1=\frac{\sum_{i=1}^{n}u_{i1}^m x_i w}{\sum_{i=1}^{n}u_{i1}^m w} .$$

Similarly, $v_2=\frac{\sum_{i=1}^{n}u_{i2}^m x_i w}{\sum_{i=1}^{n}u_{i2}^m w},\ldots,v_j=\frac{\sum_{i=1}^{n}u_{ij}^m x_i w}{\sum_{i=1}^{n}u_{ij}^m w}$.

From the above, the general form of updating center is

$$v_j=\frac{\sum_{i=1}^{n}u_{ij}^m x_i w}{\sum_{i=1}^{n}u_{ij}^m w} . \quad \textbf{(17)}$$

3.4 Clustering Steps for the Proposed BFS-VMFCM Approach

The steps of the BFS-VMFCM clustering can be described as follows.

Step 1. Input the fuzzy weight index *m*, the iteration stopping threshold epsilon ε, the classification similarity threshold *S* and the maximum number of iteration $N = 50$.
Step 2. Calculate the sample similarity matrix according to Equation 5.
Step 3. BFS clustering as shown in section 3.1. Based on threshold S and the sample similarity matrix achieved in step 3, conduct steps 3, 4, and 5 in BFS clustering and obtain different dynamic clustering results.
Step 4. Determine cluster centers and the number of clusters according to Equation 12.
Step 5. Calculate attribute weights according to Equation 13.
Step 6. Update the membership matrix *U* according to Equation 16.
Step 7. Update the cluster center *V* according to Equation 17.
Step 8. Calculate the distance between two adjacent cluster centers. If the distance of the original cluster centers differs from the updated cluster centers by less than epsilon ε, then stop; otherwise repeat step 6 through 8.

4 Experimental Results

In this section, in order to evaluate the performance of the proposed FCM algorithm, we have conducted extensive experiments on IRIS dataset and Wholesale Customer dataset from UCI database [34], X_{12} with noise data and artificial large data.

4.1 Test on Convergence Speed and Accuracy

4.1.1 Results of Experiment Using IRIS Dataset

The convergence speed and accuracy of the BFS-VMFCM algorithm is tested on the IRIS dataset and compared with FCM and GFCM (Glowworm Swarm Optimization –Fuzzy C-means) algorithms [18]. The IRIS dataset contains 3 classes of 50 instances each, where each class refers to a type of iris plant. Each instance has 4 attributes, respectively, sepal length and width, and petal length and width. In all four algorithms, the parameters were chosen as $m = 2$, the maximum number of iterations to be 50 and $\varepsilon = 0.00001$. This paper used the number of iterations and iteration time to determine the convergence speed of the algorithm, and used the average error score, average accuracy, the error sum of squares between the algorithm cluster centers and the actual cluster center to determine the accuracy of the algorithm.

Table 1. Cluster centers achieved from three algorithms for IRIS

	v_{FCM}			v_{GFCM}			$v_{BFSVFCM}$	
5.00	5.90	6.84	5.00	5.90	6.65	5.00	5.92	6.62
3.41	2.74	3.07	3.40	2.74	3.15	3.41	2.75	3.00
1.45	4.39	5.76	1.49	4.30	5.70	1.45	4.25	5.57
0.24	1.43	2.06	0.22	1.35	2.00	0.24	1.32	2.04

Table 2. Cluster results of three algorithms for IRIS

Algorithm	Average error score			Average accuracy (%)	Iterations	Runtime (s)	Error sum of squares
	Setosa	Versicolo	Virginica				
FCM	0/50	6/50	10/50	89.33	15	9.5	0.154
GFCM	0/50	2/50	3/50	96.67	10	7.5	0.066
BFSVCFCM	0/50	2/50	2/50	97.33	10	7.0	0.004

Fig. 2 and Fig. 3 show the cluster results for IRIS dataset run 25 times by FCM and BFSVFCM algorithms. We can see that the cluster prototype has changed and Fig. 3 is much better than Fig. 2. There are two main reasons. First, traditional FCM algorithm assumes by default that all features have the same impact on the clustering results, while the proposed algorithm adds more weights to attribute that has strong adaptability. Second, the initial cluster centers are randomly selected in FCM algorithm, while obtained by a breadth-first search algorithm with a global search capability in the proposed algorithm.

