

Similarity Technique Effectiveness of Optimized Fuzzy C-means Clustering Based on Fuzzy Support Vector Machine for Noisy Data

Hoda Khanali *, Babak Vaziri

Department of Industrial Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran

Abstract *Fuzzy VIKOR C-means (FVCM)* is a kind of unsupervised fuzzy clustering algorithm that improves the accuracy and computational speed of *Fuzzy C-means (FCM)*. So it reduces the sensitivity to noisy and outlier data, and enhances performance and quality of clusters. Since FVCM allocates some data to a specific cluster based on similarity technique, reducing the effect of noisy data increases the quality of the clusters. This paper presents a new approach to the accurate location of noisy data to the clusters overcoming the constraints of noisy points through *fuzzy support vector machine (FSVM)*, called FVCM-FSVM, so that at each stage samples with a high degree of membership are selected for training in the classification of FSVM. Then, the labels of the remaining samples are predicted so the process continues until the convergence of the FVCM-FSVM. The results of the numerical experiments showed the proposed approach has better performance than FVCM. Of course, it greatly achieves high accuracy.

Keywords Fuzzy partition clustering, Noisy data, Fuzzy support vector machine, Accuracy.

AMS 2010 subject classifications 62H30, 91C20, 90C70

DOI: 10.19139/soic-2310-5070-1035

1. Introduction

A data set divides through partition clustering algorithms into non-overlapping subsets so that each data belongs to a subset. In a fuzzy clustering, a point belongs to each cluster with a weight between zero and one. Fuzzy clustering has important applications in image processing, pattern recognition, object recognition, and so on [1, 2, 3]_other similar techniques as fuzzy c-means [4], *possibilistic fuzzy c-means (PFCM)* [5], *credibility fuzzy c-means (CFCM)* [6].

FVCM [7] is an improved FCM algorithm to solve the problem of the sensitivity to noisy data in FCM. FVCM has a good performance in detecting noisy data. FVCM improves the accuracy and computational speed. The sensitivity to noisy and outlier data is one of FCM problems that a robust clustering approach called TCLUST [8], *fuzzy c-means-relaxed constraints support vector machine (FCM-RSVM)* [9], *Relative entropy fuzzy c-means (REFCM)* [10], and algorithms based on type-2 fuzzy sets such as [11, 12] have recently been proposed to solve this problem. In addition, a comparison of partition algorithms is presented by Khanali and Vaziri in [13]. Sefidian and Daneshpour in [14] by combining grey system theory concepts and FCM to provide the clustering accuracy called *grey based fuzzy c-means and mutual information (GFCMI)* based feature selection imputation. For each missing record it determines a set of similar records, and estimates a missing value by using the regression imputation. Bu in [15] uses the tensor canonical polyadic decomposition in FCM to compress the attributes of each object and the amount of the data which run in *Internet of Things (IoT)* systems. Further, it called the high-order tensor fuzzy c-means (HOFCM) and enhances the clustering efficiency. Ramos et al. in [16] provide a diagnostic algorithm

*Correspondence to: Hoda Khanali (Email: hod.khanali.eng@iauctb.ac.ir). Department of Industrial Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran, Tel.: +9844600220.

for the problem of the novel faults such that density oriented fuzzy c-means decreases the confusion by removing the outlier data, and kernel fuzzy c-means classifies the data to reduce the classification errors. Heidari et al. in [17] offer a joint formulation by combining fuzzy c-means clustering and distance metric learning which reduces sensitivity to initializing, improves non-linear optimization problem and increases accuracy.

Moreover, the FSVM [18] classification algorithm is a developed species of *Support Vector Machine* (SVM) [19] so that each of the training samples has different degrees of importance using the fuzzy concepts. This issue is considered in the learning process. Of course, FSVM has a stronger theory than neural networks. It also achieves overall optimality due to the quadratic structure. In addition, FSVM is resistant to over-fitting and benefits from legible and geometric interpretations, and it uses a little space in storing the prediction model. Using the fuzzy membership function, FSVM can reduce the problem of classifying noisy and outlier data and increase the accuracy of allocation of border data. Due to various applications of SVM, recently an *entropy based on fuzzy least squares support vector machine* (EFLSSVM-CIL) and *entropy based on fuzzy least squares twin support vector machine* (EFLSTWSVM-CIL) are proposed by Gupta and Richhariya in [20] to decrease class imbalance impact in binary class data sets. The class imbalance is a common problem because the class that is very important in scope of application (the minority class) contains fewer samples than the class that has not particular importance (the majority class). Yan and Wang in [21] introduced matching decision method based on the improved SVM to optimize classifying small, uncertain input, and unbalanced samples using triangular fuzzy theory. Further, Hamidzadeh and Moslemnejad in [22] have proposed *belief function and fuzzy rough set-boundary samples* (BFFR-BS) to train procedure identifying the data points at the boundary of the classes and detecting noisy data points. Zhou et al. in [23] have introduced a new membership function to overcome the dependence of FSVM so that reduces the sensitivity to noisy and outlier data. To solve the problem of class imbalance, Samma et al. in [24] have used *particle swarm optimization* (PSO) and FSVM. In other case, for developing the linear and nonlinear FSVM, Tang in [25] has defined a new fuzzy membership function where the accuracy is improved and the negative impact of the outliers is reduced. FSVM with a hybrid kernel function and *genetic algorithm* (GA) is used by Zhou et al. in [26] to achieve better learning ability and performance called *control chart patterns recognition* (CCPR) method. Since the FVCM allocates some points to a particular cluster based on similarity technique, sometimes they may be attributed with low membership values. On the other hand, FSVM has more tolerant to noisy points and a short-term response by using a fuzzy membership functions. This is due to the fact that the classification method in choosing the decision boundary attempts to maximize the minimum distance to each of the classes, and how to select a boundary based on points called support vectors. Therefore, samples with high membership rates are identified for training in the classification of FSVM, and the labels of the remaining samples are predicted using the FSVM classification. The allocation of the noisy points is done with the correct replacement of them, and this process continues until the convergence of FVCM-FSVM.

