FUZZY C-MEAN IMAGE CLUSTERING USING LDA AND INFINITE FEATURE SELECTION

A Dissertation Submitted in Partial Fulfilment of the Requirement for the Award of the Degree of

MASTER OF ENGINEERING

In

Electronics and Communication Engineering

Submitted By

PURNIMA

Roll No: 801661018

Under Supervision of

Dr. Amit Mishra

Assistant Professor, ECED

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY), PATIALA, PUNJAB

JULY, 2018

DECLARATION

I, Purnima hereby declare that the work presented in this thesis entitled "Fuzzy C-Mean Image Clustering using LDA and Infinite Feature Selection" in fulfilment of the requirement for the award of degree of Master of Engineering (ECE) submitted at Electronics and Communication Engineering Department, Thapar Institute of Engineering & Technology (A Deemed to be University), Patiala is an authentic record of work carried out under supervision of Dr. Amit Mishra (Assistant Professor, Electronics and Communication Engineering Department, Thapar Institute of Engineering & Technology (A Deemed to be University), Patiala) from 2016 to 2018. The matter presented in this has not been submitted either in part or full to any other university or institute for the award of any other degree.

Purnima

Date: 13/07/18

Purnima

801661018

Dr. Amit Mishra

Assistant Professor

Electronics and Communication Engineering Department

Thapar Institute of Engineering & Technology

(A Deemed to be University), Patiala, Punjab

Date: 13/7/2018

ACKNOWLEDGEMENT

I would like to express my gratitude to **Dr. Amit Mishra, Assistant Professor,** Electronics and Communication Engineering Department, Thapar Institute of Engineering & Technology (A Deemed to be University) Patiala, for his patient, guidance and support throughout this report. I am truly very fortunate to have the opportunity to work with him. I found his guidance to be extremely valuable. I am also thankful to our Head of Department, **Dr. Alpana Agarwal** and P.G. Coordinator, **Dr. Amit Mishra**. I would like to thank the entire faculty and staff of Electronics and Communication Engineering Department and then friends who devoted their valuable time and help me in all possible ways toward successful completion of this work. I thank all those who have contributed directly or indirectly to this work.

Lastly, I would like to thank my parents for their years of unyielding love and encourage they have always wanted the best for me and I admire their determination and sacrifice.

ABSTRACT

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters. Fuzzy c-mean is the most popular and commonly used technique for image clustering. FCM algorithm is a distinctive clustering algorithm and it makes use of the squared-norm to determine the similarity between prototypes and data points. Furthermore, several algorithms are developed by numerous authors based on the FCM with the aim of clustering more general dataset.

In this research work, features are classified with the help of linear discriminant analysis technique. Then after features representation, the features are ranked by using infinite feature selection approach, in which least rank feature which would not make any contribution of algorithm are neglected. After that image clustering is done on the higher ranked features thus decreasing the computational load of the program and improving the efficiency of the algorithm. The performance parameter like accuracy, sparseness and time complexity will be measured after stimulating proposed algorithm.

TABLE OF CONTENTS

LISTS OF THE TABLES

LISTS OF FIGURES

LISTS OF ABBERVATIONS

SVM	*Support Vector Machine*
ML	*Machine Learning*
FCM	*Fuzzy C-Mean*
EM	*Expectation Maximization*
LDA	*Linear Discriminant Analysis*
2DLDA	*2-Dimensional Linear Discriminant Analysis*
SVD	*Singular Vector Decomposition*
FD	*Fisher Discriminant*
IFS	*Infinite Feature Selection*
PCA	*Principal Component Analysis*
SED	*Squared Euclidean Distance*

CHAPTER 1

INTRODUCTION

1.1. PREAMBLE

Machine Learning approach was developed from computational learning theory and recognition pattern in artificial intelligence [1]. Moreover, it explores the construction of algorithms which could learn from the data and are also capable of making predictions on available data. Machine learning approach has wide range of tasks associated with computer like programming as well as designing algorithms having good performance, which is a quite difficult task.

It is closely associated with computational statistics that focus on prediction approach by the help of computers. This method has sturdy stalemates regarding optimization of mathematics to delivers theory, application and methods to the field. It can also be used with various other methods such as data mining [2] which is also known for un-supervised learning [3]. Whereas machine learning is also based on unsupervised learning approach and it can be used to create and learn the basic behavioral outlines for many different entities.

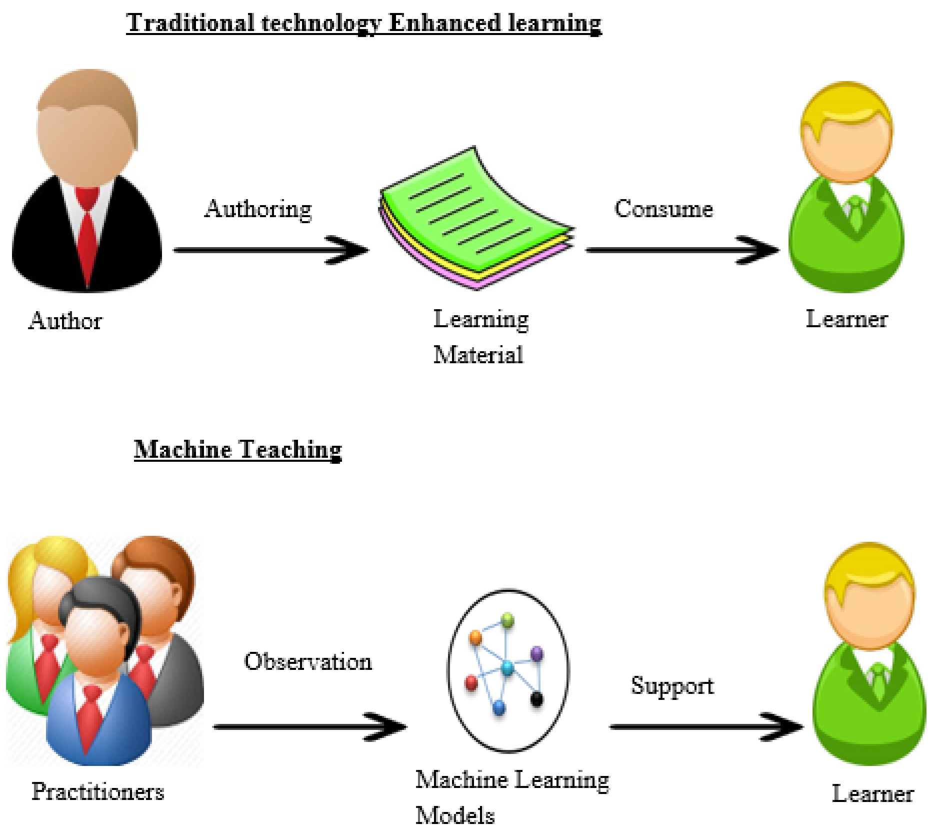


Figure 1.1: Visualization of Machine Learning method as a substitute to traditional learning approach [4].

In case of data analytics field, machine learning approach is used for contriving the complex algorithms and models to be predictive analytics. So, these analytics models allow folks to make reliable results or decisions. These models are also capable of revealing the insights which were hidden lately, by learning from the historical trends or relationships in the data [4]. Figure 1.1 demonstrates the visualization of machine learning approach as a substitute to traditional learning approach. As machine learning is a way different from traditional learning approach which use different learning material from machine learning.

1.2. INTRODUCTION TO MACHINE LEARNING

Machine learning based algorithms are applied to various problems such as in case of explicit, hard or un-desirable hand-coded solution is to be obtained. In this section, the various machine learning algorithms are described such as supervised machine learning and unsupervised machine learning approaches. Supervised and unsupervised machine leaning approaches are further classified into individual sub-classes and are represented in the figure 1.7.

The machine learning involves the concept of creating a machine which would be able to learn from the experience or learn from paradigms to decide. It also deal with interpretation of human learning and decision-making capability, thus its goal is to automate the course of solving problems.

All predictive learning-based algorithms comprises of two phases –

- To study or approximate the un-known dependencies in model from provided set of data samples.
- To make prediction based on the estimated dependencies of new output for every future value of input in model.

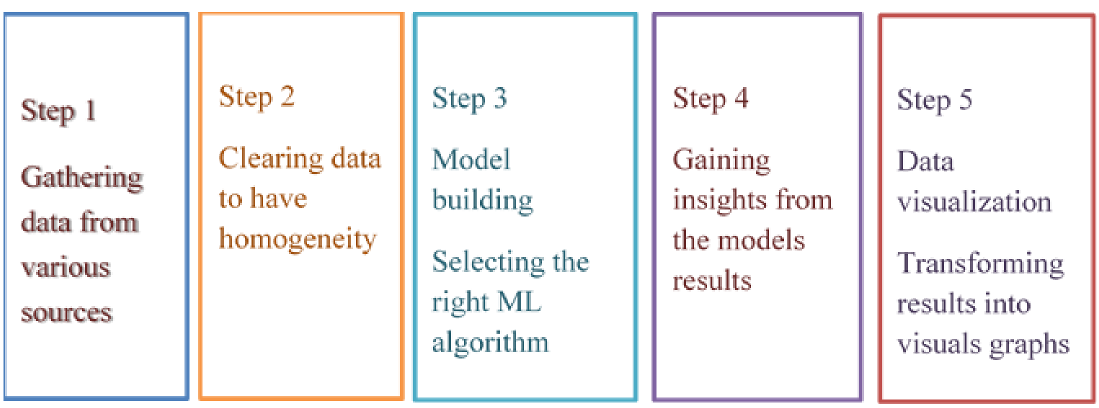


Figure 1.2: The process of Machine Learning

The above mentioned two steps relates to the classical interference which is named as deduction i.e. to give input values to the specific kind of output values and induction that means developing of a general mapping model from the particular training cases of data. These processes are formalized and implemented with the help of various learning techniques [6].

The Machine learning process includes 5 steps as depicted in figure 1.2. Firstly, the data is gathered from various sources, in second step the data is cleared to have similarity in the model, in third step model is built by selecting the appropriate machine learning (ML) algorithm, in fourth step the insights are gained from the model results and in the final step, visualization of data is done by transforming results into visuals graphs.

1.3 LEARNING MODES

Machine learning technique could be classified in three categories- the supervised learning approach, unsupervised learning approach and reinforcement learning approach, all are based on the obtain ability of the labeled training-set. However, the selection of appropriate learning models is a troublesome task and the choice is completely dependent on the type of learning problem.

1.3.1 Supervised Learning

This is the most common approach used in various fields. In supervised learning, input is map to the output labels and is done in classification perspective, when the input needs to be mapped to the continuous output. In supervised learning, ‘supervised’ refers to the output values of the training samples [7].

Suppose x is the input variable and y is the output variable, and the mapping function can be given as –

$$y = f(x) \tag{1.1}$$

The aim is to estimate the mapping function so appropriately that whenever new value of input x is added, model could be able to foretell the output variable y for that data. The basic model of supervised learning is shown in figure 1.3.

It is a process of algorithm gaining knowledge from the training data-sets and these algorithms generate predictions iteratively on the training data-set. This procedure of learning ends when algorithm accomplishes a satisfactory performance level. When operating the supervised learning approach, the main focus is on the bias-variance trade-off and complexity of model. Moreover, both of these are related to each other [8]. Complexity of model can be stated as the function complexity which is trying to learn, and nature of training data is used to determine the complexity of model. Low complexity model can be used when there is small quantity of data or data is spread inconsistently through-out the various possible scenarios. Because of this the models of high complexity would over-fit when operated for small quantity data-points [8].

Here over-fitting could be defined as a learning function which suits the training data but also doesn't generalize with other data-points. It could also learn to produce the training data with-out learning about actual structure in the dataset.

