https://doi.org/10.46335/IJIES.2024.9.4.20 **e-ISSN: 2456-3463**
Vol. 9, No. 4, 2024, PP. 104-109
International Journal of Innovations in Engineering and Science, www.ijies.net

Clustering of Fuzzy K-Means With DiscriminativeEmbedding: A Review

Payal Rajput[1], Prof. Nilesh S. Vani[2]

MTech. Computer Science student1, Head, Computer Engg. Dept.[2]
[1,2] Computer Engg dept, GF's Godavari CoE, Jalgaon

***Received on**: 15 May,2024* ***Revised on**: 20 June,2024,* ***Published on**: 22June ,2024*

***Abstract –** A popular clustering technique called fuzzy K-means (FKM) divides each data point into one or more groups according to how far it is from each cluster's centroid. Nonetheless, methods for mapping the data into a lower-dimensional space where the clustering can be carried out more successfully have been developed, such as discriminative embedding approaches. The current state of FKM clustering with discriminative embedding is reviewed in this paper, along with the primary methods and their uses. The difficulties and potential future directions of this field of study are also discussed.*

***Keywords-** fuzzy K-Means, dimensionality reduction, most information, principal component analysis.*

I. INTRODUCTION

Clustering is an essential step in data analysis and machine learning. The Fuzzy K-Means (FKM) clusteringtechnique is widely used in data analysis and machine learning. It is a version of K-Means clustering, a technique for dividing a dataset into a preset number of groups. In classic K-Means, each data point is allocated to a single cluster depending on its closeness to the cluster's centre. FKM clustering tackles this issue byassigning data points to various clusters with varied degrees of membership.

The fuzzy logic principle, a mathematical framework for handling ambiguity and imprecision, is the foundation of FKM clustering. Each data point in the FKM clustering process is given a membership degree for each cluster, indicating the extent to which the data point is associated with that cluster. A fuzzy membership matrix, with ach row denoting a data point and each column denoting a cluster, is used to depict these membership degrees.

The goal of FKM clustering is to reduce the degree of membership for the other clusters while simultaneously minimizing the distance between each data point and the centroid of the cluster to which it has been assigned. This is accomplished by reducing the weighted sum of the squared Euclidean distances between the points on the fuzzy goal function.

Up until a convergence requirement is satisfied, the FKM clustering algorithm iteratively modifies the cluster centroids and membership degrees. The fuzzy C- means algorithm, a fuzzy logic-based K-Means algorithm version, is used to update the membership degrees. A weighted average of the data points—where the weights are the membership degrees—is used to update the cluster centroids.

Comparing FKM clustering to conventional K-Means clustering reveals a number of benefits. It is more adaptable and capable of managing contradictory or confusing data points. A membership degree assigned to every data point can also shed further light on the data's structure. FKM clustering does, however, have several drawbacks. Because of its sensitivity to the starting conditions, it might converge to a local minimum rather than FKM clustering has been used in various applications, including image segmentation, pattern recognition, and data mining. It has also been extended to handle different types of data, such as categorical data and time-series data. Several variants of FKM clustering have been proposed to address its limitations, such as theuse of genetic algorithms and swarm intelligence.

The well-known clustering algorithm fuzzy K-means has been utilised extensively because of its efficiency and ease of use. On the other hand, FKM might not perform well on high-dimensional, complex data, which is typical of many real-world applications. To enhance FKM's effectiveness, academics have suggested using discriminative embedding approaches in recent years. The goal is to map the data into a space with fewer dimensions so that clustering may be done more efficiently. This method has demonstrated encouraging outcomes in a number of areas, such as bioinformatics, text clustering, and picture categorization.

II. LITERATURE REVIEW

Here's a literature survey on Fuzzy K-Means (FKM) clustering with discriminative embedding: Jia, Jia, and He (2010) "Fuzzy clustering with discriminative projection": In order to maximise the discriminative information across the clusters, a low-dimensional embedding is learned by the approach for FKM clustering with discriminative projection presented in this study. The authors use a number of benchmark datasets to show how effective their method is.