The rest of this paper is arranged as follows: in section 2, the general field of data mining and the main concepts of FSVM and FVCM are described. Section 3 presents the structure of the proposed algorithm. The experimental results are drawn in Section 4. Finally, conclusion and future work are presented in Section 5.

2. Background

Data mining helps process a large amount of data and discover the knowledge contained in data sources. Model learning methods of data exploration are categorized based on two groups, predictive (supervised) methods and descriptive (unsupervised) methods. Figure 1 shows the model learning methods.

The predictive methods describe the values of some attributes to predict the value of a specified attribute. Classification, regression, and anomalies detection are among the methods of learning a model with the nature of prediction. The different types of classification algorithms are neural networks, support vector machines, decision trees, and Bayesian networks.

The descriptive methods explain the relationships between the data regardless of the label or external variable. Clustering [13], association rules [27], and sequential patterns are among the methods of learning a model with the nature of descriptive. The clustering method is a model learning method that has a descriptive nature. In all

clustering algorithms, the main goal is to minimize cluster density and maximize the separation of clusters from each other. Various methods of clustering are divided into hierarchical clustering, density-based clustering, grid-based clustering, incremental clustering, and partitional clustering.

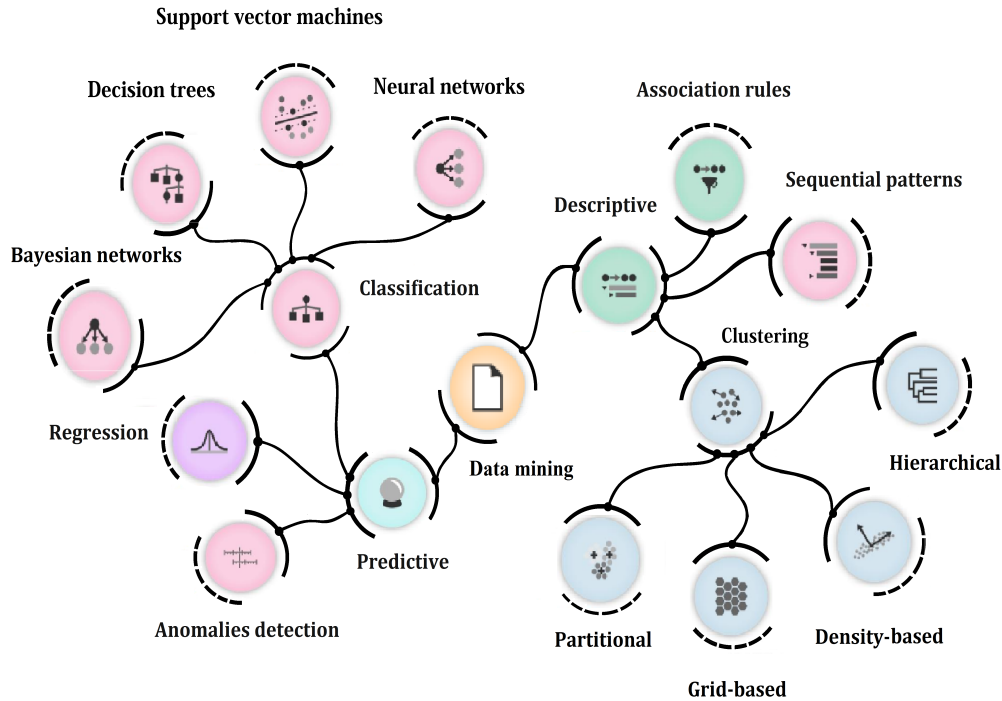


Figure 1. Model learning methods in data mining.

2.1. FSVM

FSVM improves the accuracy of the classification. There is a training sample set of labeled, $(x_1, y_1, s_1), \dots, (x_n, y_n, s_n)$, where $\{x_i \in R^p\}_{i=1}^n$, p is the dimension of the training samples, with labels $y_i \in \{-1, 1\}_{i=1}^n$ and s_i is the fuzzy membership $0 < s_i \leq 1$. The *optimal separating hyperplane* (OSH) problem solves the following problem:

$$Minimize \quad \frac{1}{2}(W^T.W) + C \sum_{i=1}^n s_i \xi_i \tag{1}$$

$$Subject \ to \quad y_i(W^T.\varphi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, i = 1, 2, \dots, n \tag{2}$$

In the above equations, W is a normal vector of the separating hyper plane and b is a bias in the pair (W, b) . C is a constant to control the misclassification errors, a smaller s_i for the corresponding point x_i reduces the effect of ξ_i which is an error measure.

