Optimal structure for large-scale data clustering based on support vector machine and fuzzy rules

Amin Feli, Rezvan Khalaji, Sakineh Kadaei, Mojtaba Banifakhr

Optimal structure for large-scale data clustering based on support vector machine and fuzzy rules

Amin Feli, *Master of Computer Engineering, Arak Azad University,Iran,Arak.*
Rezvan Khalaji, *Master of Computer Engineering, Malayer Azad University, Iran, Malayer.*
Sakineh Kadaei, *Master of Computer Engineering, Bushehr Azad University, Iran, Bushehr.*
Mojtaba Banifakhr*, *Ph.D. of Electrical Engineering, Yazd University, Iran, Yazd.*

***Abstract*— Data clustering is a method for classifying similar data that used in various sciences for many years and many algorithms designed in this field. Recent clustering research has led to hybrid methods that are more robust and accurate. Combined clustering first tries to produce primary clustering that is as scattered as possible and then combines the results by applying an agreement function. In this research, a combination of fuzzy clustering and support vector machine used for classification. The hybrid network (FS-FCSVM) is an efficient fuzzy clustering operation performed on the input data. The network parameters trained with SVM, achieves a network with high generalizability. The number of rules in such systems is smaller than fuzzy systems and cussed a lower computation time.**

Data clustering is a technique for identifying related data that has been utilized for many years in a variety of fields. Numerous algorithms have been developed in this area.

Recent advances in clustering research have produced hybrid techniques that are more reliable and precise.

In order to combine the findings, combined clustering first attempts to create primary clustering that is as dispersed as feasible.

In this study, categorization was done using a hybrid of fuzzy clustering and support vector machines.

The input data are effectively fuzzy clustered using the hybrid network (FS-FCSVM).

With SVM-trained network parameters, a highly generalizable network is produced.

These systems have fewer rules than fuzzy systems and need less computing time.

In this study, the reduction clustering method used before fuzzy clustering. The main idea of reduction clustering is to search for high-density regions in the data space characteristic. Each point that has the largest number of neighbors selected as the center of the cluster. In other words, the reduction clustering technique used to select feature points that are more different and less similar to other points. In this paper, the idea is to use differential clustering to find the exact center points of clusters and the number of clusters, which reduces the number of repetitions of fuzzy clustering and use these central points as part of the training data and the second part of the work.

Prior to using fuzzy clustering in this investigation, reduction clustering was employed.

Searching for high-density areas in the data space is the main goal of reduction clustering.

The center of the cluster is chosen for each point with the most neighbors.

To choose feature points that are more distinct from one another and less similar to one another, in other words, is the goal of the reduction clustering method.

In this study, the goal is to utilize differential clustering to determine the precise cluster centers and the number of clusters, hence reducing the number of fuzzy clustering repeats. These central points will then be used in the training data and the second stage of the study.

Another part of the training data selected. To select them, we have used the belonging matrix obtained from fuzzy clustering. By determining a numerical range of data far from the center of each data, we also selected it as another part of the data. Reduce the amount of training data significantly. The results of experiments performed on the large data set of the UCI database show that in addition to reducing training time with proper data selection.

a different portion of the training data was chosen.

We utilized the belonging matrix discovered using fuzzy clustering to choose them.

We also chose it as another portion of the data by identifying a numerical range of data that was distant from the center of each data.

Significantly reduce the volume of training data.

The outcomes of tests conducted on the substantial UCI database data set demonstrate that smart data selection may also shorten training times.

***Index Terms*—Fuzzy clustering, Support vector machine, data mining, big data.**

Manuscript received June XX, 20XX; revised July XX, 20XX. (Write the date on which you submitted your paper for review and the last revision date.)

1-Amin Feli, Master of Computer Engineering, Arak Azad University,Iran,Arak,amin.feli.2010@gmail.com.

2-Rezvan Khalaji, Master of Computer Engineering, Malayer Azad University,Iran,Malayer,khalaji@gmail.com.

3-Sakineh Kadaei, Master of Computer Engineering, Bushehr Azad University,Iran, Bushehr, saki.kadaeiii1998@yahoo.com.

Mojtaba Banifakhr*, *Ph.D. of Electrical Engineering*, *Yazd University, Iran, Yazd.banifakhr.mojtaba@stu.yazd.ac.ir.*

I. INTRODUCTION

Data clustering is one of the most common data mining techniques. Clustering is one of the most widely used methods in data analysis. Clustering is an automated process in which samples divided into groups whose members are

similar, which called clusters. In other words, a cluster is a set of objects in which objects are similar to each other and are not similar to objects in other clusters.

Clustering used in many fields, including pattern recognition, machine learning, data mining, information retrieval, and bioinformatics. The purpose of clustering is to provide the end user with a good view of what is happening in the database. Another application of clustering can used to determine data that differs significantly from other data.

One of the most popular data mining approaches is data clustering. One of the most used techniques in data analysis is clustering. Clustering is an automated method that divides samples into groups with similar constituents, or clusters.

In other terms, a cluster is a collection of items that are related to one another but not to objects in other clusters.

Clustering is utilized in a wide range of disciplines, such as bioinformatics, pattern recognition, machine learning, data mining, and information retrieval. To give the end user a clear picture of what is occurring in the database, clustering is used. Finding data that varies considerably from other data is a further use of clustering.

In clustering, an attempt is made to divide a set of data without observers into clusters that maximize the similarity of the data within each cluster and minimize the similarity between the data within the different clusters [1,2].

Clustering algorithms [3,4] separate data objects (designs, entities, samples, observations, units) into a certain number of clusters (groups, subcategories or articles). In the case of clustering, Orrit (2001) states that clustering is a set of similar entities, but the entities of different clusters are not the same.

Different criteria can be considered for similarity, for example, the distance criterion can be used for clustering and objects that are closer to each other can be considered as a cluster, which is also called distance-based clustering.

In clustering, an effort is made to partition a set of unobserved data into groups that maximize similarity within each cluster while minimizing similarity between clusters [1,2]. Data objects (designs, entities, samples, observations, units) are divided into a predetermined number of clusters using clustering algorithms [3,4]. (groups, subcategories or articles). According to Orrit (2001), clustering refers to a collection of items that are related but not identical across various clusters. Different criteria for similarity may be taken into consideration. For instance, the distance criterion may be used for clustering, and objects that are closer to one another may be taken into consideration as a cluster (this is known as distance-based clustering).

In fact, clustering algorithms are often such that a series of initial representatives considered for the input samples and then from the degree of similarity of the samples with these representatives, it is determined which cluster the sample belongs to and after this step, new representatives calculated for each cluster. Against the samples compared with these representatives to determine which cluster they belong to, and this repeated until the representatives of the clusters do not change.

Fuzzy clustering can considered as a part of fuzzy data analysis, which has two parts: one is fuzzy data analysis and the other is definite data analysis using fuzzy techniques.