Fig. 4 shows the proposed algorithm and GFCM algorithm converges in 10 iterations, while the FCM algorithm is close to convergence in 10 iterations but converges in 15 iterations. In the experiment, 25 iterations were performed for the three algorithms and the iteration time of the proposed algorithm and GFCM algorithm are7.0s and 7.5s, respectively, while the iteration time of FCM algorithm is 9.5s, indicating that the convergence speeds of the new algorithm and GFCM are faster than FCM. But iteration time of this algorithm is shorter than that of the GFCM algorithm, indicating that the performance of this new algorithm is a certain advantage over GFCM algorithm and FCM algorithm. Table 2 shows that the proposed algorithm is also better than the other two in classification accuracy. As a result, this novel algorithm has better clustering results on IRIS dataset than FCM algorithm and GFCM algorithm.

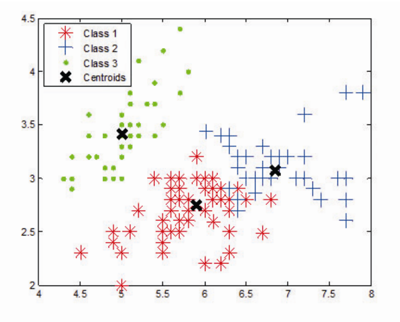


Fig. 2. Cluster result of FCM

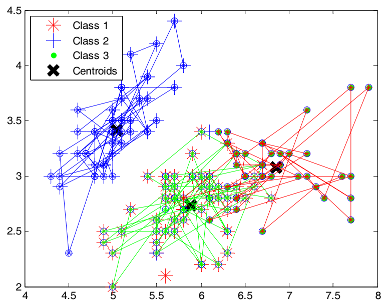


Fig. 3. Cluster result of BFS-VMFCM

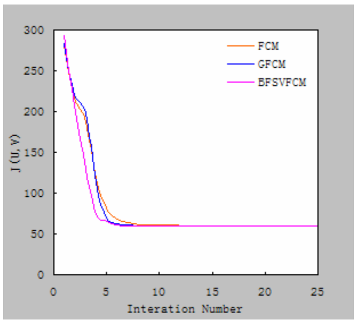
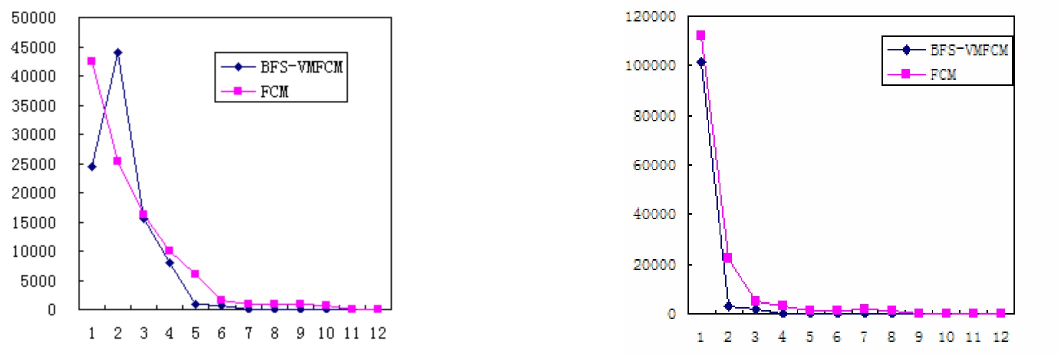