The Lagrangian multiplier using a kernel function is represented as:

$$Maximize \quad \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{3}$$

$$\text{Subject to } \sum_{i=1}^n y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq s_i C, \quad i = 1, 2, \dots, n \quad (4)$$

Here $K(x_i, x_j)$ is the kernel function in $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, and $\alpha(k)$ is the Lagrange multiplier. y is the decision function to the predicted class label for x [18].

$$y = \text{sign}\left(\sum_{i \in S} \alpha_i y_i K(x_i, x) + b\right) \quad (5)$$

2.2. FVCM

The principal idea of the fuzzy clustering algorithm is related to the decision-making concept of extended VIKOR [28] so that various alternatives are evaluated with criteria to achieve the best solutions.

In designing this algorithm to evaluate clusters, not only the similarity of the clusters is considered through internal validation measures; but also the algorithm uses cluster quality through external validation measures. According to Algorithm 1, Dunn's index, Davies-Bouldin index, entropy, and density and means functions are considered as the decision-making criteria. The set of alternatives ranked with the extended VIKOR are actually improved points. This algorithm then continues until convergence is achieved.

Algorithm 1: FVCM algorithm [7].

- Step 1:** Initialize k (number of clusters), Initialize m (fuzziness parameter)
 - Step 2:** While $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ do.
 - Step 3:** Determine a set of alternatives: Replacing membership degree as alternative of each sample.
 - Step 4:** Update a set of criteria (Dunn's index, the means, the density, Davies-Bouldin index and the entropy)
 - Step 5:** Rating the clusters and calculating the vectors of the centers $C^k = [c_j]$ with $U^{(k)}$ using extended VIKOR.
 - Step 6:** Update $U^{(k)}, U^{(k+1)}$.
 - Step 7:** End while.
-

3. Proposed FVCM-FSVM algorithm

In the proposed method, the FVCM function is improved by combining it with the classification of constraints of fuzzy support vector machines. Therefore, this method is called FVCM-FSVM. First, The FVCM has a structure of partition clustering so that there is no special class in this structure so the data points are unlabelled samples as inputs, and the data are only clustered based on the similarity criterion. Then, FVCM clustering done that the cluster centre and the fuzzy membership values are determined in the data clustering. Figure 2 shows the general process.

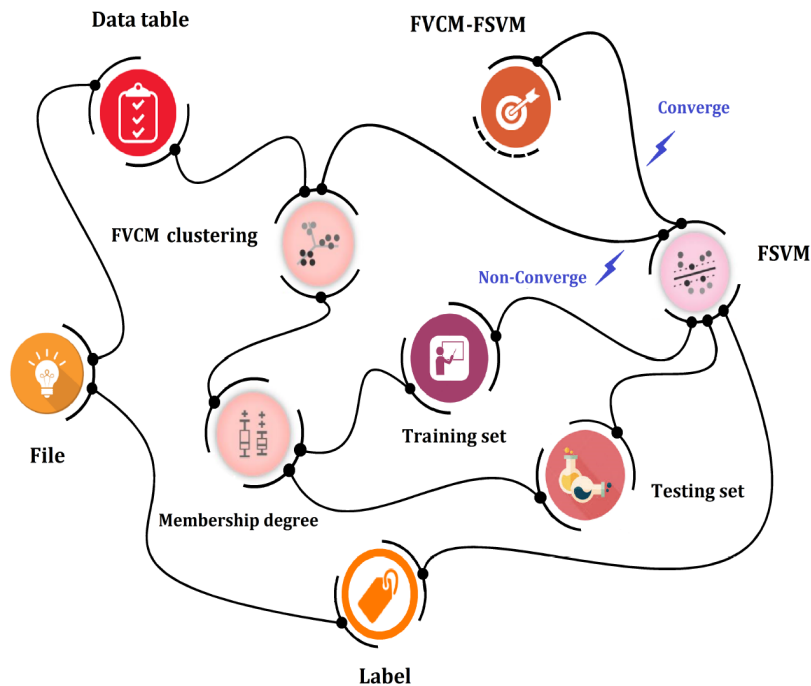


Figure 2. The general trend of FVCM-FSVM algorithm

In the next step, each sample has a specific label which is used as input to FSVM. The purpose of the learning algorithm which consists of two stages of training and testing is to find the order of the types of labels based on other features of the samples. In this way, the data set is divided two parts of the training and test data set. In the training phase, the membership of each data point is determined based on the membership values of each data point to the clusters, and a part of the data point of each cluster is considered as a training set for learning and model building. These samples should naturally have high membership rates. In the following, another part of the data is used to estimate the accuracy of the classification model at the test stage. The noisy points gradually achieve the minimum membership values while leading to class balance. Ultimately, this process recursively continues until convergence conditions are provided. At each stage, the quality of the clusters is improved by the evaluation criteria of FVCM, and such an interaction between the two algorithms is the reason for the increased accuracy. This procedure is shown in Algorithm 2.