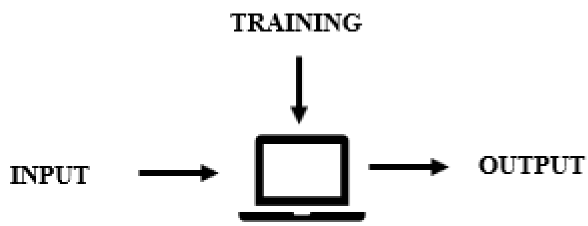


Figure 1.3: The basic model of supervised learning

The bias-variance can be defined as the variance, which is the quantity by which error can differ among various training datasets and the bias that is a constant error expression. So, this way model generalization is related to bias- variance [9]. In case of high bias and low variance, the model would be incorrect about 20% of the time but in case of high variance and low bias, then the model would be incorrect 5-50% of the time which depends on the data we use as training. Now it could be seen that bias and variance are indirectly proportional to each other as increase in variance leads to decrease in bias and with increase in bias there is decrease in variance. Normally, decrease in variance has consequences such as models with pledged base-line performance level which is considered as critical in particular tasks.

The variance of the model has to be scale along with complexity and size of training data, in order to generate a model that would generalize fine. Moreover, simple and small data set requires lower variance whereas complex and large data set require higher variance to learn the data structure [10].

Supervised Learning approach can be classified in two groups –

- Regression

  A regression is based on real output values such as weight or dollars. The aim of regression is to discover a particular relationship of the input which allows to efficiently generating the appropriate output. As it is known that, the appropriate output is determined from training data only. On other hand incorrect or noisy output will lessen the efficiency of the model.

- Classification

A classification is when the output category is variable, for instance any color such as blue, black or red or any disease or no disease. Some general form of problems based on regression and classification are time series prophecy and recommendation. Support Vector Machine (SVM) for classification, linear regression for regression and random forest is for both regression and classification, are some famous supervised based machine learning algorithms.

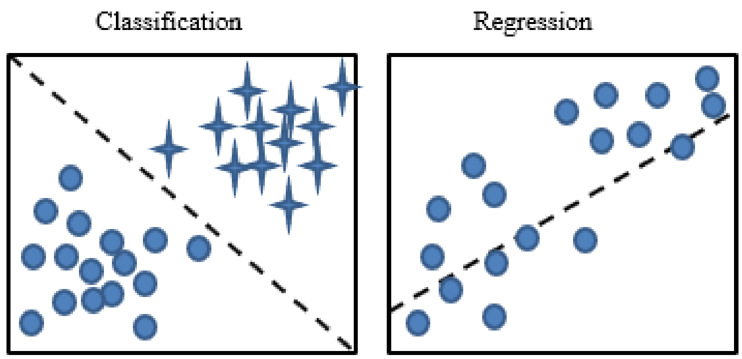


Figure 1.4: Difference between Classification and Regression.

Table 1.1 shows the more detailed difference between regression and classification based supervised learning algorithms. As classification can be describe as grouping of output into a class whereas regression is described as the prediction of the output value by using training data.

The classification based supervised learning and regression based supervised learning are quite different from each other as it can be seen from figure 1.4.

There are various supervised learning algorithms such as –

- Support vector machine
- Decision tress
- K-Nearest neighbors
- Logistic regression
- Linear regression

1.3.2 Unsupervised Learning

In this the training data is unlabeled and it is used to discover the under-lying structure of the data that depends on statistical properties like frequency. It is also used for analysis based on exploratory to discover the un-realized patterns. Moreover, it is used in case of lack of labeled training-based documents.

Table1.1: Difference between Classification and Regression

Difference between Classification and Regression	
Classification	Regression
• Classification can be described as grouping of output into a class.	• Regression is described as the prediction of the output value by using training data.
• It is used to predict the type of say tumor for example i.e. either harmful or not harmful using training data.	• It is used to predict, for instance the house price from the training data.
• If we have discrete or categorical variables, then it is classification-based problem,	• If we have real number or continuous variables, then it is regression-based problem.

Unsupervised learning can also be defined as when, there is input data x but no equivalent output variables. The aim of unsupervised learning is to acquire more about data by modeling under-lying structures present in the data.

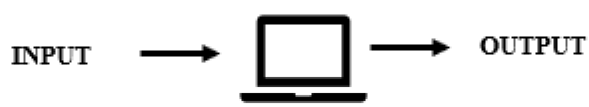


Figure 1.5: Basic model of unsupervised learning

The basic difference between supervised and unsupervised learning is that, the unsupervised learning does not have any teacher or appropriate answers. In this method the algorithms are on their own to find and portray the fascinating distribution of the data. Figure 1.5 display the basic model of unsupervised learning as there is no training data or labels.

The most popular errands of unsupervised learning are representation learning, estimation of density and clustering. In all these mention errands, there is no use of offered labels instead we learn from the intrinsic part of the data structure [11]. Generally, some algorithms in unsupervised learning uses principal component analysis (PCA), k-means clustering or auto-encoders, hence there are no label data so there is no exact way to evaluate the performance of models.

Unsupervised learning can be classified into two groups such as –

- Clustering– A clustering-based problem are those problems in which integral grouping of the data is to be determined, for instance grouping of customers according to their shopping behavior.

- Association– An association-based problem is when, there is requirement to determine the rule that would illustrate the large segment of the data, for instance if folks purchase x will also tends to purchase y.

The most generally used cases of unsupervised learning are dimension reduction and exploratory analysis.

- Dimension Reduction– In this method, the data is represented by using few features or columns which is achieved by using unsupervised approaches. The desire to discover the relationship among individual features in representation learning, allows us to epitomize data with the help of latent features which can inter-relate with the initial features. Moreover, this latent feature which is sparse in nature is often epitomized by using lesser features in comparison of the amount users initially began with. This is why further processing of data is less intensive and could eradicate the redundant features.

- Exploratory Analysis– This approach is capable of automatically identifying the structure present in data. For instance, when an analyst segments his consumers with the help of unsupervised clustering method would achieve a good starting point for his analysis. Unsupervised learning can offer initial insights when it's impossible for person to provide trends in data, which can be used as test personal hypotheses.

Most common exemplars of unsupervised learning methods are- k-means clustering-based problems, maximum entropy, a priori algorithm for association learning rule problems and hierarchical clustering-based problems.

Table 1.2: Difference between Supervised and Unsupervised Learning

	Supervised Learning	Unsupervised Learning
Input Data	• Uses known and labeled data input	• Uses unknown input data
User Role	• Select representative samples to be used as training data and arrange training data with label	Identity the no. of clusters to be generated, label each cluster and merge cluster of same label
Computer role	• Generate clusters based on training data and unclassified labels are not represented in the training data.	• Generate anonymous clusters
Real Time	• Uses offline analysis.	• Uses real time analysis of data.
Number of Classes	• Number of classes are known	• No of classes are unknown
Computational Complexity	• Complex computation	• Less complex computation
Accuracy	• Accurate and reliable results	• Moderately accurate and reliable results
Algorithms	• Support vector machine • Decision tress • K-Nearest neighbors • Logistic regression • Linear regression	• Clustering • K-Means • Dimensionality reduction • Principal component analysis • Neural networks

Unsupervised and supervised machine learning are quite different from each other as described in table 1.2. The difference between supervised and unsupervised learning is discussed based on various parameters such as- input data, real time analysis, computational time, accuracy, number of classes and various algorithms.

1.3.3 Reinforcement Learning

In reinforcement learning, the training data is specified only as feed-back in dynamic environment to the program actions, for instance playing a game with opponent or driving a vehicle [3]. It used to discover the finest actions to accomplish goals or to maximize reward. Moreover, in reinforcement learning machine acquires how to performance in a particular environment. The basic model of reinforcement model is shown in figure 1.6.

In this learning there is no preferred type of signal provided despite of that there is feedback of that uncertain category to be either wrong or right. For instance, it is equivalent to a reviewer that only utters something about being wrong or right instead of being specific about how that is wrong [12].

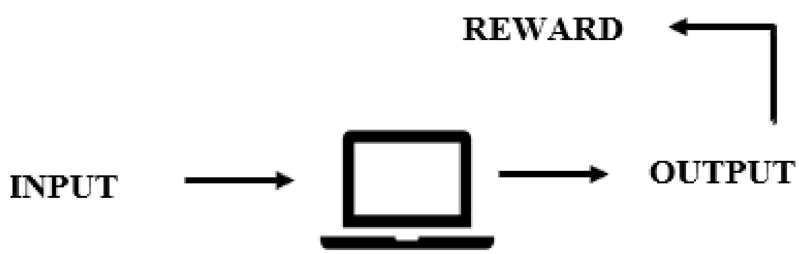


Figure 1.6: Basic model of Reinforcement Learning

Reinforcement Learning approach have various algorithms based on it such as-

- Monte Carlo methods
- Temporal difference learning
- Markov decision processes
- Learning tasks
- Robot navigation
- Real time decisions
- Skill acquisition

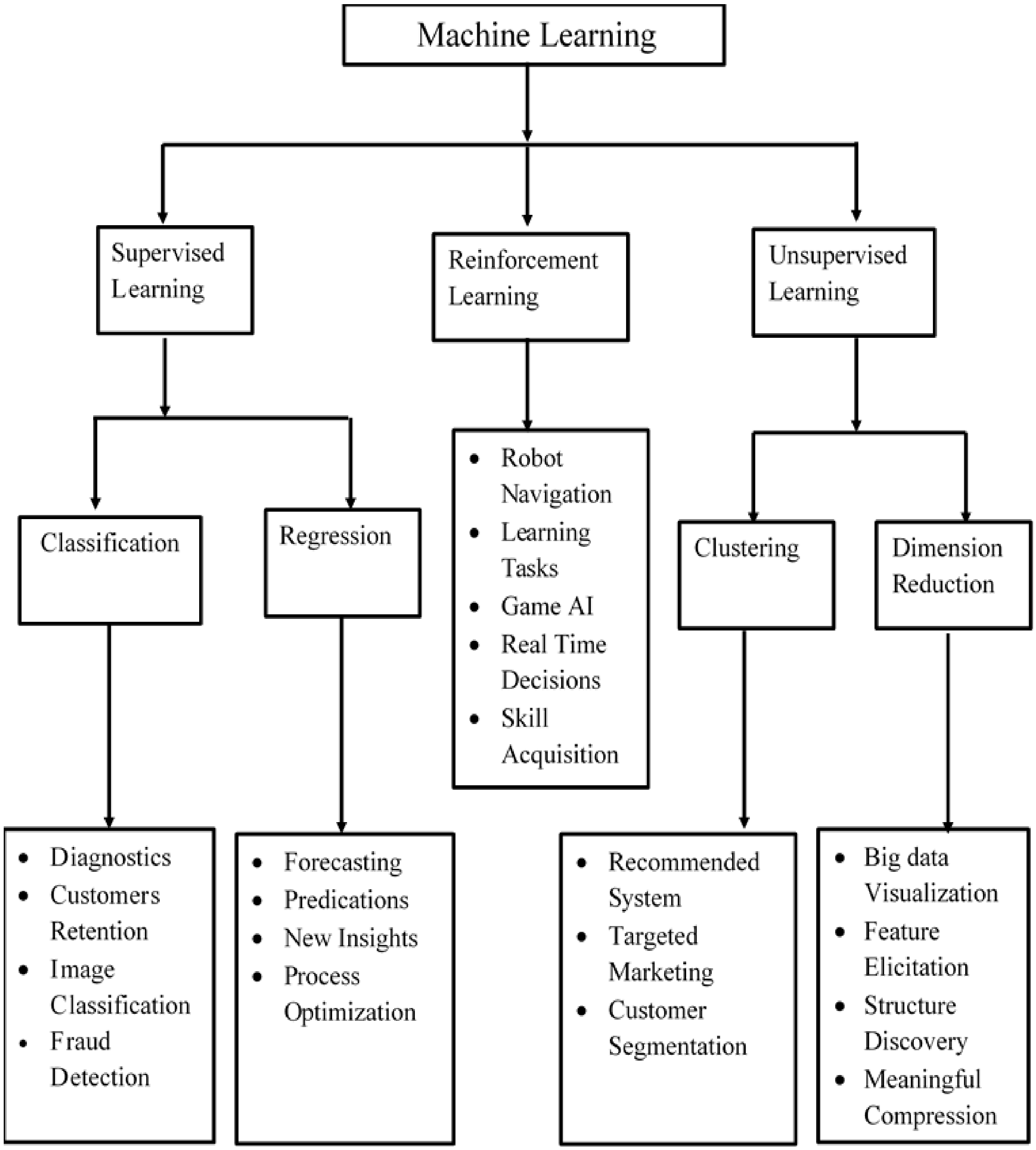


Figure 1.7: Classification of Machine learning algorithms

1.4 APPLICATIONS OF MACHINE LEARNING

Machine learning technique is widely used in various fields and its application is almost limitless. Few of machine learning applications are listed below-

- Marketing and Sales- Now a days digital marketing is most common used, It could be used to analysis customers likes and dislikes and also use that to improve the sales of the products. Marketing and scale department now can easily analysis the data, the customer pattern to purchase and can understand how they could enhance the buyer's journey.
- Financial Services-This sector has grown so much so it the complexity, the financial companies have to evaluate the wide and diverse data-sets. Machine learning is used to detect the market risks, opportunities for trading, to detect activities related to fraud and to identify the investment.
- Data Security- In recent years, data security has come out to be the first priority as there are ransom ware security attacks, for instance Petya and Wanna-cry. The malware tends to be dependent on previously used architecture along with some minor technical variations and changes in the code. So, machine learning approach can analysis the malware and search for the pattern or variations in the code and it capable to stop malware with prodigious accuracy in short period of time as well.