Fig. 4. Cluster iterative process of three algorithms

4.1.2 Result of Experiment Using Wholesale Customer Data Set

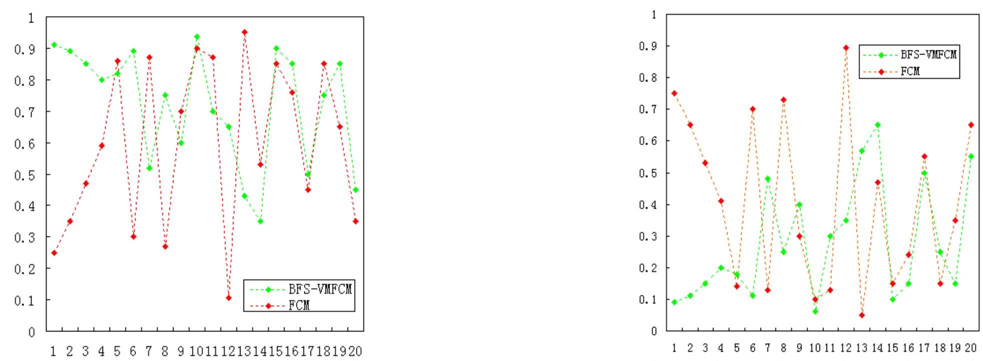
Wholesale Customer data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units on diverse product categories. Each sample contains six characteristic properties, namely fresh, milk grocery, frozen, detergents paper, and delicatessen. Wholesale customers can be classified into two categories: restaurant industry customers and retail customers. FCM algorithm and BFS-VMFCM algorithm were used. The results are as follows:

As can be seen from Fig. 5, the convergence of centers by successive iterations of the BFS-VMFCM algorithm is faster than FCM algorithm. From Fig. 6(a) to Fig. 6(b), which compared the memberships of both clusters that were obtained using two methods, the proposed methods produced cluster 1 and 2 that differ greatly and outperforms the standard FCM in this perspective. Fig. 7 and Fig. 8 show the results of the standard fuzzy C-means and the proposed BFS-VMFCM method with Wholesale Customer classification. Fig. 7 and Fig. 8 also show that the proposed BFS-VMFCM algorithm is superior to the standard FCM. As a result, the BFS-VMFCM algorithm has better convergence and clustering effect.



Successive centers of cluster 1 Successive centers of cluster 2

Fig. 5. Convergence of centers



(a) (b)

Fig. 6. Membership by successive iteration. The color red represents membership updated by FCM and green represents membership updated by BFS-VMFCM

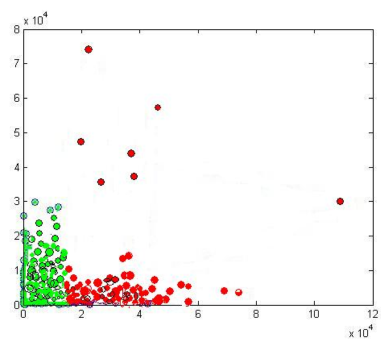


Fig. 7. Cluster result of FCM

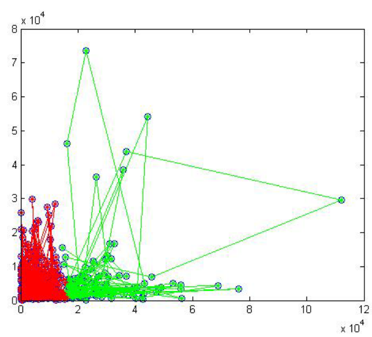


Fig. 8. Cluster result of BFS-VMFCM

4.2 Test on Noise Immunity

To test the effect of the proposed algorithm on noisy data set, the paper applied noisy data set to conduct experiments. X_{12} is a two–dimensional data set consisting of 12 data points. Among them, 10 data points are divided into two classes, while the other two data points x_6 and x_{12} are noises. Experiment parameters were set to be $\varepsilon = 0.00001$, $m = 2$, $S = 0.85$, and a maximum number of iterations was 50. The cluster centers achieved from this proposed algorithm, PCM, and PFCM, which discussed in the literature [35] are shown in the following table:

Table 3. Cluster centers of each algorithm

Actual cluster centers		v_{PCM}		v_{PFCM}		$v_{BFSVFCM}$	
-3.34	3.34	-2.17	2.17	-3.01	3.01	**-3.34**	**3.34**
0.00	0.00	0.03	0.03	0.18	0.18	**0.00**	**0.00**

As can be seen from Table 3, the results from the proposed algorithm are the same as the real cluster centers and thus the error sum of squares is zero, indicating that the BFS-VMFCM algorithm outperforms the others in achieving more accurate cluster centers on data set X_{12}.