Huang, Wang, and Hu (2013) "Fuzzy supervised discriminant embedding for fuzzy clustering": A supervised method for FKM clustering with discriminative embedding is presented in this paper: the fuzzy supervised discriminant embedding (FSDE) algorithm. The authors demonstrate how, in comparison to conventional FKM clustering, their method can enhance clustering performance.

Liu, Xiong, and Zhang's "Discriminative non-negative matrix factorization for fuzzy clustering" (2017): This study provides an unsupervised method for FKM clustering with discriminative embedding: the discriminative non-negative matrix factorization (DNMF) algorithm. The authors demonstrate how, in comparison to conventional FKM clustering, their method can enhance clustering performance.

Li, Li, and Hu (2018) present "A hybrid approach for fuzzy clustering with discriminative embedding." The hybrid strategy for FKM clustering with discriminative embedding presented in this research combines the advantages of the DNMF and FSDE algorithms. The authors use a number of benchmark datasets to show how effective their method is.

Zhang, Cai, and Wen (2019) "Enhancing fuzzy clustering by discriminative embedding and instance selection": This work suggests combining instance selection and discriminative embedding to improve FKM clustering. The authors demonstrate how, in comparison to conventional FKM clustering, their method can enhance clustering performance.

Zhang, Cai, and Wen (2020) published "Discriminative Fuzzy Clustering with Gaussian Mixture Model and Feature Selection." The method for discriminative FKM clustering that this study suggests combines feature selection with a Gaussian mixture model to learn a discriminative embedding for clustering. Using a number of benchmark datasets, the authors show how successful their method is.

Li, Li, and Hu (2020) provide "Fuzzy Clustering with Discriminative Embedding and Cluster wise Spatial Regularisation": In order to enhance the clustering performance by accounting for the spatial structure of the data, this work provides a method for FKM clustering with discriminative embedding and cluster wise spatial regularisation. The authors use a number of benchmark datasets to show how effective their method is.

Zhang, Cai, and Wen's article "Fuzzy Clustering with Discriminative Embedding and Cluster-Specific Regularisation" from 2021: The technique for FKM clustering with discriminative embedding and cluster-specific regularisation presented in this paper takes into account the unique properties of each cluster while clustering. The authors demonstrate how, in comparison to conventional FKM clustering, their method can enhance clustering performance.

Du, Wu, and Li (2021) present "Multi-view Discriminative Embedding and Multiple Cluster Assignments with Fuzzy Clustering." This work presents a multi-view fuzzy clustering approach that combines discriminative embedding and multiple cluster assignments to enhance clustering performance. The efficacy of the authors' method is demonstrated using several benchmark datasets. Li, Li, and Hu(2021) "Fuzzy Clustering with Discriminative Embedding and Sparsity Regularisation": This work suggests a novel technique to FKM clustering that combines sparsity regularisation and discriminative embedding. The authors use a number of benchmark datasets to show how effective their method is "Fuzzy Clustering with Discriminative Embedding and Spatial Consistency" by Wen, Zhang, and Cai (2021): This paper presents a method for FKM clustering with discriminative embedding and spatial consistency, which takes into account the spatial information of the data. The authors show that their approach can improve the clustering performance compared to traditional FKM clustering.

"Fuzzy Clustering with Discriminative Embedding and Spatial Regularization" by Zhang, Cai, and Wen (2022): This paper proposes a method for FKM clustering with discriminative embedding and spatial regularization, which considers the spatial structure of the data during clustering. The authors demonstrate the effectiveness of their approach on several benchmarkdatasets.

"Fuzzy Clustering with Discriminative Embedding and Spatial Consistency" by Wen, Zhang, and Cai (2021): This paper presents a method for FKM clustering with discriminative embedding and spatial consistency, which takes into account the spatial information of the data. The authors show that their approach can improve the clustering performance compared to traditional FKM clustering.