According to the above, the similarity technique of FVCM-FSVM based on FSVM increases the quality of the clustering. Figure 3 shows the behaviour of the proposed method to clarify the issue. There is a synthetic data set with two clusters. Cluster 1 is red (training data) and purple (classified data) and cluster 2 is green (training data) and phosphorous (classified data). In this type of clustering, the points with higher membership are trained and used to separate the noisy points. In this figure, the effect of FVCM-FSVM is shown to improve the quality of the clusters in the presence of noisy points.

Step Algorithm 2: FVCM-RSVM algorithm

-
- Step 1:** Initialize k (number of clusters), Initialize m (fuzziness parameter)
- Step 2:** While $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ do.
- Step 3:** Determine a set of alternatives: Replacing membership degree as alternative of each sample.
- Step 4:** Update a set of criteria (Dunn's index, the means, the density, Davies-Bouldin index and the entropy)
- Step 5:** Rating the clusters and calculating the vectors of the centers $C^k = [c_j]$ with $U^{(k)}$ using extended VIKOR.
- Step 6:** Update $U^{(k)}, U^{(k+1)}$.
- Step 7:** Assign the 90% of the most membership values in the training set and the remaining 10% of samples in the test set.
- Step 8:** Insert the class labels
- Step 9:** Train one-against-all FSVM
- Step 10:** End while
-

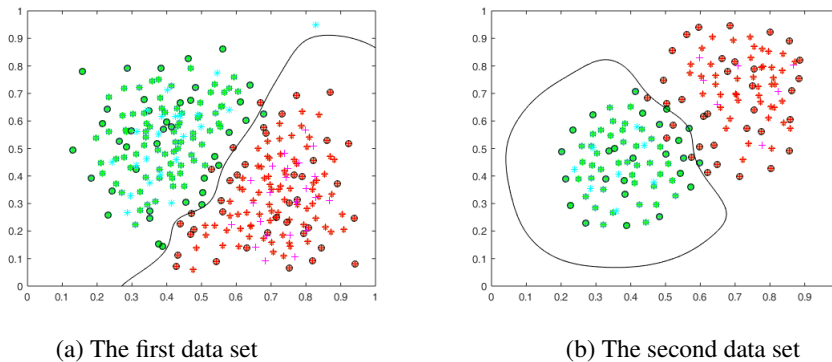


Figure 3. The results of FVCM-FSVM in a synthetic data set with two clusters. (a) and (b) Red and green points (training data), purple and phosphorous points (classified data), hollow black circles (noise points), RBF kernel.

4. Experiment

FVCM and FVCM-FSVM are evaluated in three weighted modes. In this research, all experiments were run *MATLAB R2015b* software, and executed on a computer with an *Intel CoreTM i5* processor (1.80GHz), and 4G RAM(3.88GB). These experiments are repeated 20 times for each of the three UCI data sets [29]. They are as follows:

1 Iris data set contains 150 instances described by 4 features and classified into three classes.

2 Glass Identification data set contains 214 instances described by 10 features and classified into seven classes.

3 Haberman's Survival data set contains 306 instances described by 3 features and classified into two classes.

These data sets are randomly divided into two parts which are the training section (90% of data) and the test section (10% of data). In this case study, the multi-class classification method is used one-against-all and RBF kernel strategies.

4.1. Experiment 1

the quality of the clusters of the two algorithms (FVCM and FVCM-FSVM) is evaluated based on six criteria; such as Dunn's index, Davies-Bouldin index, the entropy, data processing time (seconds), the number of iterations, and accuracy (percent).

Dunn's index identifies the sets of clusters that are compact and well separated. The main goal of this measure is to maximize intercluster distances whilst minimizing intracluster distances [30]. The index is defined in [31].

$$D_{nc} = \min_{i=1, \dots, nc} \left\{ \min_{j=i+1, \dots, nc} \frac{d(c_i, c_j)}{\max_{k=1, \dots, nc} \text{diam}(C_k)} \right\} \quad (6)$$

Where $d(c_i, c_j)$ is the dissimilarity function between two clusters c_i and c_j and also $\text{diam}(c)$ is the diameter of a cluster [31]. As such large values of D_{nc} correspond to good clusters [30]. The implications of the Dunn's are the considerable amount of time required for its computation, and the sensitive to the presence of noise in data sets [31].

Davies-Bouldin index, As the Dunn's index, aims at identifying sets of clusters that are compact and well separated [30]. A similarity measure R_{ij} between the clusters c_i and c_j is dened based on a measure of dispersion of a cluster c_i and a dissimilarity measure between two clusters d_{ij} . The R index is dened as follows [31]:

$$R_{ij} = (s_i + s_j) / d_{ij} \quad (7)$$

Where in nc is the number of clusters, s_i is the average distance between the cluster data to cluster center, and d_{ij} is the distance between the centers of the clusters. Then the Davies-Bouldin index is dened as [31]:

$$DB_{nc} = \frac{1}{nc} \sum_{i=1}^{nc} \max_{j=1, \dots, nc, i \neq j} (R_{ij}), \quad i = 1, \dots, nc \quad (8)$$

DB_{nc} is the average similarity between each cluster C_i , ($i = 1, \dots, nc$) and its most similar one [31]. Therefore, the cluster configuration that minimizes DB is taken as the optimal number of clusters [30].