Apart from of these, there are few more common applications of machine learning such as-

- Trading derivations and stocks
- Factory diagnostics maintenance
- Search engine
- Determining about which voters could be canvassed in election
- Identifying human genes
- Programmatically reading random text from photograph
- Anti-virus software
- For insurance companies
- Autonomous, self-driving cars
- Predicting music tastes
- Predicting movies taste
- Programmatically face recognition
- For real-estate firms predicting house prices

1.5 ORGANIZATION OF THE DISSERTATION

The dissertation is organized in chapter formation as mentioned below-

Chapter 2: **Literature Review**– In this chapter, a comprehensive study on various methods and algorithms available are deliberated. The historical background of fuzzy c-means image clustering and linear discriminant analysis along with the contribution of eminent researchers in this field is also discussed. This chapter also includes objective and research gaps of this dissertation.

Chapter 3: **Proposed Methodology–** In this chapter various image clustering techniques are summarized, and various performance parameters are also deliberated. The proposed method is enlightened in detail for comparative study.

Chapter 4: **Results and Discussions–** Simulated results of proposed method are mentioned along with comparison among other existing algorithms in chapter 4. All the performance parameters like accuracy and sparseness is also discussed.

Chapter 5: **Conclusion and Future Scope-** In this chapter all the work done on proposed algorithm is summarized. Future scope is also discussed based upon the observation drawn from the results.

CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION

In this chapter, the discovery of various architecture, methods and scheme are mentioned which has been expounded by various researches from time to time. The various algorithms and methods have been explained briefly that give an insight of Fuzzy C-Means (FCM) image clustering approach. This literature survey delivers copious emphasis in the image clustering field and thus specifies the objective of this dissertation. This chapter establishes the research gaps and defines the statement of problem.

2.2 LITERATURE REVIEW

Bezdek *et al.* [12] in 1984, in this FCM is performed on FORTRAN-IV. In this program author produce fuzzy partitions and prototypes of numeric data. The produced partitions were useful in verifying or evoking known sub-structures present in data which was unexplored. The clustering criteria author used to combine sub-sets is the least- square cost function and researchers include three norms that are diagonal, mahalonobis and euclidean, these were used to control the sensitivity regarding noise, quantity of clusters, output which would include various measure of validity of cluster and variable acceptance.

Hung *et al*. [13] in 2001, author proposed a novel algorithm providing better efficiency than general FCM algorithm called psFCM algorithm. This algorithm divide the data-set into various unit blocks and the centroids of these unit blocks would replace the pattern and formed new data-set which would be the more simplified data-set. There proposed psFCM algorithms would have computing time complexity from $\mathrm{O}(\mathrm{fNC})$ to $\mathrm{O}(\mathrm{f}\,N_{ps}\mathrm{C})$ for every repetition and also for the large data-set their algorithm have shown improved performance as compared to FCM algorithm. Author also shown that in their algorithm, an initial center of cluster chosen from unit-block which has higher density is close to the actual centroids of cluster.

Howland *et al*. [14] in 2004, in this research work researchers proposed generalizing approach for discriminant analysis with the help of generalized singular value decomposition (SVD) method. Discriminant analysis can be expressed as a problem based on optimization involving co-variance matrices which would represent the within cluster scatter and in between cluster scatter. One of these matrices needs to be non-singular, which apply limitation its application that is to consider data-sets with specific relative dimensions. Researchers used generalized SVD to evade the requirement of non-singularity. Moreover, the generalized discriminant analysis could be used for sample size even smaller in comparison to sample data dimension. The J_1 criterion efficiently optimizes the classification in diminished dimensional space and their algorithm also avoids the problems regarding numerical which are inherent in explicitly establishing a scatter matrix.

Their proposed algorithms also have some disadvantages such as the GSVD have high computational cost. The cost of orthogonal decomposition of k was $O\,(\mathrm{nmt})$ in case of $\mathrm{m} \leq \mathrm{n}$ and $O\,(m^2\mathrm{t})$ in case of$\mathrm{m} > n$. The orthogonal centroid is less expensive than LDA/GSVD algorithm, by solving the simple eigen value-based problem and by evading computations based on eigen values.

Wehrens *et al.* [15] in 2004, researchers used model-based clustering and applied it directly to the large data-sets when is proven to be slow for every practical application. The most common approach for this was to select the cluster out of samples which are chosen randomly, of a not too large size and then apply the clustering model obtain by this technique to classify the remaining objects. Authors prove that this easiest method leads to un-stable results. So, to overcome this researcher proposed a more stable method which perform better by just making two modifications – first is to identify the various uncertain models from the sample but despite of only one and the second is to use several EM steps instead of using a single E- step to do classification of whole data-set.

Ye. J. *et al.* [16] in 2005, researcher forthput a novel approach of 2-Dimensional Linear Discriminant Analysis (2DLDA). This approach has overcome the problem of singularity as well as has achieved high efficiency. The basic difference between LDA and 2DLDA is that LDA deal with factorized data representation where as 2DLDA deals with matrix data representation. 2DLDA is used for further dimension reduction, but in order to achieve this it was combined with LDA thus forming 2DLDA+LDA algorithm which was later compared to traditional combination of LDA+PCA. Researchers proved that their algorithm requires lower memory, have lower computational time and can used for large amount of data.

Li. T. *et al.* [17] in 2006, researchers worked on discriminant analysis for multi-class classification. The support vector machine (SVM) is very significant for binary class classification but researchers can't be prolonged to their multiclass counter-parts. SVM are comparable to the fisher's discriminant (FD) as the result of decision hyper-planes of classification in case of binary are quite equivalent. The discriminant analysis is used to learn the discriminative transformations of feature in any statistical based pattern recognition as well as it can be extended to multiclass very easily. Researchers applied the discriminant analysis to multiclass problem and thus obtained quite efficient results and also the time consumption of discriminant analysis is much lesser than SVM.

Ding *et al.* [18] in 2007, authors combined linear discriminant analysis (LDA) with K-means clustering to a coherent frame-work in order to choose the subspace which is most discriminative. This research paper used k-means to have the class labels and LDA for the subspace selection. Researchers integrate the clustering process with the process of sub-space selection and then concurrently the data is clustered, meanwhile feature sub-space is chosen. The proposed method is performed on real-world data-sets to prove the effectiveness of the algorithm. Moreover, the proposed algorithm is rich in LDA-k mean frame-work by comparing their relationship with other approaches and by examining their variants.

Krinidis *et al.* [22] in 2010, researchers proposed an algorithm in new fuzzy way along with gray level information and spatial local information named as Fuzzy Local Information C-Mean (FLICM) technique. Their proposed algorithm sure the robustness to noise, guarantee the preservation of image details, improve the clustering performance and overcome the drawbacks of FCM algorithm. In this, also a novel factor G_{ki} was introduced as local similarity measure to make proposed algorithm robust from outliners and noise. Therefore, in absence of noise in algorithm, make FLICM technique an efficient approach for clustering application. In this algorithm, the gray level and spatial information are merged together by G_{ki} factor and thus making FLICM robust from every outliners and noise.

Their proposed algorithm provides parameter free determination and balance among details in image and noise, simultaneously improving clustering performance. FLICM algorithm performs clustering on original image while other algorithms perform on images which were pre-computed. The experiments of FLICM were carried out on real world and synthetic images and this algorithm is completely free from adjacent parameter empirically, earlier which was used in every FCM algorithm.

Kannan *et al.* [19] in 2012, researchers deal with an effective quadratic entropy fuzzy c-means approach in which the combination of regularization-based function, mean distance, kernel distance and quadratic terms are used. In standard FCM algorithms computation of standard cost function is a task because of un-certainty in data objects and huge amount of data. Moreover, FCM algorithm is also incapable of setting optical parameters in clustering approaches. So, researchers proposed a generalized FCM clustering algorithm named as quadratic entropy based FCM (QEFCM) technique to deal with complex data.

Researchers establish an efficient way to estimate membership and to update centers through minimization of proposed cost function. For reduction of number of iteration in proposed method, a novel technique is used for initialization of the cluster centers. To obtain validity of clusters and selecting the number of clusters in proposed algorithm, the silhouette scheme is used. Moreover, to evaluate and investigate the effectiveness of the proposed technique, researchers have done experiments on the datasets based on time series which have significant role in field of engineering and science.

Zanaty [20] in 2012, researchers calculate the appropriate validity criterion of kernelized FCM algorithm (KFCM) and this algorithm with spatial constraints (SKFCM) for computerized segmentation of magnetic resonance imaging (MRI). The original euclidean distance is substituted with Gaussian radial basis function classifier (GRBF). These proposed algorithms are executed on eighteen popular indexes, to decide if these indexes are proficient to obtain optimal clusters number. Researches also measure the segmentation performance by employing these methods individualistically on various data-sets to decide which method would give efficient results. The indexes are compared and calculated by applying them on several test-images which also include corrupted images because of noise with diverging level and MRI data-set which should be simulated volumetric.

Havens *et al.* [45] in 2012, the researchers worked with very large (VL) data which is difficult to load on one's computer. The clustering techniques which can scale very large data are more useful and significant. In this research work a clustering approach is imposed on VL dataset. The aim was to extend FCM algorithm to large dataset. The comparison of the approaches is done, which were based on – Kernel zed FCM that give approximation depending on sampling, sampling ensued by non-iterative extension and incremental approaches which would make one pass sequentially through the data subsets. The proposed work is exhibit a very large technique on 5 billion object-based data-set plus it signify the recommendation set concerning the exploitation of various fuzzy c mean based clustering techniques.

Zaghmouri *et al*. [57] in 2012, Fuzzy c-means clustering plays an important part in all clustering-based applications despite of that sometimes FCM algorithms results are deliberated as in-complete strategy in clustering. So, the proposed technique an adapted FCM technique which is capable of creating clusters of similar size. Moreover, those points which are scatter away from every cluster were also grouped. The work includes the addition of computational data grouping in FCM algorithm to obtain similar size of clusters, plus to calculate the points belonging which are located among clusters. The aim was to propose an algorithm with high performance which would also group and sort data-set in different clusters number which is then used in managing and controlling those clusters.

Gong *et al.* [21] in 2013, this research work represent an improved and more efficient FCM algorithm for segmentation of image while presenting kernel metric and a weighted trade-off fuzzy factor. The latter one is based on the distance of space of every neighboring pixel as well as their gray level variance concurrently. Due to this fizzy factor, their proposed algorithm is capable of estimating accurately the extent of damping of neighboring pixels and the kernel distance is also introduced to measure its cost function in order to improve its noise and outlier's robustness. The proposed technique is capable of determining the parameter of kernel by means of fast bandwidth miscellany rule which depends on variance distance of every data-points from the collection and both kernel distance and trade-off weighted fuzzy factor are free from determination of parameter which permits automation dependent applications.

Researchers used unsupervised FCM technique dependent on kernel metric. This technique is used for segmentation of novel trade-off weight which is commonly based on local information distribution and in neighbors, the damping amount of pixel is affected by the constraint of local spatial. Moreover, the proposed algorithm was examined on natural images, synthetic images and medical images. There novel algorithm has proven to improve the robustness to noise and performance of segmentation of images.