Fuzzy clustering explores fuzzy models of data.

In reality, clustering algorithms frequently work by first considering a number of initial representatives for the input samples, determining the cluster to which the sample belongs based on how similar the sample is to these representatives, and then calculating new representatives for each cluster. Once the representatives of the clusters have remained the same, samples were compared with these representatives to identify which cluster they belong to.

Fuzzy data analysis comprises two components: fuzzy data analysis and definite data analysis using fuzzy techniques. Fuzzy clustering is one of these components.

Fuzzy clustering investigates data with fuzzy models.

The basic idea in fuzzy clustering is to assume that each cluster is a set of elements. Then, by changing the definition of element membership in this set, from the state where an element can only be a member of a cluster to the state that each element can be included in several clusters with different membership degrees, we present the categories that are more relevant to reality.

In classical clustering, each input instance belongs to one and only one cluster and cannot be a member of two or more clusters, and in other words clusters do not overlap. Now consider a case where the similarity of a sample with two or more clusters is the same.

Assuming that each cluster consists of a set of items is the fundamental tenet of fuzzy clustering.

Then, we offer the categories that are more applicable to reality by modifying the definition of element membership in this set from the condition where an element may only be a member of a cluster to the state where each element can be included in several clusters with varying membership degrees. In classical clustering, clusters do not overlap since each input instance may only be a member of one cluster and cannot be a member of more than one cluster.

Consider the situation where a sample with two or more clusters is comparable in all respects.

In classical clustering, it must be decided which cluster this sample belongs to. The main difference between classical clustering and fuzzy clustering is that a sample can belong to more than one cluster [5]

Numerous applications of fuzzy clustering in data analysis and pattern recognition as well as existing research areas in this field, including its use in solving routing problems, allocation and scheduling, highlight the need to study existing algorithms and improve and modify them [6].

In the reference [7], FS-FCSVM is a fuzzy system that is built with if_then rules and the result of such systems is a single value. Parameters in FS_FCSVM are generated via an SVM support vector device, which has fewer rules and less computational time than fuzzy systems.

Which cluster this sample belongs to in a traditional

cluster analysis must be determined. A sample can belong to more than one cluster with fuzzy clustering, which is the major distinction between the two methods [5].

Numerous uses of fuzzy clustering in pattern recognition and data analysis, as well as active research topics in this area, such as its application to scheduling, allocation, and routing issues, highlight the need to examine current methods and enhance and alter them [6].

According to the reference [7], the FS-FCSVM is an if then rule-based fuzzy system that produces a single value. SVM support vector devices, which have fewer rules and need less processing than fuzzy systems, are used in FS FCSVM to create parameters.

Fuzzy neural network and support vector machine are two types of computational methods with high efficiency and capacity. In FNN the only training parameter is minimizing the amount of error so generalization performance may be poor. SVM is a new and powerful network whose formula for learning is based on minimizing the amount of error. The fuzzy rules here are equal to the number of support vectors (SV), which is usually very large. The test results with FS_FCSVM are much better than FNN, but the disadvantage of FS_FCSVM is that it has so many rules that this problem can be solved with fuzzy clustering.

In [7] as fuzzy clustering with support vector machine have been used to segment human skin cancer.

Support vector machines and fuzzy neural networks are two examples of computing techniques with great capacity and efficiency. The performance of generalization may be subpar in FNN as the primary training parameter is minimizing the amount of error. SVM is a brand-new, potent network whose learning algorithm is centered on reducing the amount of error. Here, the number of support vectors (SV), which is typically quite big, is equal to the fuzzy rules. The test results with FS FCSVM are significantly better than those with FNN, however the drawback of FS FCSVM is that it has so many rules that fuzzy clustering can be used to fix this issue.

In [7], support vector machine and fuzzy clustering were used to segment human skin cancer.

[8] is a method based on proposing fuzzy weights using interpolation reasoning, which provides a scattered fuzzy weighted interpolation method that allows the previous variables to appear in fuzzy laws of different weights.

In [9] the FCM fuzzy mapping method is based on fuzzy logic and aspects of the neural network it inherits. The main advantage of FCM is fuzzy network modeling and decision making and improves their performance.

FCM is used to gather recent advances in algorithmic learning. Can be represented by a numerical matrix operation. Because FCM is a set of interconnected concepts, the propagation of a response from the primary node is very similar to a set of neural network nodes.

[8] is a method that uses interpolation reasoning to propose fuzzy weights and offers a dispersed fuzzy weighted interpolation approach that enables the prior variables to show up in fuzzy laws with various weights.

Fuzzy logic and elements of the neural network it inherits form the foundation of the FCM fuzzy mapping method in [9]. Fuzzy network modeling and decision-making are the key benefits of FCM, which boosts their effectiveness.

The most current developments in algorithmic learning are gathered using FCM. is anything that a numerical matrix operation can represent. The propagation of a response from the primary node is quite similar to a set of neural network nodes since FCM is a collection of interrelated ideas.

In [9,10,11] researches on some data, clustering by fuzzy SVM method was performed by support vector machine and fuzzy neural network and FS_FCSVM and compared with each other and it was concluded that FS_FCSVM is better than others but has more rules. Therefore, if we want to compare our proposed method with other methods, the best benchmark is to use the tested data set. In [12], a data volume reduction method based on the selection of triangular data within fuzzy clustering (FCM) is proposed.

Support vector machine, fuzzy neural network, and FS FCSVM were used to conduct clustering by fuzzy SVM technique on certain data in [9,10,11] studies. These methods were evaluated, and it was determined that FS FCSVM was superior than the others despite having more rules. Therefore, the examined data set serves as the greatest standard for comparing our suggested strategy to other methodologies. The selection of triangular data inside fuzzy clustering (FCM) is the foundation of the data volume reduction technique introduced in [12].

Initially, training data is clustered using fuzzy clustering method. In each cluster, the three training data that are farthest from the center of the cluster and each other are selected. Triangular data and cluster centers are used as reduced training data for SVM training. The results of experiments performed on the large data set of the UCI database show that the proposed method, in addition to reducing the training time by selecting the appropriate data, strengthens the SVM's resistance to noise and noise data and also reduces the number of selected support vectors by SVM in the database. grows.

In this article, after presenting the introduction, the basics of the proposed method are evaluated and then the proposed method is presented.

In the third part, the simulation is reviewed and finally the results are presented. The final section also presents the final results.

The fuzzy clustering approach is initially used to cluster the training data. The three training data that are separated from each other and the cluster's center by the greatest distance are chosen for each cluster. Reduced training data for SVM training are cluster centers and triangular data. The results of experiments conducted on the extensive dataset of the UCI database demonstrate that the proposed method not only shortens training time by choosing the right data, but also improves the SVM's resistance to noise and noisy data, as well as lowering the number of support vectors selected

by the SVM in the database. increases.
In this article, the introduction is followed by an analysis of the suggested technique's fundamentals before the proposed method itself is described.