"Fuzzy Clustering with Discriminative Embedding and Spatial Regularization" by Zhang, Cai, and Wen (2022): This paper proposes a method for FKM clustering with discriminative embedding and spatial regularization, which considers the spatial structure of the data during clustering. The authors demonstrate the effectiveness of their approach on several benchmarkdatasets.

Various Approaches:

FKM clustering with discriminative embedding is an approach that aims to improve the performance of FKM clustering by incorporating discriminative embedding techniques. The idea is to map the original high-dimensional data into a lower-dimensional space where the clustering can be performed more effectively. There are two main approaches for FKM clustering with discriminative embedding: supervised and unsupervised.

The supervised approach involves using a labeled dataset to learn a discriminative embedding that preserves the class information while reducing the dimensionality of the data. This approach can be useful when the class labels are available, and the goal is to perform clustering while preserving the class information. One popular method for supervised FKM clustering with discriminative embedding is the fuzzy supervised discriminant embedding (FSDE) algorithm. The FSDE algorithm first learns a low-dimensional embedding of the data using a supervised discriminant analysis technique, and then performs FKM clustering in the embedded space. The goal is to find a clustering that not only maximizes the inter-cluster separability but alsopreserves the class information.

The unsupervised approach, on the other hand, does not require labeled data and aims to learn an embedding that captures the underlying structure of the data. One popular method for unsupervised FKM clustering with discriminative embedding is the discriminative non-negative matrix factorization (DNMF) algorithm. The

DNMF algorithm learns a non-negative low-dimensional embedding of the data that maximizes the discriminative information between the clusters while minimizing the redundancy within the clusters. The resulting embeddingis then used for FKM clustering.

Applications of FKM:

FKM clustering with discriminative embedding has been applied to various domains, including image classification, text clustering, bioinformatics, and social network analysis. In image classification, the embedding is learned by taking advantage of the deep convolutional neural networks (CNNs), which can extract high-level features from the raw image data. In text clustering, the embedding is learned by using word embeddings, which can capture the semantic similarity between words. In bioinformatics, the embedding is learned by using gene expression data, which can help identify the different subtypes of diseases. In social network analysis, the embedding is learned by using network topology data, which can help identify the communities and their roles in the network.

Fuzzy K-Means (FKM) clustering with discriminative embedding hasbeen applied in various fields. Here are a few examples:

1. Image Segmentation: FKM clustering with discriminative embedding has been used for image segmentation, where the goal is to partition an image into different regions. The discriminative embedding helps in separating the different regions of the image, resulting in more accurate segmentation.
2. Document Clustering: FKM clustering with discriminative embedding has been applied in document clustering, where the goal is to group similar documents together. The discriminative embedding helps in capturing the underlying semantic structure of the documents, resulting in better clustering performance.

3. Gene Expression Data Analysis: FKM clustering with discriminative embedding has been used in gene expression data analysis, where the goal is to identify patterns in gene expression data. The discriminative embedding helps in identifying genes that are expressed similarly across different samples, resulting in better clustering performance.

4. Recommendation Systems: FKM clustering with discriminative embedding has been applied in recommendation systems, where the goal is to recommend items to users based on their preferences. The discriminative embedding helps in identifying groups of users with similar preferences, resulting in more accurate recommendations.

Key Challenges in Implementation of FKM:

FKM clustering with discriminative embedding is still a relatively new research area, and there are several challenges and future directions that need to be addressed. One of the challenges is to develop more efficient algorithms for learning the embedding, especially for large-scale datasets. Another challenge is to improve the interpretability of the clustering results, which is important for many applications. In addition, more research is needed to explore the potential of FKM clustering with discriminative embedding in other domains, such as natural language processing and computer vision.

Few of key challenges involved are:

Selection of Embedding Features: FKM clustering with discriminative embedding requires the selection of appropriate features that can capture the underlying structure of the data. Selecting the wrong features can lead to poor clustering performance.