Entropy determines cluster quality. The lower entropy means better clustering. Besides, the quantity of disorder is found by using the entropy [32]. To consider a data set containing the classes, S is clustered into k clusters. Let n_k be the number of data points in k th cluster and n_{sk} be the number of data points from the s th class in the k th cluster. The entropy of the k th cluster is given by [9]:

$$Ent_k = - \sum_{s=1}^S \frac{n_{sk}}{n_k} \log\left(\frac{n_{sk}}{n_k}\right) \quad (9)$$

The total entropy for the set of k clusters is dened as [9]:

$$Entropy = \sum_{k=1}^k \frac{n_k}{n} Ent_k \quad (10)$$

Where n is the number of data points in the data set.

The accuracy is the certain percentage of the data which are correctly clustered by an examined method.

$$Accuracy = \frac{\text{The number of correctly classified samples}}{\text{The total number of samples}} \quad (11)$$

To consider have the correct assessment, the conditions for all experiments are similar to [7], and to compare the results of FVCM in [7] with experiment results of FVCM-FSVM are presented in tables. With regards to

[33], internal validation measures, namely Dunn's index, Davies–Bouldin index, and so on, achieve better results. In addition, Dunn's index and Davies–Bouldin index have more effective results in recognition of the noisy and outlier data than other evaluation criteria [34] so higher weights are assigned to these values. As stated above, extended VIKOR alternatives allocate weights. Therefore, three weight modes, including minimum (#1), average (#2), and maximum (#3), are considered which are shown in the Table 1. In each row of states (#1, #2, #3), the sum of the values is equal to one, but the values less than 0.1 are ineffective.

Table 1. The weight of alternatives in extended VIKOR.

Weighted states	Density function	Davies-Bouldin index	Dunn's index	Entropy	Average function
#1	0.3	0.1	0.1	0.3	0.2
#2	0.2	0.2	0.2	0.2	0.2
#3	0.1	0.35	0.35	0.1	0.1

4.1.1. *Iris data set* In this subsection, the efficiency of the algorithms is described in Iris data set. FVCM and FVCM-FSVM are evaluated based on three weight modes and three clusters ($k=3$), and the results are shown in Table 2 where the best results are bold.

Table 2. Clustering results of FVCM and FVCM-FSVM with Iris data ($k=3$).

FVCM states	Davies–Bouldin index	Dunn's index	Entropy	Time (seconds)	The number of iterations	Accuracy (percent)
FVCM#1	0.407394	0.65298	0.12691	0.909073	5.9	98.6905
FVCM#2	0.410200	0.64967	0.11950	0.966171	4.3	97.9336
FVCM#3	0.409152	0.65010	0.13754	0.988342	4.7	98.7566
FVCM-FSVM#1	0.380763	0.72518	0.12017	1.270013	6.6	98.804
FVCM-FSVM#2	0.375902	0.73919	0.13716	1.591237	6.1	98.913
FVCM-FSVM#3	0.389978	0.74105	0.13560	1.355884	6.9	98.895

Best results in the ranking of clustering algorithms are shown in bold.

Given the values of Table 2, the use of evaluation criteria in both methods can affect the status of membership values of the elements relative to the status of the membership values of the neighbouring points. On the other hand, targeted displacements in extended VIKOR method make the FVCM and FVCM-FSVM optimal with more accuracy. Because of the sustainability of the results in various evaluations, convergence conditions are provided to resist noisy data. Moreover the fuzzy feature and error correction possible of misallocation of samples in both methods has caused the least number of iterations in clustering. In total all, due to FSVM three modes of FVCM-FSVM have more run time than all three modes of FVCM.

Since the results of FVCM and FVCM-FSVM are based on the evaluation of DB and Dunn indexes, the intra-cluster distance with the inter-cluster distance has been compared. In evaluating the quality of the clusters, the FVCM-FSVM has obtained the inter-cluster distance and the inter-cluster diameter better than other. The smaller DB index values and the larger Dunn index values indicate FVCM-FSVM clusters denser so that it has led to a better separation of clusters.

4.1.2. *Glass data set* In this subsection, the efficiency of the algorithms is described in Glass data set. FVCM and FVCM-FSVM are evaluated based on three weight modes and seven clusters ($k=7$), and the results are shown in Table 3 where the best results are bold.

Given the values of Table 3, In FVCM and FVCM-FSVM, evaluation criteria are defined as the measure of similarity and proximity of data, and extended VIKOR is also considered to fit this area. Since clusters of Glass

Table 3. Clustering results of FVCM and FVCM-FSVM with Glass data (k=7).