The simulation is examined and the findings are then reported in the third section.
The final results are also shown in the part after that.

II. BASICS

One of the first fuzzy clustering methods based on the objective function and the use of Euclidean distance proposed by Don in 1974 and then generalized by Bozdak. The algorithm detects spherical clouds of points in a dimensional P space. These clusters assumed to approximately the same size. Each cluster represented by its center. This way of representing clusters called a model or sample, because it often thought of as representing all the data assigned to the cluster. In selecting the center of the cluster, the average value is used.

One of the earliest approaches to fuzzy clustering, developed by Bozdak after Don introduced it in 1974 using the objective function and Euclidean distance. In dimensions P space, the technique finds spherical clouds of points.

These clusters were thought to be about the same size.

By its center, each cluster is represented. Because it is frequently believed to reflect all the data attributed to the cluster, this method of describing clusters is known as a model or sample.

The average value is used to determine the cluster's center.

To calculate the center of the cluster, the sum of the membership degrees of each element divided by the power m in itself multiplied by the power m of the degrees of membership. The problem with this algorithm is that the algorithm cannot identify clusters of different shapes, sizes, and densities. Instead of the same matrix, other matrices, such as the diagonal matrix, can used to identify elliptical clusters to identify other shapes. One of the advantages of this algorithm is its ease, which leads to a reduction in computational time. In practice, with a few repetitions, an almost final solution can reached. Yang then conducted an overview of fuzzy clustering methods.

The total of the membership degrees for each element is multiplied by the power m of the degrees of membership to determine the cluster's center. The issue with this approach is that it is unable to distinguish between clusters of various sizes, densities, and forms. Other matrices, such as the diagonal matrix, can be used in place of the same matrix to detect elliptical clusters and other forms.

This algorithm's simplicity, which reduces calculation time, is one of its benefits. In practice, after a few iterations, a nearly perfect answer can be found.

Yang then performed a summary of fuzzy clustering techniques.

By replacing the Euclidean distance with another meter (created by a symmetric and definite matrix), elliptic clusters can identified.

In this method, each cluster in addition to the center of the cluster is determined by a symmetric, definite and positive matrix. This matrix creates a soft for each cluster. It should also be borne in mind that by arbitrarily selecting matrices, the distances can arbitrarily reduced. To avoid minimizing the objective function with matrices with almost zero inputs, we need a constant value for clusters with matrices with determinants one. So now only the shape of the clusters is changing, not their size.

Elliptic clusters can be located by substituting another meter (produced by a symmetric and definite matrix) for the Euclidean distance. This approach uses a symmetric, definite, and positive matrix to identify each cluster as well as its center. Each cluster is given a soft by this matrix.

Additionally, keep in mind that the distances may be arbitrarily decreased by choosing matrices.

We require a constant value for clusters with matrices with determinants one in order to prevent minimization of the objective function with matrices with virtually zero inputs. Therefore, at this point, just the clusters' form, not their size, is altering.

Gustafson and Kessel also made possible different shapes for clusters by introducing a constant value of e for each matrix A, and in general det (A) = e. Although the choice of constants also requires prior knowledge of clusters has it. The quality of the result obtained from this method is highly dependent on the available data. It also remains a problem with GK. If the data clustered with the feasibility approach, development factors for the clusters can used to detect defects in the images. Position and direction can obtained from the center of the cluster and the matrix. Compared to FCM, GK Effective in reducing repetition steps and increasing convergence speed is the start of the GK algorithm with the results of one FCM.

Gustafson and Kessel introduced a constant value of e for each matrix A, and in general det (A) = e, making it feasible for clusters to take on diverse forms. Although previous knowledge of clusters is also necessary for the selection of constants. The quality of the results produced by this approach is strongly influenced by the data at hand.

GK is still having issues with it.

When using the feasible strategy to cluster the data, development factors for the clusters may be utilized to identify image flaws.

The cluster's and the matrix's center may be used to determine position and direction.

In contrast to FCM, GK

The start of the GK algorithm using the output of one

FCM is efficient in decreasing repeat steps and enhancing convergence speed.

The possibility of fuzzy cloning at the first combined by Kim and Krishnapuram in 1993. This method based on the weight criterion η, in which the modification of the fit function and the addition of input weights are used to reduce the effect of remote data on the centers of the clusters.

the potential for fuzzy cloning at the first 1993 combination by Kim and Krishnapuram. This approach, which is based on the weight criteria, uses the addition of input weights and the modification of the fit function to lessen the impact of remote data on the cluster centers.

$$\eta_j = \frac{\sum_{j=1}^{n} \mu_{ij}^m d_{ij}^2}{\sum_{j=1}^{n} \mu_{ij}^m} \quad (1)$$

In relation (1), μ is the membership function of each data sample, d is the Euclidean distance, n is the number of data samples, m is a real number greater than one. The fit function of this algorithm is considered as Equation (1-2).

In relation (1), n is the number of data samples, d is the Euclidean distance, and m is a real integer higher than one. is the membership function of each data sample. Equation is used to represent this algorithm's fit function (1-2).

$$J(u,v) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^m d_{ij}^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1-\mu_{ij})^m \quad (2)$$

$$\mu_{ij} = \frac{1}{1+(\frac{d_{ij}^2}{\eta_i})^{1/(m-1)}} \quad (3)$$

In the above relation, c is the number of final clusters, η is the weight of each cluster according to Equation (1-2), J is the fitness function (target).

In the above relation, c is the number of final clusters, η is the weight of each cluster according to Equation (1-2), J is the fitness function (target).

In Algorithm C, the average probability, taking into account η weight and using the relation (1-3) belonging (membership function) for each data, minimizes the effect of remote data to the centers of the clusters and the sum of the components of each data is a number between 0 and 1, if in algorithm C the fuzzy mean of the sum of the components of each data is one.

In Algorithm C, the average probability minimizes the impact of distant data on the centers of clusters while using the relation (1-3) belonging (membership function) for each data and the sum of each data's components is a number between 0 and 1, if in Algorithm C the fuzzy mean of the sum of each data's components is one.

$$v_k = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m (x_j - v_i)^2 + \frac{1}{c} \sum_{i=1}^{c} (v_i - \bar{v})^2}{\min_{i \neq j} (v_i - v_j)^2} \quad (4)$$

Using Equation (3), the membership function is calculated for each data and Equation (4) of the cluster centers. The fit function of the C-mean algorithm gives a higher probability value than the fuzzy mean C algorithm. But in algorithm C, the average number of possible steps to find the final cluster increases. If in this algorithm, the initial random cluster is not considered and the centers of the clusters, the final clusters of the fuzzy mean algorithm are considered, the greater the value of the fit function of the mean algorithm C is clearly seen.