Table 4 shows that the membership of noise point x_6 and x_{12} are 0.5 after running FCM algorithm. Yet in fact, the value of the membership of x_6 is bigger than x_{12} because x_6 is closer to the cluster centers. FCM algorithm is sensitive to the noise. The typical values of x_6 and x_{12} after running PCM algorithm are 0.62 and 0.08. x_6 is more typical than x_{12}, because of higher value. Thus, PCM algorithm reduced the effects of the noise. The typical values of x_6 and x_{12} after running BFS-VMFCM algorithm are 0.18 and 0.03, respectively. Compared with the PCM algorithm, the BFS-VMFCM algorithm has higher noise immunity and is more suitable for processing data sets with noise.

Table 4. The coordinate values of data set and membership values after running various algorithms

Number	x	y	FCM		PCM		BFSVFCM			
			u_1	u_2	t_1	t_2	u_1	u_2	t_1	t_2
1	-5.00	0.00	0.94	0.06	0.48	0.14	0.96	0.04	0.34	0.03
2	-3.34	1.67	0.97	0.03	0.67	0.20	0.97	0.03	0.62	0.04
3	-3.34	0.00	0.99	0.01	0.83	0.18	0.99	0.01	0.89	0.03
4	-3.34	-1.67	0.90	0.10	0.68	0.21	0.93	0.07	0.30	0.03
5	-1.67	0.00	0.92	0.08	0.94	0.37	0.94	0.06	0.54	0.08
6	**0.00**	**0.00**	**0.50**	**0.50**	**0.62**	**0.62**	**0.95**	**0.05**	**0.18**	**0.18**
7	1.67	0.00	0.08	0.92	0.37	0.95	0.50	0.50	0.08	0.54
8	3.34	1.67	0.03	0.97	0.17	0.68	0.07	0.93	0.04	0.62
9	3.34	0.00	0.01	0.99	0.24	0.81	0.01	0.99	0.03	0.89
10	3.34	-1.67	0.10	0.90	0.18	0.65	0.07	0.93	0.03	0.30
11	5.00	0.00	0.06	0.94	0.15	0.47	0.04	0.96	0.03	0.34
12	**0.00**	**10.0**	**0.50**	**0.50**	**0.08**	**0.08**	**0.50**	**0.50**	**0.03**	**0.03**

4.3 Test on Large Data Set

We use artificial data (ad) to test on large data set. The scale, dimensions, the number of classes, and other information of ad are shown in Table 5. In this paper, we take the value of F and time of the clustering as the evaluation index of large data clustering results.

Table 5. The characterization and cluster result of artificial data set after running three algorithms

data set	scale	d	k	FCM		PCM		BFSVFCM	
				F	(T,s)	F	(T,s)	F	(T,s)
ad2_1e+4	10k	2	6	0.8374	0.0652	0.8374	0.0652	0.8374	0.0645
ad2_1e+5	100k	2	6	0.8372	0.5004	0.8372	0.5004	0.8372	0.5004
ad2_5e+5	500k	2	6	0.8338	10.8246	0.8338	10.8246	0.8345	10.8012
ad2_1e+6	1M	2	6	0.8337	24.5862	0.8337	24.5862	0.8335	18.5356
ad2_6e+6	6M	2	6	0.8332	40.2471	0.8332	40.2471	0.8332	30.4624
ad2_1e+4	500K	20	6	0.8336	25.2365	0.8336	25.2365	0.8336	12.0045
ad20_5e+5	1M	20	6	0.7008	50.8745	0.7008	50.8745	0.8336	18.4428
ad20_2e+6	2M	20	6	0.6958	50.8924	0.7174	50.8924	0.8236	20.8524
ad40_5e+5	500K	40	6	0.6948	48.5124	0.7125	48.5124	0.8228	12.6237
ad40_1e+6	1M	40	6	0.6945	52.2548	0.6925	52.2548	0.8225	28.5652

The advantage of the novel algorithm is evident compared to FCM and PCM on large data sets. It can be seen from Table 4 and Table 5, for data sets above 500k with 20 dimensions, this algorithm is much faster than FCM and PCM algorithms. When increased to 40 dimensions, its advantage is more obvious. The higher the value of F is the better clustering. F of this algorithm is significantly higher than the other two algorithms for large data. Therefore, the proposed algorithm is also able to get better clustering results in dealing with large data sets.