Fahad *et al.* [23] in 2014, author provide a survey on clustering algorithms which are powerful tool for meta learning to efficiently analyze large amount of information produced by modern applications. The aim was to provide a categorize data in form of clusters in such a way that data points are grouped in similar cluster. The major issue in application of clustering technique is huge amount of data which cause lack of formal categorization and lack of definition agreement of their properties. Author introduces algorithms and concepts based on clustering, and also provide comparison based on empirical as well as on theoretical basis.

From empirical perspective, author conducted experiment for comparison of various algorithms based on their performance for huge datasets. Where as in theoretical perspective, author established a framework depending on the important properties mentioned in previous studies. The performance comparison is carried out on the basis of various parameters such as stability, scalability measure, runtime and number of external as well as internal validity metrics.

Abdu *et al.* [24] in 2014, author provide a survey on fuzzy c- mean algorithm in which the fuzzy logic is used in classifying image objects into objects which are whether changed or unchanged and is an arrangement of c-means in the grouping process of pixels into an object. Image registration deals with problem regarding misalignment that is solved with the help of FCM algorithms. Moreover, the result of change detection accuracy not only depends on image registration but also on classification and segmentation accuracy as the error in image registration is resolved in the phase of classification and segmentation.

In FCM clustering algorithm image pixels are divided into fuzzy clusters with the help of iterative optimization of the cost function along with the update of center of cluster and membership. Much of the FCM dependent algorithm deals with the errors of image registration. This research work, review algorithms based on image registration errors and for classification and segmentation phase, the fuzzy-statistics -based -affinity -propagation c-means clustering technique. The FCM algorithm also considers the entire cluster similarly due to which the poor solution instigated by hard decisions is prevented during clustering process. Because of evading mis-classification of small clusters, provides good segmentation and classification quality results.

Arora *et al.* [25] in 2014, this research work provides a brief survey on various clustering algorithms in the case of huge data analysis. The data mining is described as a method in which extraction of hidden relationships and useful information is done. As the analysis for huge data is problematic and traditional data mining methods are not capable of directly working on huge datasets because it cannot analyze huge amount of data. Clustering algorithms are widely used for data mining approaches, which perform mining process by discovering clusters of similar data. This work provides a comprehensive analysis of algorithms and also provides the appropriate clustering-based technique on the basis of comparison done on various parameters such as cluster quality and execution time.

Rajapriyadharshini *et al.* [26] in 2015, researchers worked on synthetic aperture radar (SAR) image analysis. Digital images plays an important part in day to day routine application for instance geographic information system, satellite television, area of remote sensing and magnetic resonance images. The data-set collected by sensor based on image, usually contain noise and the noise which affect the image is called as granular noise which is also named as speckle noise.

The major issue in SAR image processing technique is the speckle noise and to tackle with this issue, the researchers used Linear Discriminant Analysis. The clustering approach is used for denoising SAR images, researchers cluster the noisy image in various dis-joint local sections, which involve spatial structure which are similar in nature and then every spitted section is again de-noised with the help of wiener filter in LDA. Further, in LDA domain it verify for LMMSE shrinkage. The experiment outcomes display denoised patches of all clusters which are then used for reconstructing of image which is noise free.

Memon *et al*. [27] in 2016, researchers proposed a generalized FCM clustering technique along with local information (GFCMLI) to overcome the standard FCM techniques disadvantages. Much of FCM algorithms are for image segmentation that include neighbor information in their cost function to deal with the problems regarding poor performance and sensitivity to noise. The other FCM based algorithm suffers from demerits such as- robustness to noise, limitation to single feature input information, effectiveness depends on crucial parameter α and difficulties in determining the optimal values of crucial parameter i.e. α, that is chosen experimentally. The P_{ki} fuzzy factor is used in the form of a local similarity measure in data to improve the performance of clustering algorithms. It also provides robustness regarding noise and outliners. The proposed algorithm GFCMLI does not require pre-defined parameter α instead of this; researchers used automatically calculated distance weightD_i. The GFCMLI algorithms overcome the drawback of poor performance of standard FCM algorithm by using various sizes and densities of clusters. The test of performance of proposed algorithms is carried out by performing it on real world images and noisy datasets to exhibit the efficiency, noise robustness and effectiveness of the algorithm.

Saxena *et al*. [28] in 2017, in this author provide a comprehensive review on various clustering methods. Clustering is usually defined as an un-supervised learning approach in which similar object are grouped. Clustering technique is further categories in parts for instance- partitional, density based, grid, model based and hierarchical approach. Author provides a detailed discussion on these methods along with applicability and their state of art respectively. The application of clustering are also discussed in this research work such as character recognition, image segmentation, data mining and object recognition. Different kind of similarity criteria used in clustering have been discussed in this research work and after cluster formation, the clustering methods are evaluation on the basis of accuracy and performance parameter.

In hierarchical clustering approach, by using iterations division of patterns to from clusters in bottom-up or top-down manner whereas in partition-based clustering data is assigns into k-clusters by doing optimization of criterion function. Most popular method to find euclidean distance among the points is by availability of every cluster and by assigning every point minimum distance in the cluster. Data objects are supposed to be produced with respect to various probability distributions in mixture density-based clustering scheme and obtained from various kind of density function can be t-distribution or multivariate gaussian distribution or from several parameters having similar families.

Chen *et al.* [29] in 2018, researchers used FCM algorithms with multi objective optimization. FCM algorithm is proven to be an effective and efficient algorithm for segmentation. The main drawback of FCM algorithms are there sensitivity, to noise so the researchers proposed a new FCM algorithm with multi objective optimization to over the FCM sensitivity to noise. Researchers adjoined a novel parameter λ in the fuzzy measurement formula of distance to have improved multi objective optimization. This parameter could manage pixel weights of local information. Moreover, in proposed technique to improve the multi objective mathematical model, the neighboring pixel of local correlation is merged to optimize clustering cent. Matlab 7.0 is used as programming tool and to prove the efficiency of the proposed algorithm, researchers performed the algorithm with different kinds of noise such as salt and pepper noise and gaussian noise. The performance of proposed algorithm is also compared to various other clustering algorithms.

2.3 GAPS AND OBSERVATION

FCM clustering techniques are most commonly used unsupervised technique which have many applications in various field of study for instance medicine, pattern detection, bioinformatics, data mining application. Researchers have developed various algorithms as some of them are describe in above section, on the basis of their aim for clustering.

- In comparison to various other clustering algorithm fuzzy c mean is more efficient, robust and reliable approach in particular applications or say cases according to its performance [30].

- FCM clustering algorithm can improve their performance in terms of computational time.
- FCM algorithms have more sensitivity to initial guess such as local minima or speed [44].
- This clustering algorithm provides high sensitivity to noise and there is requirement of lower membership degree for noisy points or outliners.
- LDA technique seen as efficient dimension reduction approach, hence this technique transform original data or information in lower dimensional space which is then calculated with the help of different $n-1$ eigen-vectors, n represent quantity of different classes.
- To have more efficient LDA as dimension reduction technique, there is need of better understanding of its interpretations.
- FCM algorithms are proven to be more suitable in case of clustering overlapping based errands.
- FCM clustering algorithms have limitations such as the sum of membership value is 1, because of which higher membership value points are on the cluster's outer. That is the reason why FCM algorithm faces difficulties in managing the points which are outliner [44].
- The value of data point membership in cluster directly dependent on other clusters along with their value of membership, because of that unwanted results are received.

The FCM technique was used to specify the similarity among data points and prototypes. Moreover, fuzzy c-means algorithms assign membership on the basis of distance between data point and cluster center, to every data point related to each center of cluster. More the data near to the center of cluster more would be its membership value of that center of cluster. This clustering method is an iterative process and the cluster center and the membership value after every iteration is updated. There is some suggestion mentioned below to help to improve the performance of the FCM algorithms.

- To improve computational time of fuzzy c-means algorithm, decision tree scheme can be used which would extract data sequentially and accurately.

- To further improve the performance of FCM algorithm, the noise sensitivity of these algorithms should be reduced.

2.4 OBJECTIVE

The main objective of this dissertation work is specified below–

- To study the various existing works on image clustering techniques.

- To propose a better algorithm to improve the performance of fuzzy c means clustering algorithm.
- To increase the efficiency.
- To provide robustness to noise in FCM algorithms.
- To compare the existing work with the proposed method in terms of sparseness and accuracy.

CHAPTER 3

PROPOSED METHODOLOGY

3.1 INTRODUCTION

Clustering is regarded as unsupervised classification method which is used to group data into clusters depending on their similarities [31]. One of the important tasks for image clustering is to effectively represent the raw data. Moreover, there are various clustering-based methods used for different applications such as image indexing [32], image classification [33], image annotation [34], image segmentation [35] and image retrieval [36].

Clustering has also been used for feature construction in which group of similar variables is replaced by cluster centroids which then become a feature. The hierarchical clustering and k-means are some of quite popular algorithms. In clustering, unsupervised learning is used to innovate supervision to gain some more discriminant features. Grouping of similar entities simultaneously could aid to profile attributes of dissimilar groups or in other words it could give intuition about basic patterns of diverse groups. These grouping of similar data found various applications such as for segmentation of customers, marketing of different groups for maximizing revenue or for identifying dissimilar groups.

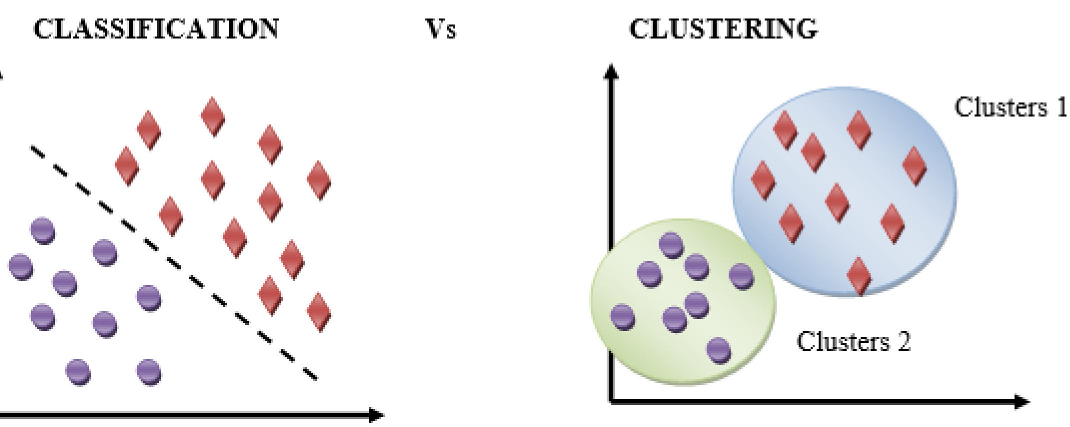


Figure 3.1: Difference between Classification and Clustering.

Generally clustering is mistaken for classification. So, to avoid this confusion elaboration of clustering and classification is done through figure 3.1. Clustering and classification are two different approaches as clustering is unsupervised learning which is also known as self-learning whereas classification is based on supervised learning which is reliable on pre-defined labels. Moreover, clustering works on un-labeled data because it does not need training sets whereas classification works with both labeled as well as un-labeled data. The comparison between classification and clustering is mentioned in table 3.1.

Table 3.1: Comparison between Clustering and Classification

Criteria	Classification	Clustering
Prior knowledge of classes	Yes	No
Use Case	Classify new sample into already known classes	Suggestion of groups based on patterns present in data
Algorithms	Bayesian classifiers, Decision Tree	Expectation Maximization, K-means
Data Needs	Labeled samples	Un-labeled samples
Prediction	Deals with prediction	Does not deals with prediction
Complexity	Less complex	More complex

There are basically two subgroups of clustering such as –

- Hard Clustering- In this every data point completely belongs to cluster or does not belongs to any cluster.
- Soft Clustering- In this clustering, it’s not likely to put every data point in distinct cluster despite that a likelihood of that data point is assigned to be in those clusters.