FVCM states	Davies–Bouldin index	Dunn’s index	Entropy	Time (seconds)	The number of iterations	Accuracy (percent)
FVCM#1	1.39615	0.890213	0.280534	0.919017	13.4	98.152
FVCM#2	1.45508	0.874113	0.239460	0.919709	8.9	98.478
FVCM#3	1.41748	0.931684	0.301115	0.934185	11.1	97.915
FVCM-FSVM#1	1.07628	0.683563	0.116975	1.164328	13.6	98.573
FVCM-FSVM#2	1.02901	0.682637	0.192851	1.080629	15.3	98.515
FVCM-FSVM#3	1.06923	0.701928	0.216640	0.951103	10.7	98.224

Best results in the ranking of clustering algorithms are shown in bold.

data set have various densities, in the evaluation of the comparison between the clusters, the evaluations show that the FVCM-FSVM in DB index provides denser clusters. This means a good separation of clusters from each other. However Glass data set is the standard set of data that contains clusters of different sizes, the best entropy value belongs to FVCM-FSVM#3 because the method of data distribution and the weight of the alternatives are the effective factors of such a result. This property has greatly contributed to the accuracy of the algorithm.

4.1.3. Haberman’s survival data set In this subsection, the efficiency of the algorithms is described in Haberman’s survival data set. FVCM and FVCM-FSVM are evaluated based on three weight modes and two clusters (k=2), and the results are shown in Table 4 where the best results are bold.

Table 4. Clustering results of FVCM and FVCM-FSVM with Haberman’s survival data (k=2).

FVCM states	Davies–Bouldin index	Dunn’s index	Entropy	Time (seconds)	The number of iterations	Accuracy (percent)
FVCM#1	0.235186	0.68810	0.63172	0.29175	8.1	98.058
FVCM#2	0.229104	0.71978	0.91617	0.52141	10.2	98.163
FVCM#3	0.238535	0.77701	0.69318	0.29910	13.8	98.391
FVCM-FSVM#1	0.2498293	0.71309	0.62791	1.21990	9.8	98.751
FVCM-FSVM#2	0.2432701	0.74725	0.62989	1.24015	8.5	98.196
FVCM-FSVM#3	0.2544935	0.75783	0.62011	1.12518	9.2	98.455

Best results in the ranking of clustering algorithms are shown in bold.

Given the values of Table 4, by comparing the intra-cluster distance with the inter-cluster distance, FVCM produces dense clusters that are well-separated so the best DB index belongs to this algorithm. This is the ability of this algorithm to accurately locate the cluster centres and thus lead to good clustering efficiency. The diameter of the FVCM-FSVM clusters can be very influential if there are noisy data, this parameter can be effective for detecting cluster density. And as a result, by implementing this algorithm with Haberman’s survival data set which has non-spherical clusters, Dunn’s index is somewhat more desirable than other case. Thus, FSVM is the main cause of the resistance of the algorithm to noisy data. The homogeneous elements of each cluster are measured as 0.62011. Since it is likely that some noisy points in the FVCM-FSVM are assigned to a specific cluster with low membership values, the use of FSVM constraints has been effective in improving fuzzy clustering performance. As a result, the higher accuracy of the algorithm is 98.751%.

4.2. Experiment 2

According to the results of Experiment_1, The weight modes of the algorithms that achieved the best results about the accuracy (in Table 2, Table 3, and Table 4) are considered so that Figure 4, Figure 5, and Figure 6 illustrate

the behaviour of the objective functions than the number of iterations for both algorithms. FVCM-RSVM shows more flexibility in fuzzy inequalities. That is, the distribution of samples is considered even in the early iterations. On the other hand, FVCM-RSVM has better clustering accuracy which indicates the proper positioning of noisy data. Using extended VIKOR evaluation criteria as alternatives leads to optimal clustering with FVCM and FVCM-FSVM in the first iterations so that the first iterations of the objective function decrease to the convergence level, and they do not have the problem of local optimality.

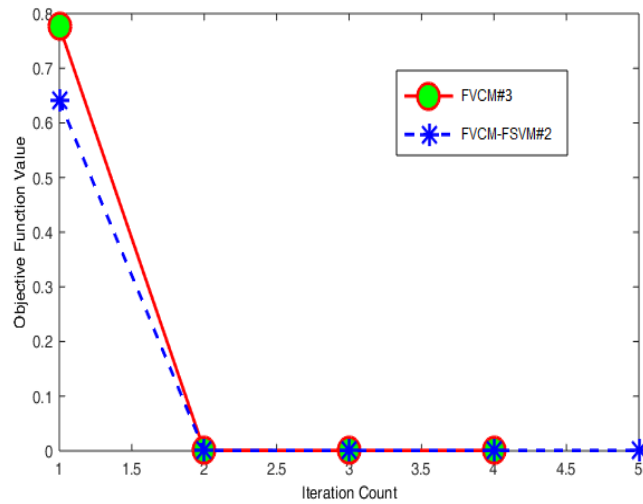


Figure 4. Convergence characteristic of the different clustering algorithms with Iris data