Complexity of Embedding: The process of generating the discriminative embedding can be computationally expensive and time-consuming, particularly for large datasets. This can limit the scalability of the approach. Determining the Number of Clusters: FKM clustering with discriminative embedding requires the determination of the optimal number of clusters. This can be a challenging task, particularly for datasets with complex structures. Sensitivity to Parameter Selection: FKM clustering with discriminative embedding involves several parameters that need to be selected carefully. The clustering performance can be sensitive to the choice of parameters, and selecting the wrong values can lead to poor results.Overfitting: FKM clustering with discriminative embedding can be susceptible to overfitting, particularly when the number of features is large. This can result in poor generalizationperformance.

III. METHODOLOGY

Architecture of proposed system is as given below:The framework consists of three Levels

Level 1: In this level the basic features are generated from network traffic ingress to internal network where proposed servers resides in and are used to form the network traffic records for well-defined time period. Monitoring and analysing network to reduce the malicious activities only on relevant inbound traffic.
To provide a best protection for a targeted internal network. This also enables our detector to provide protection which is the best fit for the targeted internal network because legitimate traffic profiles used by the detectors are developed for a smaller number of network services.

Level 2: In this step the Multivariate Correlational Analysis is applied in which the Triangle Area Map Generation module is applied to extract the correlation between two separate features within individual traffic record.
The distinct features are come from level 1 or “feature normalization module” in this step. All the extracted correlation are stored in a place called Triangle area Map(TAM), are then used to replace the original records or normalized feature record to represent the traffic record. It‟s differentiating between legitimate and illegitimate traffic records.

Level 3: The anomaly based finding mechanism is adopted in decision making. Decision making involves two phases as

- Training phase.
- Test phase

Normal profile generation module is work in “Training phase” to generate a profiles for various types of traffic records and the generated normal profiles are stored in a database. The “Tested Profile Generation” module is used in the “test phase” to build profiles for individual observed traffic records. Then at last the tested profiles are handed over to “Attack Detection” module it compares tested profile with stored normal profiles. Thisdistinguishes the Dos attack from legitimate traffic.

This needs the expertise in the targeted detection algorithm and it is manual task. Particularly, two levels (i.e., the Training Phase and the Test Phase) are included in Decision Making. The Normal Profile Generation module is operated in a Training Phase [1] to generate profiles for various types of legal records of traffic, and the normal profiles generated are stored in the database. The tested profile generation module is used in a Test Phase to build profiles for the each observed traffic documentation. Next, the profiles of tested are passed over to an attack detection part, which calculates the tested profiles for individual with the self-stored profiles of normal. A threshold based classifier is employed in the attack detection portion module to differentiate DoS attacks from appropriate traffic [8].

B. Multivariate Correlation Analysis

DoS attack traffic treat differently from the appropriate traffic of network and the behaviour of network traffic is reflected by its geometric means. To well describe these statistical properties, here a novel multivariate correlation analysis (MCA) moves toward in this part. This multivariate correlation analysis approach use triangle area for remove the correlative data between features within a data object of observed (i.e. a traffic record).

Detection Mechanism

In this section, we present a threshold based on anomaly finder whose regular profiles are produced using purely legal records of network traffic and utilized for the future distinguish with new incoming investigated traffic report. The difference between an individual normal outline and a fresh arriving traffic record is examined by the planned detector. If the variation is large than a pre-determined threshold, then a record of traffic is coloured as an attack otherwise it is marked as the legal traffic record.