3.2 VARIOUS CLUSTERING MODES

There are various clustering models shown in figure 3.2 and are explained below -

- Connectivity Model- This model is also known as hierarchical clustering where the data points are much correlated to adjacent data point than those data points which are far away. In these algorithms data points are connected to make clusters depending on their distance. These models stick to two tactics: first one is that the categorization of every data point is done into different clusters and as the distance decreases they aggregate them and in second tactics every data point is categorize in one single cluster and as the distance increases they segregate them.

Moreover, the decision of distance function to increase or decrease is individual. All the connectivity-based algorithms only differ by the way of calculating distances and from the choice of linkage criterion as every cluster have multiple data points and various users to calculate distance. Some of most widely used linkage criterion is complete linkage, Un-weighted Pair Group Method with Arithmetic Mean (UPGMA) also acknowledged as average linkage clustering and single linkage clustering.

- Centroid Based Model- In this iterative clustering techniques are used in which idea of similarity is formed by closeness of the data point of the clusters to centroid. The most well-known algorithm k-means also known as Lloyd's algorithm is based upon this model. These algorithms discover only local optimum and operate on multiple iteration with discrete random initialization. Optimization in these algorithms could be done by- choosing multiple iteration that corresponds to k-medoids, by not picking initial centers randomly (K-means++), by selecting medians that is k-means clustering or by permitting a fuzzy cluster assignment that is fuzzy c-means clustering. Moreover, in k-means algorithms data space is divided into a Voronoi diagram-based structure and then it theoretically near to closest neighbor classification.

  Despite of being popular it has drawbacks for instance- it need number of cluster 'k' to be stated in advance and the size of cluster should be similar so that they could assign a data point to the closet centroid which result in incorrect cluster borders as this algorithm only focus on optimizing cluster centers but not cluster borders.
- Distribution Models- These models depend on the idea of all data point belongs to same distribution in a cluster such as gaussian or normal distribution. Moreover, clusters in this model are demonstrated by statistical distribution for instance expectation maximization technique uses multi-variate normal distribution. The drawback of these models is that they are over fitted.
- Density Models- In these models, the data space is explored for diverse density of data points. It also insulates numerous regions of density and in the same cluster data points are assigned within these regions only. OPTICS and DBSCAN are popular density model.
- Group Models- In these some algorithm only provides information about grouping and does not deliver a sophisticated model for results.
- Subspace Models- In these, clusters are demonstrated with relevant attributes and cluster members in co- clustering which is also known as two-mode clustering or bi-clustering.
- Graph Based Model- A node subset known as clique in graph that attaches every two nodes by an edge is taken into account as proto-typical cluster form. In HCS clustering technique the easing of connectivity requirement such as section of edges being lost, named as Quasi Cliques.

- Neural Models- These models are used to categorize as same to one or more above mentioned models and these are eminent un-supervised based neural network. When neural networks are executed from Independent Component Analysis (ICA) or Principal Component Analysis (PCA), they include sub-space models.

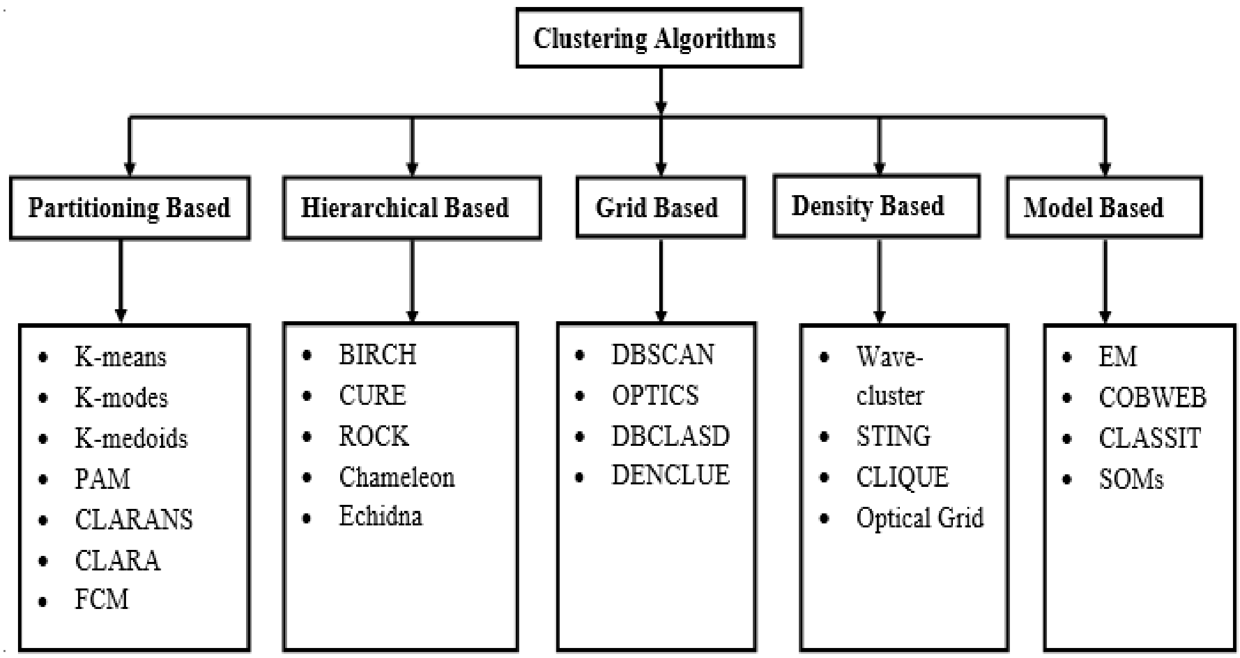


Figure 3.2: Type of Clustering Algorithms

3.3 METHODOLOGY

The proposed algorithm includes working with linear discriminant analysis and infinite feature selection for fuzzy c-means clustering. Both the above-mentioned approaches are quite efficient with respect to their performance. The linear discriminant analysis use for the classification of features from the datasets and after then with the help of infinite feature selection process, the best rank features are selected from the dataset and least rank features are neglected which would not have contributed in the performance of algorithms. The detailed explanation of these approaches is discussed in this section.

3.3.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is generally used in dimension reduction methods. The aim is to protrude data sets in low dimensional space and to evade over fitting with good class separability and diminish computational costs. The basic concept behind it is to discover linear transformation which could best differentiate between classes and then classification done in transformed space dependent on metric for instance Euclidean Square Distance (ESD) [38]. LDA have achieved great performance in application such as object and face recognition. Original LDA was defined as two classes problem and later also describe as multi-class version of Fisher's linear discriminant [38]. Two-class setting have demerit such as it was difficult for variable selection as compared to multi- class, because of the huge number of classes which does not provide virtuous separation for random-set of features. Whereas multi-class setting could over represent easily or abundant distinguishable classes in case of un-even distribution among clusters.

Table 3.2: Algorithm of Linear Discriminant Analysis

Procedure of Linear Discriminant Analysis
1. Calculate d-dimensional mean vectors for various classes from data set.
2. Calculate scatter matrix i.e. in-between-class & with-in-class scatter matrix.
3. Calculate the eigen-vectors $(v_1, v_2, \dots \dots \dots v_d)$ and their equivalent eigen-values $(e_1, e_2, \dots \dots \dots e_d)$for scatter matrix.
4. Categorized the eigen-vectors by declining eigen-values and then select k eigen-vectors with largest eigen-values to create $d \times k$ dimensional matrix E, where every column signifies an eigen-vector.
5. Exploit $d \times k$ dimensional eigen-vector matrix to convert the samples into new subspaces. This could be summarized by matrix multiplication$Y = H \times E$, where H is a $i \times d$ dimensional matrix representing the i samples, and j is the transformed $i \times k$ dimension samples in the new subspace.
6. After achieving the desired eigen vector matrix terminate the algorithm.

LDA when use for dimensionality reduction can be another preprocessing step for a pattern classification or machine learning applications.

For computing LDA five basic steps mention in Table 3.2 are discussed below-

1. Calculating d-dimensional mean vectors:

   For this, calculate mean vectors $g_y (y = 1,2,3)$ for say 3 different classes –

$$g_y = \begin{bmatrix} \mu_{\omega y} \\ \mu_{\omega y} \\ \mu_{\omega y} \end{bmatrix} \quad , with \;\; y = 1,2,3 \tag{3.1}$$

2. Calculating scatter matrices

   Calculate two matrices of dimension4×4, one in-between class and second with-in class scatter matrix.

- With-in class scatter matrix M_W

  The within class of scatter matrix can be expressed as –

$$M_W = \sum_{y=1}^{u} M_y \quad (3.2)$$

  where $M_y = \sum_{x \in P_y}^{i} (x - g_y)(x - g_y)^T$,that is scatter matrix for every class and g_y is the mean vector calculated earlier.

$$g_y = \frac{1}{i_y} \sum_{x \in P_y}^{i} x_k \quad (3.3)$$

  On other hand to calculate class co-variance matrix we could add a scaling factor $\frac{1}{L-1}$ to within class scatter matrix, so this could be expressed as –

$$\Sigma_y = \frac{1}{L-1} \sum_{x \in P_y}^{i} (x - g_y)(x - g_y)^T \quad (3.4)$$

  and

$$M_W = \sum_{y=1}^{u} \left(L_y - 1\right) \Sigma_y \quad (3.5)$$

  where L_y is represented as sample size of corresponding classes and when all classes are of same size, then drop to $\left(L_y - 1\right)$ term. So ensuring eigen-spaces would be identical.

- Between class scatter matrix $\mathrm{M_B}$

  The between class of scatter matrix can be expressed as –

$$M_B = \sum_{y=1}^{u} L_y (g_y - m)(g_y - m)^T \quad (3.6)$$

  where m is denoted as the over-all mean, L_y is size of corresponding classes and g_y is denoted as sample mean.

3. Solving the generalized eigen value difficulty for matrix $M_B M_W^{-1}$:

   To obtain linear discriminants, solve the generalized eigen value difficulty for matrix $M_B M_W^{-1}$.

4. Select linear discriminants for novel feature sub-space:

   For this there are two step-first is to sort eigen-vector by reducing eigen-values and second is choosing k eigen-vector with largest eigen-values.

- Sort eigen-vectors by reducing eigen-values

To select which eigenvector is appropriate for low dimension sub-spaces, look at corresponding eigen-values of the eigen-vectors. The eigen-vector with lower eigen-values accept the least information about distribution of data and those are the values to be dropped. The general method is to rank the eigen-vectors according to highest to lowest values corresponding to eigen-values and then select the top k eigen-vector.

- Choosing k eigen-vector with largest eigen-values

  After the sorting of eigen-vector by reducing eigen-values is completed, then its times to contrast an eigen-vector matrix of dimensions $k \times d$and thus reducing high dimensional feature space in lower dimensional subspaces.

5. Transforming samples in novel subspace

   In this step, $k \times d$ dimensional matrix E that has been calculated to transforms samples on novel subspace which is done by below mentioned equation –

$$Y = H \times E \tag{3.7}$$

   where $H$is denoted as $i \times d$ dimension matrix which represent the$i$ samples and Y denotes transformed samples of matrix of dimensions $i \times k$ in new sub-space.

6. After achieving the desired eigen vector matrix terminate the algorithm.

3.3.2 Infinite Feature Selection

Infinite feature selection a is filter-type algorithm as shown in figure 3.3, which execute ranking step in an un-supervised way and then tracked by simple cross validation tactic for choosing the finest features. This approach estimates the significance of given feature, whereas taking in account every possible feature subset and this is considered as most interesting characteristic of this approach [39]. Moreover, IFS method is also robust by means of few training data sets, have complexity which is competitive and accomplish effectively in high ranking in most pertinent feature.

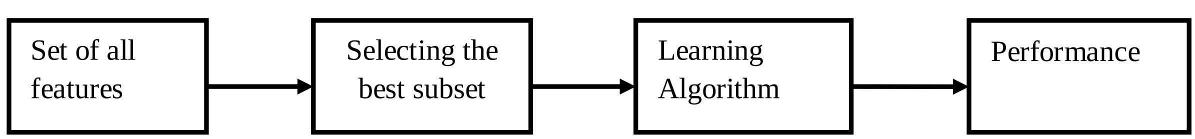


Figure 3.3: Filter based Infinite Feature Selection Method

Consider a feature distribution $F_e = \{f^1, \dots\dots, f^n\}$and $z \in R$ which denote a sample of generic distribution f. A fully linked graph is represented as $G_r = (V_e, E_d)$, where V_e denotes the set of vertices equivalent one-by-one to every feature distribution and E_d is the weighted edges that model pair-wise relations between feature distributions.