Since then, many modifications and improvements have been made to the proposed algorithms.

The membership function for each piece of data is obtained using Equation (3), and the cluster centers are calculated using Equation (4). In comparison to the fuzzy mean C method, the fit function of the C-mean algorithm provides a greater probability value. The average number of steps needed to discover the final cluster rises in method C, though. If the end clusters of the fuzzy mean algorithm are taken into account rather than the initial random cluster in this algorithm, the fit function of the mean algorithm C obviously has a higher value.

Since then, the suggested algorithms have undergone several changes and enhancements.

According to the classification made in this paper, the basic fuzzy clustering algorithms have been limited to Fuzzy C-Means and Possibilistic C-Means. Introduced as the K-Means algorithm have been extracted [1].

The Possibilistic C-Means algorithm characterizes the relative degree of probability of membership.

The relative nature of the potential membership, despite being appropriate in most cases, can sometimes cause problems [3]. As an example of this category of problems can be mentioned as follows. Point X_1 is a distance from both clusters, so its degree of membership to both clusters is 5. This is logical, but the problem occurs when the same degree of belonging to X_2 is also given while the distance between these two points of the clusters is not the same.

The fundamental fuzzy clustering methods have been restricted to Fuzzy C-Means and Possibilistic C-Means in accordance with the categorization proposed in this study. K-Means method was first introduced and extracted [1].

The relative level of membership probability is described by the probabilistic C-Means method.

Although it is usually suitable, the relative nature of the possible membership can occasionally lead to issues [3]. The following is a list of difficulties that fall under this category. Point X_1's degree of membership to both clusters is 5, as it is separated from both clusters. Although this makes sense, there is an issue when the same degree of membership to X_2 is also supplied and the distance between these two cluster locations is not the same.

The reason for this is normalization and the need for the

sum of the membership points of a point in different clusters to be equal. Normalizing membership rates can lead to adverse effects on out-of-center data presentation. If we release the normalization condition in the FCM algorithm, these adverse effects will be less. This approach is called possibility. The objective function, which previously minimized only squares of distances, does not seem to agree with the possibility approach. By removing the normalization condition by obtaining zero membership degrees for all points in each cluster, at least the objective function is obtained, and of course it is not reasonable that all clusters are empty. Therefore, a fine should be considered to keep membership rates below zero. [13]

This is due to normalization and the requirement that the total of a point's membership points in several clusters be equal. Normalizing membership rates may have a negative impact on the way out-of-center data is presented. These negative impacts will be reduced if the normalization criterion in the FCM method is removed. This strategy is known as possibility. The possibility method does not appear to be supported by the goal function, which previously exclusively reduced squares of distances. At least the goal function is attained by eliminating the normalization constraint by achieving 0 membership degrees for all points in each cluster, even though it is obviously illogical for all clusters to be empty. Thus, it should be thought about imposing a fine to maintain membership rates below zero. [13]

The k-means algorithm, proposed by McQueen, is a basic method for many other clustering methods (such as fuzzy clustering) despite its simplicity. This method is usually the fastest clustering method for large data sets. Different forms have been expressed for this algorithm. But they all have iterative routines that try to estimate the following for a fixed number of clusters:

Obtaining points as centers of clusters, these points are actually the average points belonging to each cluster.

Attribute each data sample to a cluster that has the shortest distance to the center of that cluster.

Despite its simplicity, the McQueen-proposed k-means algorithm forms the foundation for many different clustering techniques (including fuzzy clustering). For big data sets, this technique often yields the quickest clustering results. This algorithm has been expressed in a variety of formats. But given a fixed number of clusters, they all feature iterative processes that attempt to estimate the following:

When points are obtained as cluster centers, they are really the average points for each cluster.

Each data sample should be assigned to the cluster with the least distance to the cluster's center.

The K algorithm is an average as unsupervised learning in which the number of clusters is not predetermined and the clusters do not have a common chapter with each other. Different initial values for the mean K algorithm can lead to different clustering. Because this algorithm is based on the Euclidean distance, it can converge to the local minimum. This is usually true for clusters that do not separate very well. It is shown that there is no guarantee for the convergence of an iterative algorithm to a global optimization [14].

In unsupervised learning, the K method uses an average when the number of clusters is not preset and the clusters do not have a common chapter. The mean K technique might provide different clusters based on various beginning values. This approach can reach the local minimum since it is based on the Euclidean distance. This is typically true for clusters that are difficult to disentangle. It is demonstrated that there is no assurance that an iterative method would converge to a global optimization [14].

High convergence speed is one of the most important advantages of this algorithm, but there is no specific procedure for the initial calculation of cluster centers and if in the iteration of the algorithm, the number of data belonging to clusters becomes zero, there is no way to change and improve the continuation of the method.

One of the most significant benefits of this algorithm is its high convergence speed, but there is no specific method for calculating the cluster centers at the beginning of the algorithm, and if the number of data in a cluster falls to zero during an algorithm iteration, there is no way to alter or improve the method moving forward.

A. Differential clustering

This algorithm is a quick way to find the number of clusters as well as their centers when there is no clear view of the number of clusters to be specified for the data set.

The centers estimated by this method are sometimes used as starting points for other clustering algorithms. This technique is used to extract key points or distinct samples from a set of data set records, each record containing the characteristics of a key point.

When it is unclear how many clusters should be defined for the data set, this approach provides a rapid way to determine both the number of clusters and their centers.

Other clustering methods occasionally employ the centers computed by this approach as their starting locations. This method is used to extract important information or distinct samples from a collection of data set records, where each record includes information that defines a key point.

Differential clustering is essentially a modified form of the Mountain method. In the algorithm, each point considered as a potential for the cluster center. The potential measurement is obtained according to Equation (5). This method can be used both as an independent method for clustering and can be used as a precondition for other clustering algorithms. (In this case we can decide in the process)

The steps of the algorithm are as follows:

Consider a set of n data {x_1, x_2,…, x_n} in the next M space. Each point in this set, as a candidate for the cluster centers, the density measure at the x_i data points is calculated as follows:

The Mountain technique has largely been changed by differential clustering. Each location is taken into account as a potential cluster center by the algorithm. Equation is used

to acquire the potential measurement (5). Both independently and as a prerequisite for other clustering techniques, this technique can be utilized for clustering. (In this instance, we have a decision-making process.)

The algorithm's stages are as follows:

Consider a collection of n data in the following M space: x_1, x_2,..., x_n. The density measure at the x_i data points is determined for each point in this collection as a potential candidate for the cluster centers as follows:

$$D_i = \sum_{j=1}^{n} exp(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2}) \quad (5)$$

r_a is a positive fixed number that denotes the neighborhood radius. Hence a point in the data will have a high density value if it has a large number of points in the neighborhood.