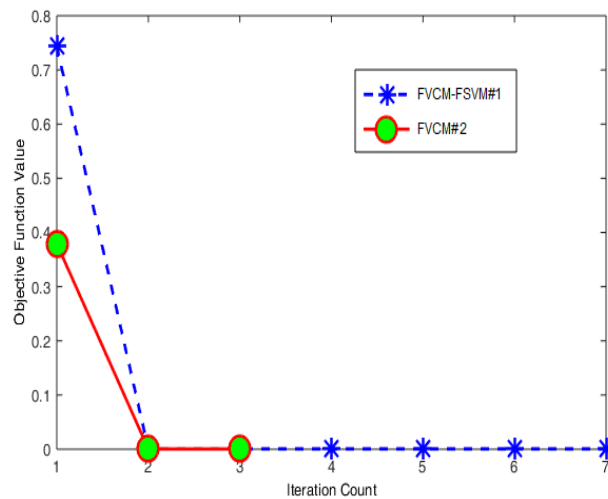


Figure 5. Convergence characteristic of the different clustering algorithms with Glass data

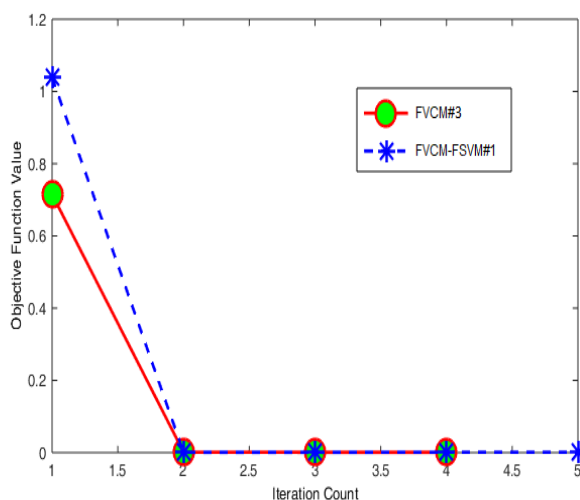


Figure 6. Convergence characteristic of the different clustering algorithms with Haberman's survival data

5. Conclusion

In this paper, FVCM-FSVM based FSVM develops FVCM to overcome the destructive effects of noisy data, and increase accuracy because FSVM is one of the classification methods which it has a good generalized. Since FVCM selects high-degree membership samples for training in FSVM, and it preserves the labels of the remaining samples using the FSVM classifier, noisy data are correctly allocated by accurately replacing them. Of course, FVCM and FVCM-FSVM consist of the same structure with different weights. Since FVCM-FSVM involves the learning phase due to the use of the FSVM method, it leads to spend more time for clustering than FVCM, but the algorithm has higher clustering accuracy than FVCM.

The proposed algorithm using FSVM has a good quality in data clustering and the proper location of cluster centres than FVCM. In addition, the good results of accuracy emphasize the better implementation of this trend. So it reduces the sensitivity to noisy data. Moreover FVCM-FSVM solves the local optimality problem, and provides the reasonable convergence. Future studies can focus on optimizing the points of each cluster through the parallelization process.

References

1. L. Panigrahi, K. Verma, and B. Kumar Singh, *Ultrasound Image Segmentation Using A Novel Multi-Scale Gaussian Kernel Fuzzy Clustering and Multi-Scale Vector Field Convolution*, Expert Systems With Applications, vol. 115, pp. 486-498, 2019.
2. E. Sert, and D. Avci, *Brain tumor segmentation using neutrosophic expert maximum fuzzy-sure entropy and other approaches*, Biomedical Signal Processing and Control, vol. 47, pp. 276-287, 2019.
3. F. Han, and R. Ellis, *Identifying Consistent Patterns of Quality Learning Discussions in Blended Learning*, The Internet and Higher Education, vol. 40, pp. 12-19, 2019.
4. J. Dunn, *A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters*, Journal of Cybernetics, vol. 3, no. 3, pp. 32-57, 1974.
5. N. Pal, K. Pal, J. Keller, and J. Bezdek, *A possibilistic fuzzy c-means clustering algorithm*, IEEE Transactions on Fuzzy Systems, vol. 13, p. 517-530, 2005.
6. K. Chintalapudi, and M. Kam, *A noise resistant fuzzy c-means algorithm for clustering*, IEEE Conference on Fuzzy Systems Proceedings, vol. 2, p. 1458-1463, 1998.
7. H. Khanali, and B. Vaziri, *An improved approach to fuzzy clustering based on FCM algorithm and extended VIKOR method*, Springer Transactions on Neural Computing and Applications, vol. 30, no. 179, p. 1-12, 2019.
8. H. Fritz, and L. Garcia-Escudero, *Robust constrained fuzzy clustering*, Elsevier Trans. Information Sciences, vol. 245, pp. 38-52, 2013.