G_r is represented as an adjacent matrix named as A_d and stipulate the class of weighted edges as every element i.e. a_{ij} of matrix A_d, $(\mathrm{i} \geq 1, \mathrm{j} \geq \mathrm{n})$ denotes a pair-wise energy term. The energies are expressed as a linear weighted combination of 2 simple pair-wise calculates linking f^j and f^i, expressed as –

$$a_{ij} = \sigma_{ij}\alpha + c_{ij}(1 - \alpha) \tag{3.8}$$

where α is represented as a loading coefficient and α ∈ [0,1] and σ_{ij} is given as –

$$\sigma_{ij} = \max(\sigma^i, \sigma^j) \tag{3.9}$$

where σ^i is the standard deviation in samples z ∈ f^i and another term is given as –

$$c_{ij} = 1 - \left| spearman\left(f^i, f^i\right)\right| \tag{3.10}$$

along with spearman demonstrating Spearman's rank cor-relation coefficient. The a_{ij} joins two feature distributions on the basis of the maximal feature-dispersion along with their cor-relation.

When pair-wise analysis of features is performed, then the energy related to sets bigger than two feature distributions is individuated. Suppose $\rho = \{v_0 = i, v_s = j, v_{(1)}, \dots . v_{(s-1)}\}$ which denotes the path of length s in between i and j vertices i.e. the $f^i$and f^i features $v_{(1)}, \dots . v_{(s-1)}$. For ease, let length of path s is lower as compared to the total quantity of features n and also the path s does not have cycles therefore features are not visited more than one time [40].

A path could be defined as a sub-set of existing features that gets involved. The energy of ρ could be define as –

$$\mathcal{E}_\rho = \prod_{k=0}^{s-1} a_{v_k, v_{k+1}} \tag{3.11}$$

here $\mathcal{E}_\rho$ are pair-wise energies of every feature pair that comprise the path. Moreover, it is supposed as the joint energy of feature sub-set. Now $P_{i,j}^s$ set is defined as comprising all the paths of length s among i and j, to explain the energy of every paths of length s and then sum up them as defined below –

$$R_s(i,j) = \sum_{\rho \in P_{i,j}^s} \mathcal{E}_\rho \tag{3.12}$$

The standard matrix algebra provides –

$$R_s(i,j) = I^s(i,j) \tag{3.13}$$

This is the power iteration of I. The R_s contains cycles with respect to feature selection and it's similar as if only a solo feature is taken into account more than one time, possibly related to self-cycle. If the path length is extended to infinity, then the likelihood of being a part of the cycle is unvarying for every feature and is consider by constructing R_s that when normalization gets involved.

Table 3.3- Algorithm of Infinite Feature Selection

Infinite Feature Selection
Input- $F_e = (f^1, \ldots\ldots., f^n), \alpha$
Output- q energy scores for every feature constructing graph
for i = 1: n
for j= 1: n
$\sigma_{ij} = max\ (std(f^i), std\ (f^j))$
$c_{ij} = 1 - \left\|spearman\left(f^i, f^i\right)\right\|$
$a_{ij} = \sigma_{ij}\alpha + c_{ij}(1 - \alpha)$
end
end
paths approach to infinity
$\mathrm{r} = \frac{0.9}{\rho(A)}$
$\mathrm{K} = (\mathrm{I} - \mathrm{r\ A})^{-1} - I$
$\mathrm{k} = \mathrm{K\ e}$
return K

Evaluation of a single-feature energy score of the feature f^i at any given path length s is done as –

$$q_s(i) = \sum_{j \in v} R_s(i, j) = \sum_{j \in v} A^s(i, j) \quad (3.14)$$

Above equation 3.14, models the characteristics of f^i feature in every type of n feature selection. More the value of $q_s(i)$, more would be the energy of the i^{th} feature. Basic impression of feature selection can be considered as the decreasing of features by q_s value by first taking m for the procurement of an effective and non-redundant set. Although the computational cost of q_s is quite high ($\mathcal{O}(\mathrm{s} - 1) \cdot n^3$)).
The order of s is similar to the order of n, therefore the computation becomes $\mathcal{O}(n^4)$, though become impractical for using large feature sets. To deal with this issue the infinite feature selection method expands

the length of path to infinity i.e. $s \rightarrow \infty$ and it uses algebra concepts to make simpler calculation for infinite circumstance. Table 3.3 shows the procedure for infinite feature selection method.

3.3.3 Fuzzy C-Means Clustering

Fuzzy C-Means (FCM) algorithm is based on k-means idea of portioning of data set into clusters. FCM is a soft clustering scheme with degree-of-belief data point is allocated to clusters. Moreover, in FCM algorithms the one data point could be connected multiple clusters with various degree of belief. In this, it tries to discover the characteristic point in every cluster which is called as the center of the one cluster and after that it calculates the membership-degree for every data point in the cluster. These algorithms also diminish intra cluster variance [41].

Unlike k-means algorithm where one data point is assign to one cluster only, the FCM algorithms assign one data point to multiple clusters but both algorithms iteratively explore the center of cluster and update the data point memberships. FCM algorithm allocates a value ranging from $0 - 1$ to a data point which determines the probability by which data point belongs to the cluster instead of taking a firm verdict about the belonging of pixel to which cluster [23]. According to the fuzzy rule, the sum of membership value of the pixel to every cluster should be 1. More the value of membership, more probable a pixel would fit to that cluster.

FCM clustering is gained by minimizing the cost function as shown below –

$$L = \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik}^{m} \left| b_i - x_k \right|^2 \tag{3.15}$$

where, L– the objective function,

$c -$ number of clusters,

$n -$ number of objects,

$k -$ the cluster in which μ_{ik} is likelihood value that is assign to i object,

$m -$ fuzziness factor having value 1, and

$\left| b_i - x_k \right| -$ Euclidean distance among i^{th} object b_i and k^{th} cluster centerx_k

The Euclidean distance could be defined as –

$$\left| b_i - x_k \right| = \sqrt{\sum_{i=1}^{n}(b_i - x_k)} \tag{3.16}$$

Centroid of k^{th} cluster is updated by using equation (3.15) –

$$x_k = \frac{\sum_{i=1}^{n} \mu_{ik}^{m} b_i}{\sum_{i=1}^{n} \mu_{ik}^{m}} \tag{3.17}$$

To calculate fuzzy membership by using original equation (3.15) –

$$\mu_{ik} = \frac{1}{\sum_{l=1}^{c}\left(\frac{|b_i - x_k|}{|b_i - x_k|}\right)^{\frac{2}{m-1}}} \tag{3.18}$$

Table 3.4: Fuzzy C-Means

Fuzzy C-Means
1. Set up the number of cluster c
2. Select initial cluster proto-type i.e. $v_1, v_2, \ldots.. v_c$ from b_i where i = 1,2, …… , n
3. Calculate the euclidean distance $\lvert b_i - x_k \rvert$ between object b_i and proto-type x_k
4. Calculate number of elements of fuzzy partition matrix i = 1,2, …… , n k = 1,2, …… , c $\mu_{ik} = \frac{1}{\sum_{l=1}^{c}\left(\frac{\lvert b_i - x_k\rvert}{\lvert b_i - x_k\rvert}\right)^{\frac{2}{m-1}}}$
5. Calculate cluster centroids k = 1,2, …… , c $x_k = \frac{\sum_{i=1}^{n} \mu_{ik}^{m} b_i}{\sum_{i=1}^{n} \mu_{ik}^{m}}$
6. Stop (if convergence is achieved or the number of iteration is exceeded else go back to 3rd step)

FCM algorithm is extended to cluster RGB images in which distance calculation is followed by equation (3.16), which is further modified as-

$$|b_i - x_k| = \sqrt{\sum_{i=1}^{n}(b_{iR} - x_{kR})^2 + (b_{iG} - x_{kG})^2 + (b_{iB} - x_{kB})^2} \tag{3.19}$$

The k-means algorithm has many disadvantages such as outliner sensitivity, describing number of clusters, memory space, un-known number of iterative steps for cluster and local optima vulnerability.

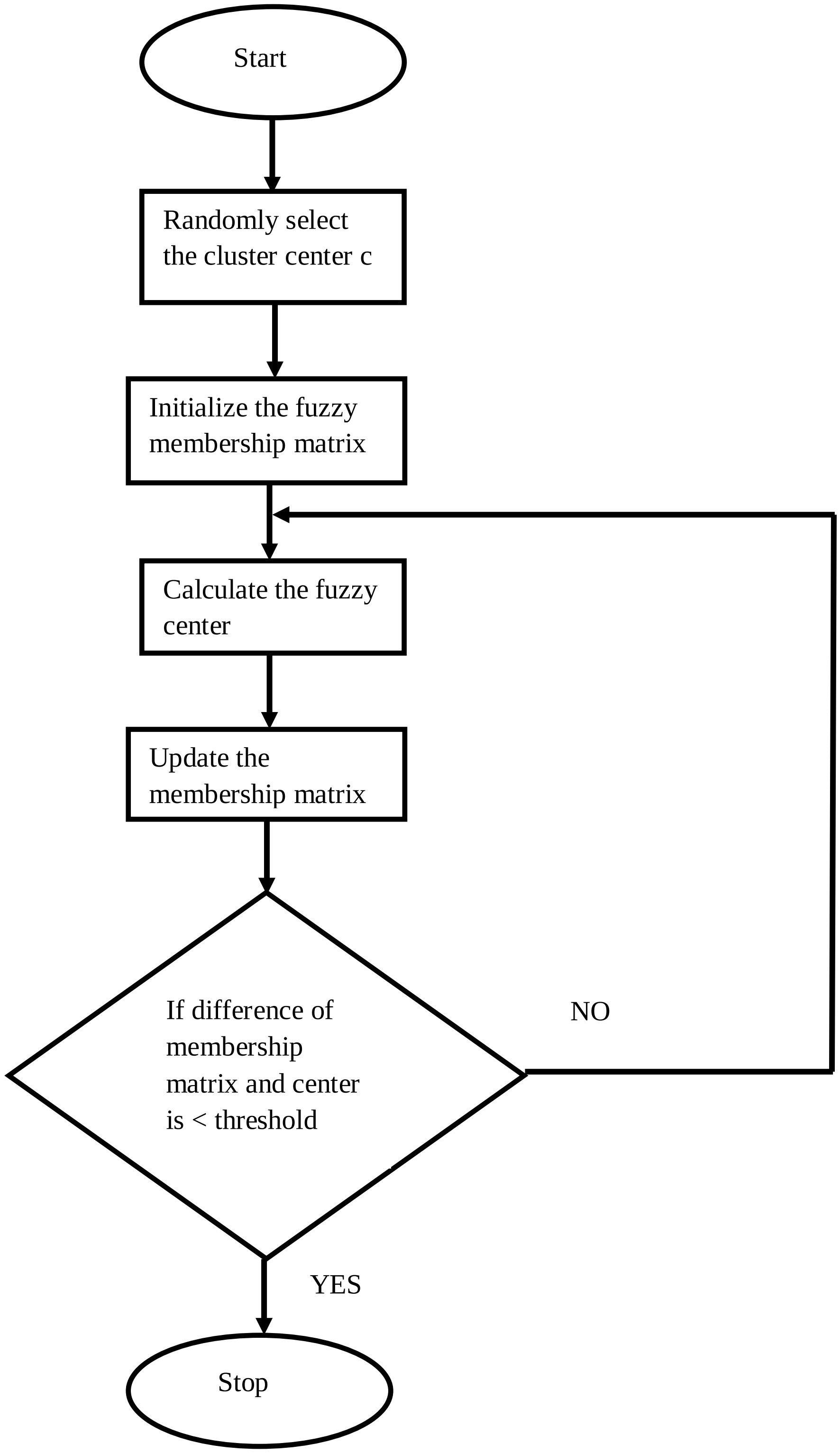


Figure 3.4: Flow chart of Fuzzy C-Means algorithm

On other hand fuzzy c-means algorithms are much more capable of handling the problems regarding the ability of patterns, interaction of human, mixed information of media, noisy or incomplete data and provide fast approximation of the solution. The procedure of FCM algorithm is described in table 3.4.