The first center of the cluster (x_c1) is selected as the point with the highest density (D_c1). Then, the density value of each point x_i is re-evaluated as follows:

The radius of the immediate area is indicated by the positive fixed integer r_a. A point in the data will thus have a high density value if there are many other points in the area. The location with the highest density (D_c1) is chosen to be the cluster's first center (x_c1).

The density value is then recalculated as follows for each location x_i:

$$D_i = D_i - D_{c1}\, exp(-\frac{\|x_i - x_{c1}\|^2}{(r_b/2)^2}) \quad (6)$$

After recalculating the density for each data point, the next x_c2 center is selected and all calculations for the data point density are corrected again. This process continues until a sufficient number of points in the centers of the clusters are produced. Decreasing clustering output is a Sugno fuzzy inference system.

The next x_c2 center is chosen, and all corrections are made to the data point density computations once each data point's density has been recalculated. This procedure is repeated until enough points are created in the cluster centers. A Sugno fuzzy inference system is the decreasing clustering output.

B. *Support machine vector*

One of the most popular methods that is currently widely used for the classification problem is the support vector machine (SVM) method. The first algorithm for classification various patterns proposed by Fisher in 1936, and its criterion for optimization was to reduce the classification error of educational patterns. Many of the algorithms and methods proposed so far, for designing pattern classifiers follow the same strategy.

In 1965, a Russian researcher named Vladimir Vapnilk took a very important step in designing classifiers and solidified the statistical theory of learning and based on it used support vector machines.

The support vector machine (SVM) approach is one of the most well-liked approaches that is now commonly employed for the classification challenge.

Fisher introduced the first classification method for different patterns in 1936, and one of its optimization criteria was to lower the classification error of instructional patterns.

The same approach is used by many of the algorithms and techniques for developing pattern classifiers that have been developed so far.

Vladimir Vapnilk, a Russian researcher, made a crucial advancement in classifier design in 1965 by establishing the statistical theory of learning and developing support vector machines on its basis.

1) *Support vector machine in integral mode*

For the inseparable state of a set of ∈_i variables called definition deficiency variables

We do so that the following condition is met:

1. Integral mode support vector machine

For a set of I variables known as definition deficit variables that are in an irreducible condition.

The following criterion is ensured by what we do:

$$y_i(w_i x + b) \geq 1 - \epsilon_i \quad (7)$$

It is clear that the higher summation of ∈_i, cussed the greater the error, so we define the problem of constrained optimization as follows:

Since it is obvious that the error increases with the summing of I we formulate the restricted optimization problem as follows:

$$\begin{Bmatrix} minimize & \frac{1}{2} w.w + c\sum_{i=1}^{N} \epsilon_i \\ subject\ to & y_i(w_i x + b) \geq 1 - \epsilon_i \end{Bmatrix} \quad (8)$$

For this problem, we form the KKT conditions at the answer point and reach the following two:

At the answer point, we create the KKT conditions for this issue and arrive at the following two:

$$\begin{Bmatrix} Minimize & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^{N} \alpha_j \\ Subject\ to & 0 \leq \alpha_i \leq C \end{Bmatrix} \quad (9)$$

As can be seen, solving the SVM problem in the integral state is similar to solving it in the separable state. After obtaining the Lagrangian coefficients, the patterns whose Lagrangian coefficients apply in the following relation are supported vectors:

The SVM issue can be solved in the integral state similarly to how it can be solved in the separable state, as can be demonstrated. The patterns whose Lagrangian coefficients apply in the following relation after being determined are supported vectors:

$$0 \leq \alpha_i \leq C \quad (10)$$

The value of W and the shape of the differentiation function will be similar to the separable case. The supernatant obtained in an inseparable state is called the supernatant with a soft border area.

Nonlinear support vector machine

The support vector machines mentioned in previous sections use linear delimiters and a hyperplane to classify the patterns of a two-class problem, and are in fact the product of the internal product of the input vector with each of the support vectors in space. The next d input is calculated.

The differentiation function's form and W's value will resemble those in the separable situation. The supernatant with a soft border region is the supernatant obtained in an inseparable state.

Support-vector nonlinear machine

The support vector machines described in earlier sections, which are really the internal product of the input vector with each of the support vectors in space, employ linear delimiters and a hyperplane to categorize the patterns of a two-class issue. The following d input is computed.

Using the concept of internal multiplication in Hilbert spaces and the Hilbert Schmidt theorem, Vapnik showed that we can first transfer the input vector x by a nonlinear transformation to a space with a large dimension, in which we perform the internal multiplication space and prove that if a If the symmetric kernel meets the conditions of the Mercer theorem, applying this kernel to a low-dimensional input space can be considered as the product of the internal multiplication of a high-dimensional Hilbert space, greatly reducing computations[15]. For example, the kernel function can be in the following forms:

Vapnik demonstrated how we may first move the input vector x via a nonlinear transformation to a space with a big dimension, in which we do the internal multiplication, and demonstrate that if a Hilbert space has internal multiplication, then a Hilbert Schmidt theorem must hold. The internal multiplication of a high-dimensional Hilbert space may be used to apply the symmetric kernel to a low-dimensional input space if it satisfies the Mercer theorem's requirements, considerably lowering computations[15]. The kernel function, for instance, can have the following forms:

$K(x,y)=(xy+1)^p \qquad p=2,3,\ldots$

$K(x,y)=\exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$, K(x,y)=tanh(xy+θ) (11)

These nuclei are polynomial nuclei, Gaussian nuclei, and hyperbolic tangent nuclei, respectively. The problem of dual optimization in the integral and nonlinear state will be as follows:

These three types of nucleus are polynomial, Gaussian, and hyperbolic tangent, respectively. The dual optimization problem in the integral and nonlinear state will look like this:

$$\begin{Bmatrix} Maximum & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)+\sum_{i=1}^{N}\alpha_I \\ Subject\ to & 0\le\alpha_i\le C \end{Bmatrix} \quad (12)$$

Support vectors are patterns whose corresponding Lagrangian coefficients apply to the relation 0≤α_i≤C. A number of support vectors whose content Lagrangian coefficients are equal to 0 <α_i <C and whose number is N_b are used to calculate b:

Support vectors are patterns whose corresponding Lagrangian coefficients apply to the relation 0≤α_i≤C. A number of support vectors whose content Lagrangian coefficients are equal to 0 <α_i <C and whose number is N_b are used to calculate b:

$b_j = y_j - \sum_{i=1}^{N_{SV}} \alpha_i y_i k(x,SV_i)$, $b=\frac{1}{N_b}\sum_{j=1}^{N_b} b_j$ (13)

The decision function will be as follows:

$F(x)=\text{sign}\left(\sum_{i=1}^{N_{SV}} \alpha_i y_i k(x,SV_i)+b\right)$ (14)

Perhaps the current popularity of the support vector machine method can be compared to the popularity of neural networks over the past decade. The reason for this is the ability to use this method to solve various problems, while methods such as the decision tree can not be easily used in various problems.