5 Conclusion

The traditional fuzzy C-means clustering algorithm is a very popular clustering algorithm with a wide variety of real-world applications, but it is sensitive to initialization resulting in local minimization and noise points. In order to solve above problems, in this paper, a novel fuzzy C-means clustering algorithm based on breadth-first search and coefficient of variation weighting (BFS-VMFCM) has been proposed. The breadth-first search algorithm with global search ability is exploited to determine the appropriate initial clustering centers, and then solve the local minimization problem. In addition, this paper introduces coefficient of variation weighting to construct a new objective function for reducing noise contributions. To test the effect of BFS-VMFCM, the clustering results are compared with FCM and GFCM on IRIS and Wholesale Customer datasets from UCI database, with noise data and artificial large data. It is found that the proposed BFS-VMFCM algorithm can avoid local minimization and reduce the impact of noise points on clustering results. In addition, it outperforms other clustering algorithms in faster convergence, more accurate clustering results and higher resistance to noises. In future, the proposed improved BFS-VMFCM can be reformed into hybridization of the current algorithm with different optimization methods, for better performance.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (No. 71774044 & No. 71672050 & No. 71272191) and Philosophy and Social Sciences Innovation Team Building Project in Heilongjiang Province universities (16GLB01).

References

[1] X.B. Gao, W.X. Xie, Advance in theory and application of fuzzy clustering, Chinese Science Bulletin 45(11)(2004) 961-970.

[2] M. Zhang, J. Yu, Fuzzy partitional clustering algorithms, Journal of Software 15(6)(2004) 858-868.

[3] J.C. Dunn, A Fuzzy relative of the ISODATA process and it is use in detecting compact well-separated cluster, Journal of cybernetics 3(1974) 32-57.

[4] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, vols. 1171, Plenum Press, New York, 1981.

[5] J.M. Keller, R. Krishnapuram, A Possibilistic Approach to Clustering, IEEE Trans. Fuzzy Systems 1(2)(1993) 98-110.

[6] N.R. Pal, J.C. Bezdek, A Possibilistic fuzzy c-means clustering algorithm, IEEE Trans. Fuzzy Systems3(4)(2005) 517-530.

[7] RP. Li, M. Mukaidon, A maximum entropy approach to fuzzy clustering, in: Proc. the 4th IEEE Intl Conf. On fuzzy System, 1995.

[8] M.G. Gong Y. Liang, J. Shi, Fuzzy c-means clustering with local information and kernel metric for image segmentation, IEEE Transactions on Image Processing 22(2)(2013) 573-584.

[9] S. Krinidis, V. Chatzis, A robust fuzzy local information c-means clustering algorithm, IEEE Transactions on Image Processing 19(5)(2010) 1328-1337.

[10] Y. Dong, Y. Zhuang, K. Chen, X. Tai, A hierarchical clustering algorithm based on fuzzy graph connectedness, Fuzzy Sets Systems 157(2006) 1760-1774.

[11] W.L. Hung, M.S. Yang, D.H. Chen, Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation, Pattern Recogntion 27(2006) 424-438.

[12] M. Kuhne, R. Togneri, S. Nordholm, A novel fuzzy clustering algorithm using observation weighting and context information for reverberant blind speech separation, Signal Process 90(2010) 653-669.

[13] P.L. Lin, P.W. Huang, C.H. Kuo, Y.H. Lai, A size-insensitive integrity-based fuzzy c-means method for data clustering, Pattern Recogntion 47(2014) 2042-2056.