Moreover, the time complexity in FCM are more as compared to k-means algorithms thus they work faster than FCM algorithms [42]. These algorithms are used in applications such as image retrieval, functional dependencies and discovering of association rules. The flow chart of the FCM algorithm is shown in the figure 3.4.

Partition-based algorithms have many advantages such as – they are appropriate for compact spherical clusters-based datasets and are quite simple as well as scalable algorithms. Apart from merits these algorithms have some demerits too, for instance – they are highly sensitive to noise, outliners and initialization phase, they comprise of poor cluster descriptors, incapable in dealing of convex clusters of various density and size and are dependent on users for specification of quantity of clusters in advance [43].

3.4 FLOW CHART OF PROPOSED ALGORITHM

The procedure of the proposed algorithm is represented in table 3.5, each and every step done during computation of proposed work in Matlab code is described in mathematical form. Moreover, the simple step involved in proposed work is described in form of flow chart for better understanding of work in figure 3.5 and also all different steps are briefly explained below.

The steps mentioned in the flow chart are briefly discussed below-

Step 1: Upload the images from particular data set.

Step 2: To have every image in data set of similar size, the resizing of the images is done.

Step 3: To extracts the feature from the dataset the linear discriminant analysis is used.

Step 4: Now use infinite feature selection approach to represent the feature in rank form and then assign weight to them and choose the best ranked value features from the data set to decrease computational time.

Step 5: Apply the fuzzy c-means clustering algorithm to those ranked features extracted in pervious step to group the similar data points into clusters and it assign membership on the basis of distance between data point and cluster center.

Step 6: Evaluate the performance matrix for proposed algorithms and plot the obtained results.

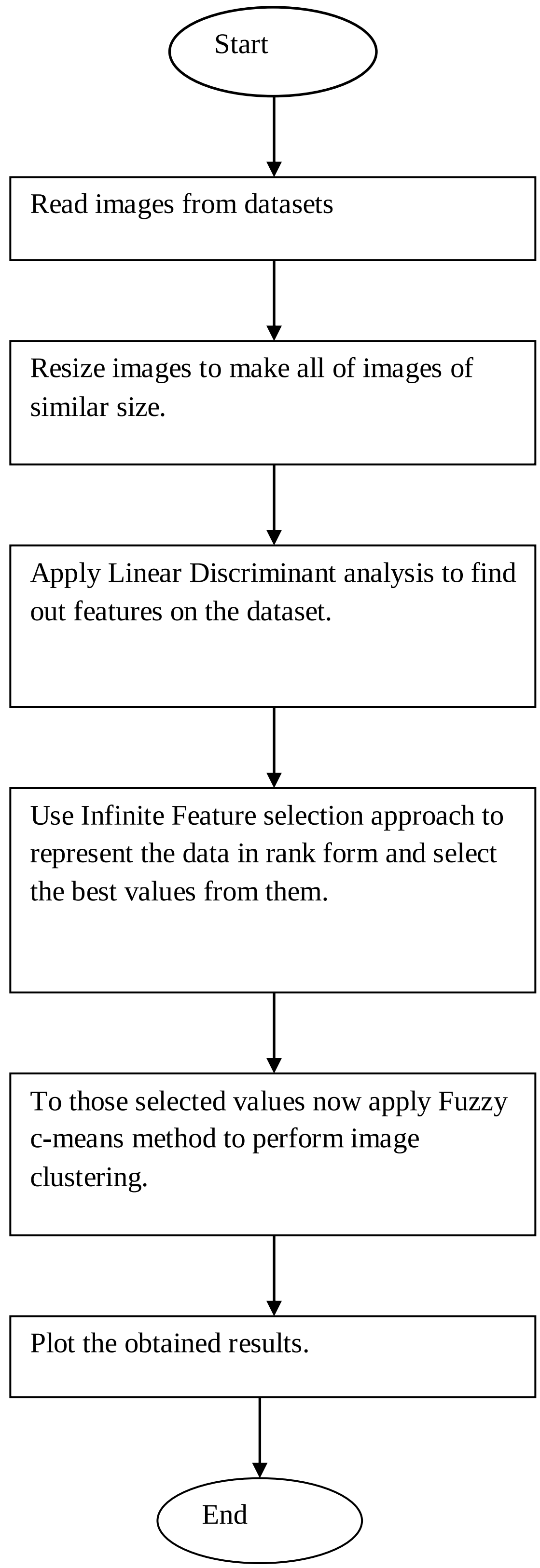


Figure 3.5: Flow chart of proposed algorithm.

Table 3.5: Algorithm for proposed algorithm

Procedure for proposed algorithm
Input-X ← images from dataset
1. **Begin**
2. Load the images from dataset
3. Find linear combination among various features using linear discriminant analysis $$Y \leftarrow \text{LDA}(X)$$
4. Represent the data in rank form and select the best values from them. By using infinite feature selection $$[R, W] \leftarrow IFS\ (Y, \alpha)$$
5. Select the best rank 10 values $$Y \leftarrow Y\big(:, R(1{:}10)\big)$$
6. Calculate the mean of the selected value $$M \leftarrow \text{mean}(Y)$$
7. Assign membership to each data-point corresponding to each center of cluster depending on the distance among cluster's center and the data point by using fuzzy c-means method $$(I, CI) \leftarrow FCM\ (M, k, t_n)$$
8. Plot the obtained results
9. **End**

CHAPTER 4
RESULTS AND DISCUSSION

4.1 INTRODUCTION

The proposed method can be used for efficiently clustering-based applications. In this dissertation, the proposed algorithm is tested to accomplish image clustering and proved the effectiveness of the proposed algorithm. The experiment results and settings are described in the subsequent section below.

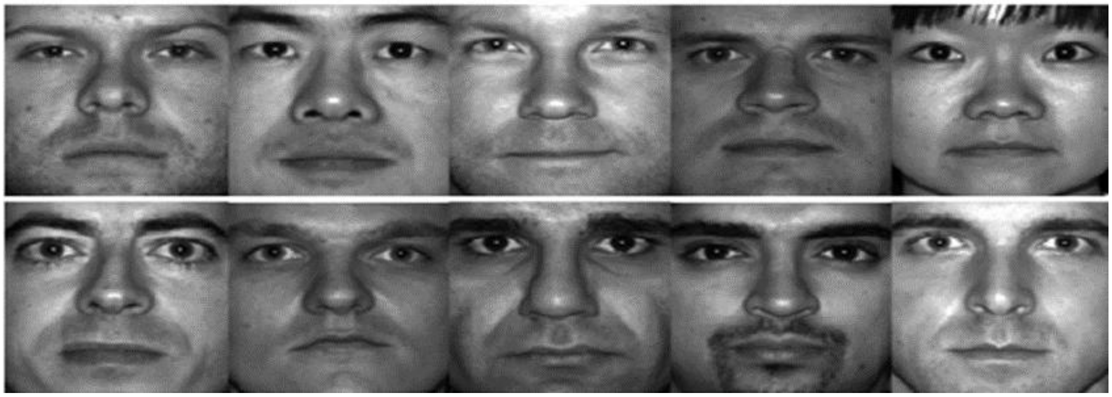
4.1.1 Data Sets

To calculate the efficiency of the proposed algorithm 3 databases has been used as mentioned in table 4.1. Some of the samples of databases are shown in figure 4.1. Detailed description of these databases is given next.

- YALE B Dataset-
  In this database there are 64 gray-scale images comprising of faces and there are 38 folks under diverse illumination and lighting conditions [47].
- FEI Dataset-
  In this data base consists of 700 RGB images from 50 folks and every folk has 14 images taken from large range of view [48].
- ORL Dataset-
  This database commonly used for face recognition application. It contains 10 gray-scale images from 40 folks under different condition such as lighting and it consists both with and without glasses along with various facial expressions of folks [49].

Table 4.1: Dataset Description

Database	Size (n)	Dimensionality	Number of Classes
Yale B	2414	1024	38
FEI	700	768	50
ORL	400	1024	40



(a)

(b)

(c)

Figure 4.1: Illustrate sample images from 3 datasets. (a) Yale B Dataset, (b) FEI dataset and (c) ORL dataset.

In figure 4.1, only the small portion of the 3 datasets is represented as sample image. It provides the idea about the images present in FEI, ORL and Yale-B datasets. In the figure 4.1 (a) the 10 sample images from Yale-B dataset is represented. In the figure 4.1 (b) the 14 sample images from FEI dataset is represented. In the figure 4.1 (c) the 10 sample images from ORL dataset is represented.

4.1.2 Evaluation Matrices

In the following section, clustering accuracy and sparseness is used for evaluating the performance of the clustering.

- Accuracy

  The objective of accuracy is to discover the one to one relationship among clusters and classes. Moreover, clustering accuracy compute the proportion of suitably classified data-points of clustering solution competed to predefined class-labels [23], which could be define as –

$$A_c = \frac{\sum_{i=1}^{n} \delta(map(u_i), g_i)}{n} \tag{4.1}$$

  where n – scale of data-set,

  $g_i -$ the truth labels offered by data set,

  $\delta(m, n) -$ the delta function, $\delta(m, n) = 1$ when $m = n$ and $\delta(m, n) = 0$ otherwise,

  $u_i -$ the cluster label, and

  $map(u_i) -$ the optimal permutation mapping function which maps the labels gained by given algorithm to the ground labels given by data-set.

- Sparseness Study

  To test the sparseness of our algorithm we used Hoyer's sparseness test [50]. The sparseness could be defined as –

$$S_p = \frac{\sqrt{n} - (\sum|x_i|)/\sqrt{\sum x_i^2}}{\sqrt{n} - 1}$$

(4.2)

  where n can define as the dimensionality of basis matrix. Moreover, the sparseness measure used here is dependent on the relationship among L_1 and L_2 norm. Table 4.5 shows the mean sparseness of the algorithms. The sparseness measure of various algorithms is shown in figure 4.5.

4.1.3 Comparative Algorithms

The proposed algorithm is compared with various other existing algorithms to prove its effectiveness. All the techniques whose performance comparison is done with proposed method are mentioned below-

- PCA- Principal Component Analysis is most commonly used method for data representation method. Moreover, in this method we could effectively extract information from raw data.
- K-means- Another popular clustering algorithm in which there is requirement of data points that would be clustered and the value of k i.e. the amount of clusters that algorithm need to produce.
- FCM- In this number of clusters has to be prespecified and then it give back the membership value and cluster centers for every data-point. It is an iterative process, dependent on minimization of cost function which describes the distance among the cluster center and data point [41].
- LDA- It is a linear transformation scheme used for dimension reduction commonly. The LDA is supervised algorithm which would indicate axes that increase the separation among various classes [38].
- IFS- It is a filter-type algorithm which execute ranking step in an un-supervised way and then tracked by simple cross validation tactic for choosing the finest features [39].

In figure 4.2, the one LDA image of all the 3 datasets each are represented respectively. The LDA images in figure 4.2 represent the values of ones and zeros present in the images. In this the proposed algorithm displays the one LDA image of last image present in the dataset. The sequence number of the original image is mentioned in the figure as it would not be possible to represent all the images from the 3 datasets in the results.

Moreover, LDA is used to protrude datasets in low dimensional space to evade the over fitting with good class separability. In figure 4.2 (a), illustration of LDA image is done for Yale-B dataset. In figure 4.2 (b), for ORL dataset illustration of LDA image is given. The illustration of LDA image is given for FEI dataset in figure 4.2 (c).