In none of these methods is the generalization property of the classifier directly involved in the cost function of the method, and the designed classifier has little generalizing property.

Perhaps the rise in popularity of neural networks over the previous ten years can be paralleled to the present success of the support vector machine approach. The rationale for this is that, unlike approaches like the decision tree, this method can be utilized to address a variety of issues, whereas others cannot.

The constructed classifier has minimal generalizing property, and none of these approaches directly link the classifier's generalization property to the cost function.

If we consider pattern classifier design as an optimization problem, many of these methods face the problem of local optimization in the cost function and fall into the trap of local optimization.

There is another problem, and that is to determine the structure and topology of the classifier before design, for example, to determine the optimal number of hidden layer nodes in MLP neural networks, the number of Gaussian functions in RBF neural networks or the optimal number of snake modes and functions in the snake model. is. All these factors make it impossible to reach an optimal classifier in practice with the previously proposed methods.

Here the learning process is done in two parts:

Many of these techniques encounter the issue of local optimization in the cost function and fall victim to local optimization if pattern classifier creation is viewed as an optimization problem.

The structure and topology of the classifier must be determined prior to design; for instance, the ideal number of hidden layer nodes in MLP neural networks, the ideal number of Gaussian functions in RBF neural networks, or the ideal number of snake modes and functions in the snake model. With the previously suggested approaches, it is hard to achieve an ideal classifier in practice because of all these issues.

The learning process is divided into two steps here:

1- Structure training: which aims to learn the partitioning structure that affects the number of fuzzy rules.

2_Learning: Here the structure of the support vector machine is taught through fuzzy clustering.

Support vector machines have the following properties :

1- Classifier design with maximum generalization

2- Achieving the global optimal cost function

3- Automatic determination of the optimal structure and topology for the classifier

4. Modeling nonlinear differentiation functions using nonlinear nuclei and the concept of internal product in Hilbert spaces (a complete vector space whose norm is defined as absolute value by multiplication of fences. In quantum mechanics, Hilbert space is complex and basically infinite next, but In certain circumstances, its dimensions can be finite.)

To learn the partitioning structure that impacts the number of fuzzy rules, there is a first step called structure training.

2 Learning: In this case, fuzzy clustering is used to teach the structure of the support vector machine.

The characteristics of support vector machines are as follows:

1- A classifier designed with the most generalization possible

Realizing the global optimum cost function is step two.

3- Automatic selection of the best topology and structure for the classifier

4. Using nonlinear nuclei and the idea of internal product in Hilbert spaces (a full vector space whose norm is determined as absolute value by multiplication of fences), nonlinear differentiation functions are modeled. Hilbert space is complicated and essentially infinite in quantum physics, but its dimensions can occasionally be limited.)

III. PROPOSED METHOD

Full In the previous section, a review of the work done in the field of FS-FCSVM was done. In most of the work done, the optimization criterion on the number of fuzzy clusters and the number of SVs used to evaluate the quality. In this chapter, a new hybrid clustering method based on two algorithms of fuzzy clustering and differential clustering as well as determining SVM parameters using GRID SEARCH algorithm presented.

The goal is to reduce the number of support vectors and thus the number of rules in SVM by selecting the right clusters. Consider a set of n data {x_1, x_2,…, x_n} in the next M space.

Step 1: In the proposed method, we calculate the density value for each data according to Equation 15.

Full A review of the FS-FCSVM research was done in the section before this one. The majority of the work was evaluated for quality using an optimization criterion based on the quantity of SVs and fuzzy clusters. This chapter introduces a novel hybrid clustering approach based on fuzzy clustering and differential clustering, as well as GRID SEARCH algorithm for SVM parameter estimation.

By choosing the appropriate clusters, it is possible to decrease the number of support vectors and, consequently, the number of rules in SVM. Consider a collection of n data in the following M space: x_1, x_2,..., x_n.

Step 1: In the proposed method, we calculate the density value for each data according to Equation 15.

$$D_i = \sum_{j=1}^{n} exp(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2}) \quad (15)$$

Step 2: Select the data with the highest density as the centers of the clusters. For this step of the numerical constant 〖r〗 _a, which is a positive constant, determine the radius of the neighborhood. Hence a point in the data will have a high density value if it has a large number of points in the neighborhood.

The first center of the cluster (x_c1) is selected as the point with the highest density (D_c1). Then, the density value of each point x_i is re-evaluated as follows:

Choose the data with the highest density to serve as the clusters' centres in step 2.

Find the radius of the neighbourhood for this application of the positive numerical constant 〖r〗 _a.

A point in the data will thus have a high density value if there are many other points in the area The location with the highest density (D_c1) is chosen to be the cluster's first centre (x_c1).

The density value is then recalculated as follows for each location x_i:.

$$D_i = D_i - D_{c1}\ exp(-\frac{\|x_i - x_{c_1}\|^2}{(r_b/2)^2}) \quad (16)$$

Step 3: After recalculating the density for each data point, the next x_c2 center is selected and all calculations for the data point density are corrected again. This process continues until a sufficient number of points in the centers of the clusters are produced.

Step 4 and 5: The most critical part of the algorithm is fuzzy clustering and the performance of this clustering is highly dependent on the conjecture of the initial parameters. (To optimize this process, differential clustering algorithm is used to determine the number of centers. This algorithm itself has parameters such as neighborhood radius that is applied as input to the algorithm. The choice of this parameter is determined by the sensitivity of the work and data structure) .

Step 3: The new x c2 centre is chosen, and all computations for the data point density are rectified once again. This is done after recalculating the density for each data point. This procedure is repeated until enough points are created in the cluster centres.

Steps 4 and 5: The fuzzy clustering, which is the most important component of the process, is extremely sensitive to assumptions made about the starting values.

(The differential clustering method is utilised to calculate the number of centres in order to improve this procedure.

Neighborhood radius is one of the parameters that this method itself accepts as input.

The sensitivity of the job and data structure dictate the choice of this value.

As mentioned, the output of this algorithm will be the

number of cluster centers which will be applied as the input of the phase clustering algorithm. The fuzzy clustering algorithm is presented below.

Initialization of c, m and u primary clusters

Calculation of cluster centers

Calculation of the belonging matrix from the clusters calculated in step 2.

If the termination condition is met (u does not change much in different iterations) the algorithm terminates. Otherwise the algorithm continues from phase 2.

The number of cluster centres that will be used as the input for the phase clustering method will be the outcome of this algorithm, as was previously described.

The algorithm for fuzzy clustering is shown below.

Initialization of the fundamental clusters c, m, and u

Cluster centre calculation. the belonging matrix is generated using the clusters from step 2 as a starting point.

The method finishes if the termination condition is satisfied (u does not vary significantly between iterations).

Otherwise, step 2 of the algorithm is carried out.