[14] L. Zhu, F.L. Chung, S. Wang, Generalized fuzzy C-means clustering algorithm with improved fuzzy partitions, IEEE Transaction on Systems Man &Cybernetics 39(3)(2009) 578-591.

[15] I. Ozkan, I.B. Turksen, Entropy assessment for type-2 fuzziness, in: Proc. the 13th International Conference on Fuzzy Systems, 2004.

[16] C. Hwang, F.C.H. Rhee, Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means, IEEE Transactions on fuzzy systems 15(1)(2007) 107-120.

[17] I. Hesam, A. Ajith, Fuzzy C-means and fuzzy swarm for fuzzy clustering problem, Expert Systems with Application 38(3)(2011) 1835-1838.

[18] D.S. Luo, X.W. Li, X.Q. Zhao, Research on fuzzy clustering algorithm based on GSO, Industrial Instrumentation & Automation 2(2013) 3-6.

[19] T.A. Runkler, Ant colony optimization of clustering models, International Journal of Intelligent Systems 20(12)(2005) 1233-1261.

[20] M.S. Yang, On a class of fuzzy classification maximum likelihood procedures, Fuzzy Sets and Systems 57(1993) 365-375.

[21] A. Siraj, R.B. Vaughn, Multi-level alert clustering for intrusion detection sensor data, in: Proc. IEEE Fuzzy Information Processing Society, 2005.

[22] H. Izakian, A. Abraham, Fuzzy C-means and fuzzy swarm for fuzzy clustering problem, Expert Systems with Applications 38(2011) 1835-1838.

[23] W. Pedrycz, P. Rai, Collaborative clustering with the use of Fuzzy C-Means and its quantification, Fuzzy Sets and Systems 159(2008) 2399-2427.

[24] M.R. Ferreira, F.A.T. deCarvalho, Kernel fuzzy c-means with automatic variable weighting, Fuzzy Sets Systems 237(2013) 1-46.

[25] B.A. Pimentel, R.M.C.R. Souza, A weighted multivariate fuzzy c-means method in interval-valued scientific production data, Expert Systems Applications 41(7)(2014) 3223-3236.

[26] A. Gosain, S. Dahiya, Performance analysis of various fuzzy clustering algorithms: a review, Procedia Computer Science 79(2016)100-111.

[27] B.A. Pimental, Multivariate fuzzy C-means algorithms with weighting, Neurocomputing 174(2016) 946-965.

[28] A. Irpino, R. Verde, F.A.T. deCarvalho, Fuzzy clustering of distributional data with automatic weighting of variable components, Information Science 406(2017) 248-268.

[29] J. Gou, F. Hou, W.Y. Chen, C. Wang, W. Luo, Improving Wang-Mendel method performance in fuzzy rules generation using the fuzzy C-means clustering algorithm, Neurocomputing 151(2014) 1293-1304.

[30] Z. Jiang, T. Li, W. Min, Z. Qi, Y. Rao, Fuzzy c-means clustering based on weights and gene expression programming, Pattern Recognition Letters 90(2017) 1-7.

[31] S. Askari, N. Montazerin, M.H. Fazel Zarandi, E. Hakimi, Generalized entropy based possibilistic fuzzy C-means for

clustering noisy data and its convergence proof, Neurocomputing 219(2017) 186-202.

[32] K. Siminski, Fuzzy weighted C-ordered means clustering algorithm, Fuzzy Sets and Systems 318(2017) 1-33.

[33] S. Ramathilaga, J. Leu, Two novel fuzzy clustering methods for solving data clustering problems, Journal of Intelligent & Fuzzy Systems 26(2014) 705-719.

[34] Center for Machine Learning and Intelligent Systems, UCI benchmark repository: a huge collection of artificial and real-word data sets. <http://www.ics.uci.edu/~mlearn>, 2017.

[35] W.C. Huang, J.L. Liu, L.Y. Xiong, A sample-feature weighted possibilistic fuzzy kernel clustering Algorithm, Computer Engineering & Science 36(1)(2014) 169-175.