9. M. Sabzekar, and M. Naghibzadeh, *Fuzzy c-means improvement using relaxed constraints support vector machines*, Elsevier Trans. Applied Soft Computing, vol. 13, no. 2, pp. 881-890, 2013.
10. M. Zarinbal, M. Fazel Zarandi, and I. Turksen, *Relative entropy fuzzy c-means clustering*, Elsevier Trans. Information Sciences, vol. 260, pp. 74-97, 2014.
11. M. Rostam Niakan Kalhori and M. Fazel Zarandi, *Interval type-2 credibilistic clustering for pattern recognition*, Elsevier Transactions on Pattern Recognition, vol. 48, no. 11, pp. 3652-3672, 2015.
12. S. Malek Mohamadi Golsed and M. Fazel Zarandi, *Multi-central general type-2 fuzzy clustering approach for pattern recognitions*, Elsevier Transactions on Information Sciences, vol. 328, p. 172-188, 2016.
13. H. Khanali and B. Vaziri, *A Survey on Clustering Algorithms for Partitioning Method*, International Journal of Computer Applications, vol. 155, no. 4, pp. 20-25, 2016.
14. A. Sefidian and N. Daneshpour, *Missing value imputation using a novel grey based fuzzy c-means, mutual information based feature selection, and regression model*, Expert Systems With Applications, vol. 115, pp. 68-94, 2018.
15. F. Bu, *An efficient fuzzy c-means approach based on canonical polyadic decomposition for clustering big data in IoT*, Future Generation Computer Systems, vol. 88, pp. 675-682, 2018.
16. A. Ramos, A. da Silva Neto and O. Llanes-Santiago, *An approach to fault diagnosis with online detection of novel faults using fuzzy clustering tools*, Expert Systems with Applications, vol. 113, pp. 200-212, 2018.
17. N. Heidari, Z. Moslehi, A. Mirzaei and M. Safayani, *Bayesian Distance Metric Learning for Discriminative Fuzzy C-Means Clustering* Neurocomputing, vol. 319, pp. 21-33, 2018.
18. C.-F. Lin and S.-D. Wang, *Fuzzy support vector machines*, IEEE Transactions on neural networks, vol. 13, no. 2, p. 464-471, 2002.
19. V. Vapnik, *The nature of statistical learning theory*, Springer, New York, 1995.
20. D. Gupta and B. Richhariya, *Entropy based fuzzy least squares twin support vector machine for class imbalance learning*, Applied Intelligence, vol. 48, no. 11, p. 4212-4231, 2018.
21. H.S. Yan and Y.F. Wang, *Matching decision method for knowledgeable manufacturing system and its production environment*, Journal of Intelligent Manufacturing, vol. 30, no. 2, p. 771-782, 2019.
22. J. Hamidzadeh and S. Moslemnejad, *Identification of uncertainty and decision boundary for SVM classification training using belief function*, Applied Intelligence, pp. 1-16, 2018.
23. H. Zhou, H. Qin and H. Shen, *An Improved Fuzzy Support Vector Machine Algorithm Based on Refactoring Class Center*, International Conference on Applications and Techniques in Cyber Security and Intelligence ATCI, pp. 91-99, 2018.
24. H. Samma, C. Lim and U. Ngah, *A Hybrid PSO-FSVM Model and Its Application to Imbalanced Classification of Mammograms*, Intelligent Information and Database Systems, vol. 7802, pp. 275-284, 2013.
25. W. Tang, *Fuzzy SVM with a New Fuzzy Membership Function to Solve the Two-Class Problems*, Neural Processing Letters, vol. 34, no. 3, pp. 209-219, 2011.
26. X. Zhou, P. Jiang and X. Wang, *Recognition of control chart patterns using fuzzy SVM with a hybrid kernel function*, Journal of Intelligent Manufacturing, vol. 29, no. 1, p. 51-67, 2018.
27. H. Khanali and B. Vaziri, *A survey on improved algorithms for mining association rules*, International Journal of Computer Applications, vol. 165, no. 9, pp. 6-11, 2017.
28. M. Tavana, R. Mavi, F. Santos-Arteaga and E. Dous, *An Extended VIKOR Method Using Stochastic Data and Subjective Judgments*, Computers and Industrial Engineering, pp. 1-29, 2016.
29. A. Asuncion and D. Newman, *UCI Machine Learning Repository*. University of California, School of Information and Computer Science, Irvine, <http://www.ics.uci.edu/mllearn/MLRepository.html>, accessed June, 2020.
30. A. Macedo-Cruz and I. Villegas-Romero, *Unsupervised Classification of Aerial Images Based on the Otsu's Method*, INTECH Open Access Publisher, 2012.
31. M. Halkidi and Y. Batistakis, *On Clustering Validation Techniques*, Journal of intelligent information systems, vol. 17, no. 2-3, pp. 107-145, 2001.
32. S. Sripada, *Comparison of purity and entropy of k-means clustering and fuzzy c means clustering*, Indian Journal of Computer Science and Engineering (IJCSE), vol. 2, no. 3, pp. 343-346, 2011.
33. E. Rendon, I. Abundez and A. Arizmendi, *Internal versus External cluster validation Indexes*, International Journal of Computers and Communications, vol. 5, no. 1, pp. 27-34, 2011.
34. Y. Liu, Z. Li and H. Xiong, *Understanding and Enhancement of Internal Clustering Validation Measures*, IEEE Trans. Cybernetics, vol. 43, no. 3, pp. 982-994, 2013.