In this algorithm, C represents the number of centers of the clusters, m represents the fuzzyness parameter, and in fact the scattering of the centers. In this work, m = 1.2 is used to increase the scattering, and U is the belonging matrix. The U-matrix shows the degree to which each data belongs to the cluster. In this work, we use this matrix to select the data by specifying a specific attribution value called K and comparing it to the degree to which each cluster actually belongs. Each row of the U matrix selects a data as a candidate for the training data. Of course, more points can be selected, but in the proposed algorithm, only one point is selected. So we will have training data for the number of clusters.

The fuzzyness parameter, m, and the dispersion of the cluster centres are all represented in this approach by the letters C and m, respectively.

In this study, the scattering is increased using m = 1.2, and U is the belonging matrix.

The U-matrix demonstrates how closely each piece of data is related to the cluster. In this study, we utilise this matrix to choose the data by defining an attribution value K and evaluating it against the degree to which each cluster genuinely belongs.

A data is chosen as a candidate for the training data in each row of the U matrix. Of course, additional points might be chosen, but the suggested method only chooses one. So we will have training data for the number of clusters.

Step 6: After selecting the training data, it is time for the SVM training. First, the GRID SEARCH algorithm used to select the kernel parameters (the proposed kernel is the RBF kernel) and SVM, thus obtaining two penalty parameters C and σ kernel width.

Grid search algorithm is a search method that examines all possible scenarios and is usually for tasks in which logic can not be found. This method involves systematically counting all possible candidates to resolve and check which candidate is able to fulfill the problem condition. For example, try all four-digit scenarios to find the ATM password. Now, if the result obtained from the Grid search algorithm is not acceptable, the algorithm can be re-evaluated with appropriate parameters and the desired result can be achieved.

Step 6: It's time to start the SVM training after choosing the training data. First, the SVM and GRID SEARCH algorithms were used to choose the kernel parameters (the RBF kernel was offered), resulting in the two penalty parameters C and kernel width.

Grid search algorithms are used to search across all situations and are typically used for problems where logic cannot be discovered. To answer the problem and determine which candidate may meet the criterion, this approach requires methodically counting all potential candidates. For instance, to get the ATM password, try each of the four-digit possibilities. Now, if the outcome of the Grid search method is unsatisfactory, the process may be reevaluated with the proper parameters to get the desired outcome.

This design is in fact an improved example of algorithms that have used only fuzzy clustering and in fact in this design the whole process including the number of clusters and the number of training data can be controlled and with these options the training process of support vector machine can be completely controlled Optimized construction. In this research, the problem of high SVM, which leads to an increase in the number of SVM rules, is solved by performing these steps.

This architecture is actually an improvement over methods that have just employed fuzzy clustering, and with these choices, support vector machine training may be fully controlled throughout the whole process, including the amount of clusters and training data used. optimised building. These procedures are used in this study to address the issue of high SVM, which results in a rise in the number of SVM rules.

IV. SIMULATION

In this section, experimental results reported to evaluate the proposed method on different data sets and parameters used. The data set used selected from UCEA data sets[16]. Also, the test results, which show the relatively high efficiency of the proposed method, are presented in this chapter.

The database specifications are used and the simulation parameters presented:

This section includes experimental findings that were used to assess the suggested approach using various data sets and input settings. The UCEA data sets were utilised to select the data set[16]. This chapter also includes the test findings, which demonstrate the relatively high efficacy of the suggested approach.

The simulation settings are supplied and the database requirements are used:

Table 1: Specifications of the data used

	Four class	Svm	Votes.84

Records	862	4000	435
Features	2	4	16
class	2	2	2

Table 2: Kernel parameters

Rbf kernel	Four class	Svm	Votes.84
gamma	1	2.6	2.4
C	500	500	650

In this section, the proposed algorithm and ordinary support vector machine and fuzzy support vector machine are compared in terms of runtime and clustering accuracy and number of support vectors:

In this part, the runtime, clustering precision, and number of support vectors of the proposed approach, as well as those of the conventional and fuzzy support vector machines, are compared:

Table 3: Comparison of the proposed algorithm and SVM in terms of time

	Four class	Svmguide1	Votes.84
Normal svm	0.276299	3.058864	0.024883
Proposed method	0.376222	1.336557	0.102236
Fs-fcsvm	0.371568	2.554789	0.065897

Table 4: Comparison of the proposed algorithm and SVM in terms of accuracy

	Four class	Svmguide1	Votes.84
Normal svm	100%	97.31%	100%
Proposed method	99.30%	96.24%	98.62
Fs-fcsvm	99.65%	96.25%	98.20

Table 5: Comparison of the proposed algorithm and SVM in terms of the number of support vectors

	Four class	Svmguide1	Votes.84
Normal svm	22	256	72
Proposed method	16	36	65
Fs-fcsvm	20	55	62

As can be seen, the results of the proposed algorithm in most cases are significantly different from the existing results of other algorithms. This result indicates the relatively high efficiency of the proposed method in optimization.

As can be observed, the results of the suggested method are typically very different from the results of other algorithms that are already in use. This outcome demonstrates the relatively high optimization efficiency of the suggested approach.

The large number of support vectors, which was one of the biggest problems of the support vector machine, was greatly reduced by the proposed method, which indicates the relatively high performance of this algorithm compared to the conventional support vector machine.

As stated in Chapter 3, after selecting the training data, it is time to train the SVM. First, the GRID SEARCH algorithm used to select the kernel parameters (the kernel suggested by the RBF kernel) and the SVM, thus obtaining two penalty parameters C and σ kernel width. Coming.

Here is an example of a GRID SEARCH algorithm for determining kernel parameters in SVM GUIDE1 data:

The suggested approach significantly decreased the vast number of support vectors, which was one of the main issues with the support vector machine. This shows that the algorithm performs rather well when compared to the traditional support vector machine.

It is now time to train the SVM, as was mentioned in Chapter 3, after choosing the training data. The first step was using the GRID SEARCH method to choose the kernel parameters (the kernel recommended by the RBF kernel) and the SVM, resulting in the acquisition of the two penalty parameters C and kernel width. Coming.

An illustration of a GRID SEARCH technique for locating kernel parameters in SVM GUIDE1 data is shown below:

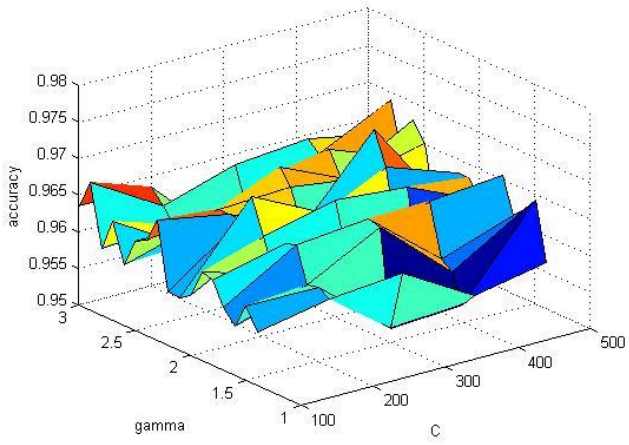

Figure 1: Execution of GRID SEARCH algorithm to determine kernel parameters in SVM

With the implementation of GRID SEARCH, the algorithm has changed a bit and the data related to the central points found from the differential search added to the points with a low percentage of belonging to the cluster, and the parameter related to the farthest value belongs to two parameters, ie the range.