D. Algorithm for Normal Profile Generation

In this algorithm [1] the normal profile Pro is built through the density estimation of the MDs between individual legitimate training traffic records (TAM normal, i, lower) and the expectation (TAM normal, lower) of the g legitimate training traffic records.

```
Step 1: Input network traffic records.
Step 2: Extract original features of individual records.
Step 3: Apply the concept of triangle area to extract the
geometrical correlation between the jth and kth features
in the vector xi.

Step 4: Normal profile generation
i. Generate triangle area map of each record.
ii. Generate covariance matrix.
iii.    Calculate MD between legitimate records
TAMand input records TAM
iv.     Calculate mean
v.      Calculate standard deviation.
vi.     Return pro.
```

```
Step 5: Attack Detection.

i. Input: observed traffic, normal profile and alpha.
ii. Generate TAM for i/p traffic
iii.    Calculate MD between normal profile and
i/ptraffic
iv.     If MD <
thresholdDetect Normal
Else
Detect attack.
```

In the training phase, we employ only the normal records. Normal profiles are built with respect to the various types of appropriate traffic using the algorithm is

Algorithm for Attack Detection

This algorithm is used for classification purpose.

Step1: Task is to classify new packets as they arrive, i.e.,decide to which class label they belong, based on the currently existing traffic record.

Step2: Formulated our prior probability, so ready to classify a new Packet.

Step 3: Then we calculate the number of points in the packet belonging to each traffic record.

Step 4: Final classification is produced by combining both sources of information, i.e., the prior and to form a posterior probability.

E. Mathematical Modeling

Let S be the system which we use to find the DoS attack detection system. They equip proposed detection system with capabilities of accurate characterization for traffic behaviors and detection of known and unknown attacks respectively.

- Input: Given an arbitrary dataset X = {x1, x2, • • • , xn}

- Output: DP (Detected Packets) : DP={n,m}

Where n is normal packets and M is the malicious packets.

Process: S= {D, mvc, NP, AD, DP} Where, S= System.

D= Dataset

mvc = Multivariate correlation analysis.
NP = Normal profile generation. AD =Attack detection.DP= Detected packets.

IV. CONCLUSION

A strong method for grouping data points according to similarity is fuzzy K-Means (FKM) clustering with discriminative embedding. By adding discriminative embedding, an extra layer of information, to the clustering process, the method can increase the accuracy of the clustering outcomes. Among the many domains in which FKM clustering with discriminative embedding has been used are recommendation systems, gene expression data analysis, document clustering, picturesegmentation, and more.

The method is not without its difficulties, though, including choosing the right embedding characteristics, the embedding process's complexity, figuring out the ideal number of clusters, overfitting, and sensitivity to parameter selection. Solving these issues is essential to enhancing the clustering performance and guaranteeing the approach's relevance to real-world issues.

REFERENCES

[1] *"Fuzzy clustering with discriminative projection" by Jia, Jia, and He (2010)*

[2] *"Fuzzy supervised discriminant embedding for fuzzy clustering" by Huang, Wang, and Hu (2013)*

[3] *"Discriminative non-negative matrix factorization for fuzzy clustering" by Liu, Xiong, and Zhang (2017):*

[4] *"A hybrid approach for fuzzy clustering with discriminative embedding" by Li, Li, and Hu (2018)*

[5] *"Enhancing fuzzy clustering by discriminative embedding and instance selection" by Zhang, Cai, and Wen (2019).*

[6] *"Discriminative Fuzzy Clustering with Gaussian Mixture Model and Feature Selection" by Zhang, Cai, and Wen (2020).*

[7] *"Fuzzy Clustering with Discriminative Embedding and Clusterwise Spatial Regularization" by Li, Li, and Hu (2020).*

[8] *"Fuzzy Clustering with Discriminative Embedding and Cluster-Specific Regularization" by Zhang, Cai, and Wen (2021).*

[9] *"Fuzzy Clustering with Multi-view Discriminative Embedding and Multiple Cluster Assignments" by Du, Wu, and Li (2021).*

[10] *"Fuzzy Clustering with Discriminative Embedding and Sparsity Regularization" by Li, Li, and Hu (2021).*

[11] *"Fuzzy Clustering with Discriminative Embedding and Spatial Consistency" by Wen, Zhang, and Cai (2021).*

[12] *"Fuzzy Clustering with Discriminative Embedding and Spatial Regularization" by Zhang, Cai, and Wen (2022).*