And the parameters related to the neighborhood radius in all simulations are considered 0.1.

The following figure shows the four class data clustering diagram using the proposed algorithm in which 16 SVMs are used for SVM training.

The algorithm has changed slightly with the implementation of GRID SEARCH, and the information pertaining to the central points discovered through differential search has been added to the points with a low percentage of cluster membership, and the parameter pertaining to the farthest value now belongs to two parameters, i.e. the range.

Additionally, in every simulation, the neighbourhood radius-related parameters are taken to be 0.1.

The four-class data clustering diagram utilising the suggested approach, which employs 16 SVMs for SVM training, is shown in the accompanying picture.

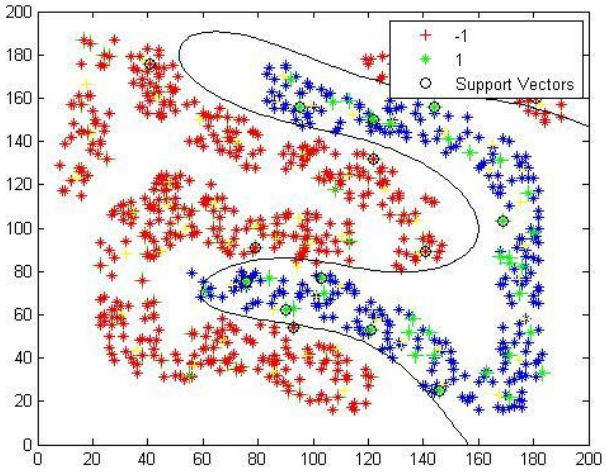

Figure 2: Clustering of Four class data using the proposed algorithm

In the last section conclusion of this paper presented.

V. CONCLUSION

In this paper, clustering based on fuzzy logic with the help of support vector machine was investigated. And SVs were getting too big, so it was not economical in terms of time and memory. Differential clustering algorithm for fuzzy clustering optimization introduced, which is both as an independent clustering method and as a preprocessing for other clustering algorithms. Fuzzy clustering algorithm that was sensitive to out-of-noise data and increased the number of SVs with the help of differential algorithm and GRID SEARCH reduces the number of out-of-date and ineffective training data and improves the quality of SVM. Thus, the number of SVs is greatly reduced Found and the category improved dramatically in terms of accuracy and timing.

This method has a relatively good performance for large databases and the results of experiments on different standard datasets show the high efficiency of the proposed method.

This study looked on fuzzy logic-based clustering with the use of support vector machines. Additionally, SVs were growing too large, making it less efficient in terms of time and memory. Introduced is a differential clustering algorithm, which may be used both as a standalone clustering approach and as a preprocessor for other clustering methods. Reduce the amount of outdated and inefficient training data and enhance the performance of SVM using a fuzzy clustering method that was sensitive to out-of-noise data and improved the number of SVs with the aid of differential algorithm and GRID SEARCH. As a result, there are a lot less SVs found, and the category has significantly increased in accuracy and timeliness.

This approach performs rather well for huge databases, and the outcomes of testing on various standard datasets demonstrate the suggested method's high efficiency.

REFERENCES

[1] Ruspini EH, Bezdek JC, Keller JM. Fuzzy clustering: A historical perspective. IEEE Computational Intelligence Magazine. 2019 Jan 11;14(1):45-55.

[2] Wang XL, Xie WX, Li LQ. Interacting TS fuzzy particle filter algorithm for transfer probability matrix of adaptive online estimation model. Digital Signal Processing. 2021 Mar 1;110:102944.

[3] R.P. Paiva, A. Dourado, Interpretability and learning in neuro-fuzzy systems, Fuzzy Sets Syst. 147 (1) 17–38.2004.

[4]J.C. Dunn; "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well Separated Clusters". Journ. Cybern. 3, 95-104, 1974.

[5] Osmar R. Zaïane: "Principles of Knowledge Discovery in Databases - Chapter 8: Data Clustering" .

[6] Mahmoudi MR, Baleanu D, Qasem SN, Mosavi A, Band SS. Fuzzy clustering to classify several time series models with fractional Brownian motion errors. Alexandria Engineering Journal. 2021 Feb 1;60(1):1137-45..

[7] Mishra R, Meher S, Kustha N, Pradhan T. A Skin Cancer Image Detection Interface Tool Using VLF Support Vector Machine Classification. InComputational Intelligence in Pattern Recognition 2022 (pp. 49-63). Springer, Singapore.

[8] Pier Luca Lanzi: "Ingegneria della Conoscenza e Sistemi Esperti – Lezione 2: Apprendimento non supervisionato".

[9] Maldonado S, López J, Vairetti C. Time-weighted Fuzzy Support Vector Machines for classification in changing environments. Information Sciences. 2021 Jun 1;559:97-110..

[10] Wiktorowicz K, Krzeszowski T. Identification of time series models using sparse Takagi–Sugeno fuzzy systems with reduced structure. Neural Computing and Applications. 2022 Jan 5:1-6.

[11] E. I. Papageorgiou, Ath. Markinos, and Th. Gemtos, "Learning Algorithms for Fuzzy Cognitive" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 42, NO. 2, MARCH 2012.

[12] Yücelbaş Ş, Yücelbaş C. Autism spectrum disorder detection using sequential minimal optimization- support vector machine hybrid classifier according to history of jaundice and family autism in children. Concurrency and Computation: Practice and Experience. 2022 Jan 10;34(1):e6498.

[13] C. Cortes, V. Vapnik, "Support-VectorNetworks", Machine Learning, Vol. 20, pp. 273-297, 1995.

[14] B. Schölkopf, A.J. Smola, Learning withKernels, MIT Press, Cambridge, MA, 2002.

[15] Fauzi IR, Rustam Z, Wibowo A. Multiclass classification of leukemia cancer data using Fuzzy Support Vector Machine (FSVM) with feature selection using Principal Component Analysis (PCA). InJournal of Physics: Conference Series 2021 (Vol. 1725, No. 1, p. 012012). IOP Publishing..

[16] Chang S, Shihong Y, Qi L. Clustering Characteristics of UCI Dataset. In2020 39th Chinese Control Conference (CCC) 2020 Jul 27 (pp. 6301-6306). IEEE.