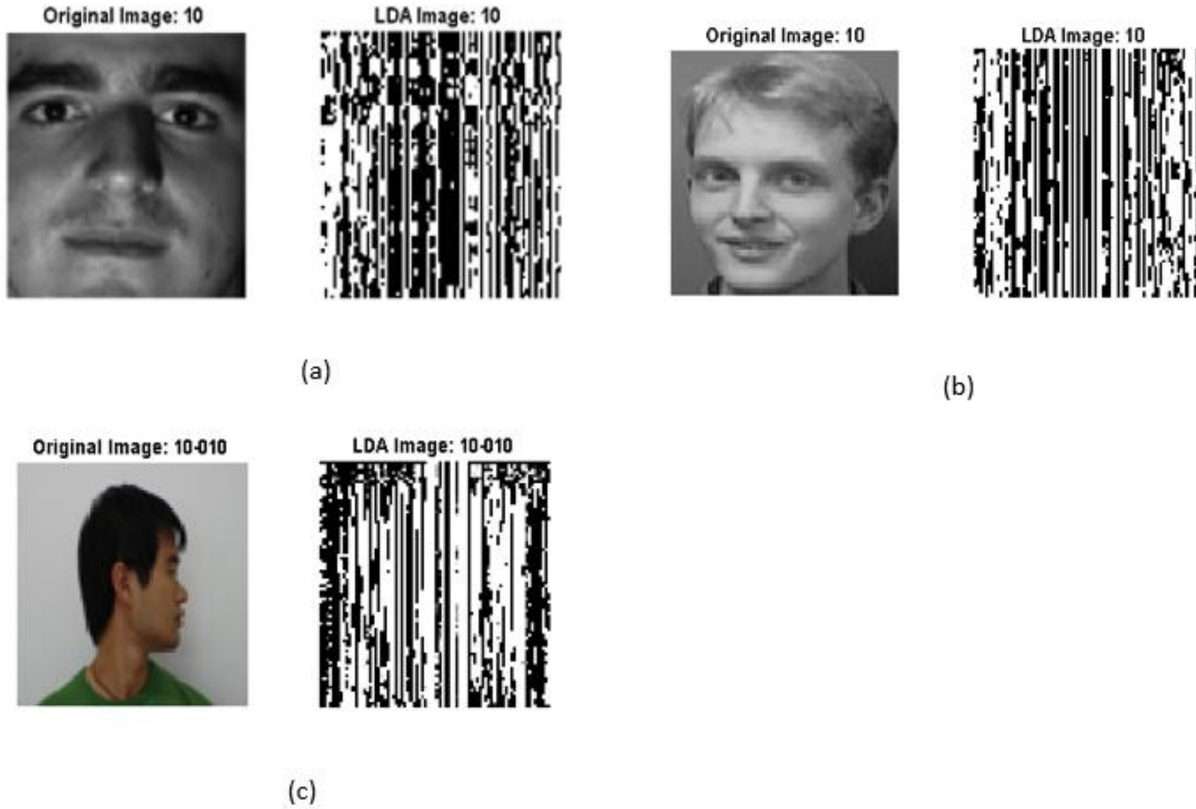


Figure 4.2: Illustration of LDA images of all the datasets. (a) YaleB Dataset, (b) ORL Dataset, (c) FEI Dataset.

4.2 CLUSTERING PERFORMANCE COMPARISON

Various parameters are used for clustering comparison of proposed algorithm and the corresponding results are also shown and briefly discussed.

4.2.1 Selection of Parameters

The selected parameters for comparison are, the rank r, the rank index τ and sparseness degree.

- Rank Index-

  Rank index can be defined as the similarity measure between two data clustering. The adjusted rank index can be defined as the form of rank index in that adjusted for the chance grouping of elements. The rank index is also considered to be related with accuracy and even applicable when class labels are not used [46].

For convenience we define rank index τ as

$$\tau = \frac{r}{min\,(f,n)} \tag{4.3}$$

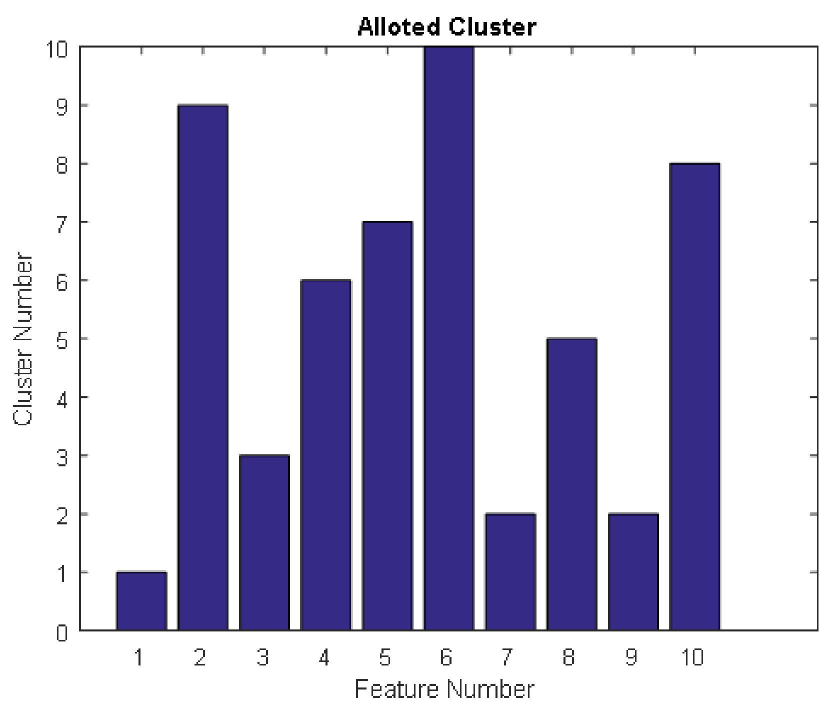
where r- represent the rank,

n -represent the number of samples,

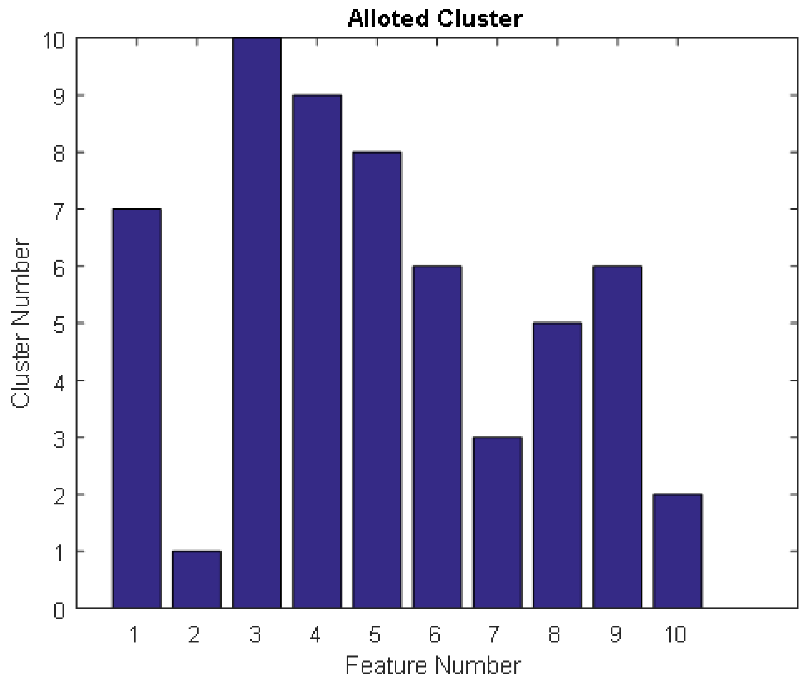
f-represent the dimension of features.

The equation (4.3) defines the mathematical term for rank index, which is used as a parameter for calculating the accuracy of proposed algorithm. The accuracy of proposed algorithm is determined for 3 different datasets.

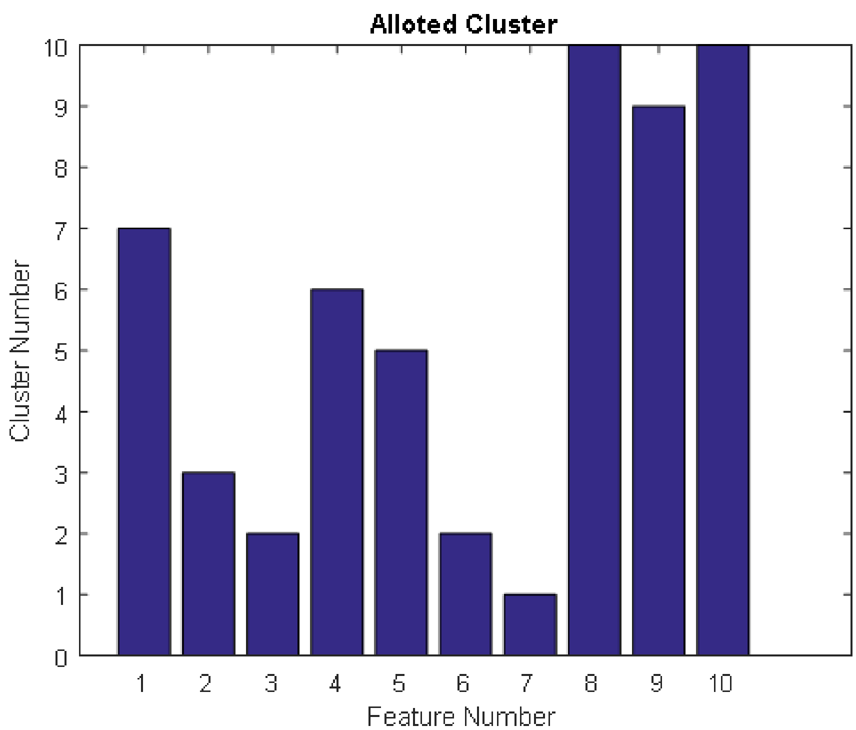
Figure 4.3 illustrate the number of clusters allocated to all three datasets. Here, the figure 4.3 illustrates the top ranked 10 features those are selected by using infinite feature selection. Corresponding to those feature the proposed algorithm represent the number of cluster allocated. In proposed work, the 10 highest ranking features are selected. The selection of top ranking features from the dataset can be increased or decreased according the requirement. Similarly, with the change in the number of features selected the allotted cluster number would also be affected.



(a)



(b)



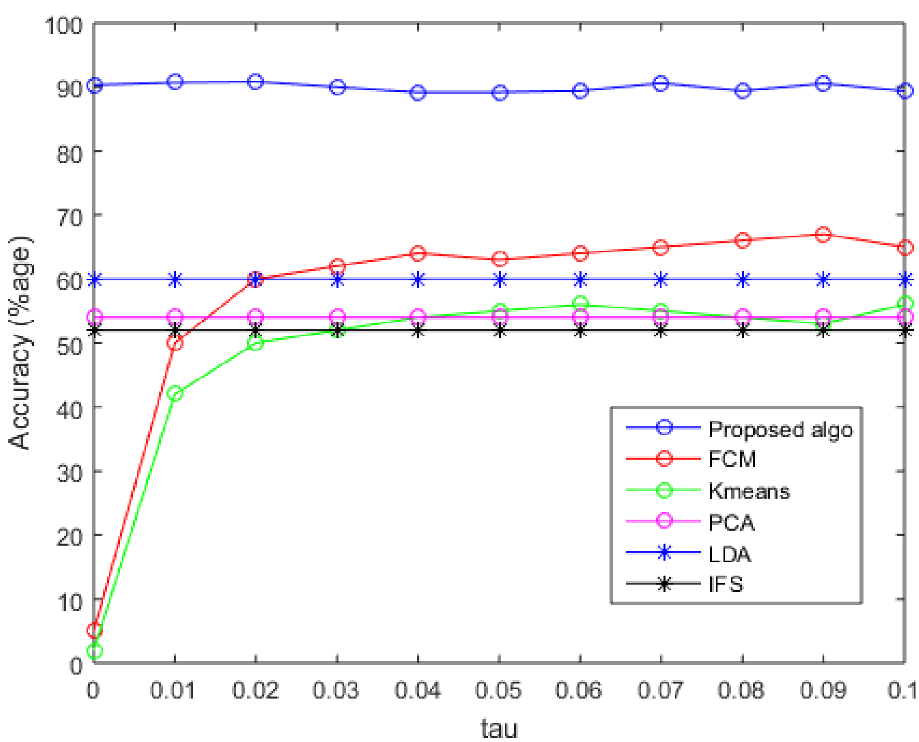
(c)

Figure 4.3: Illustration of the allotted cluster from the three datasets. (a) FEI dataset, (b) YaleB dataset, (c) ORL dataset.

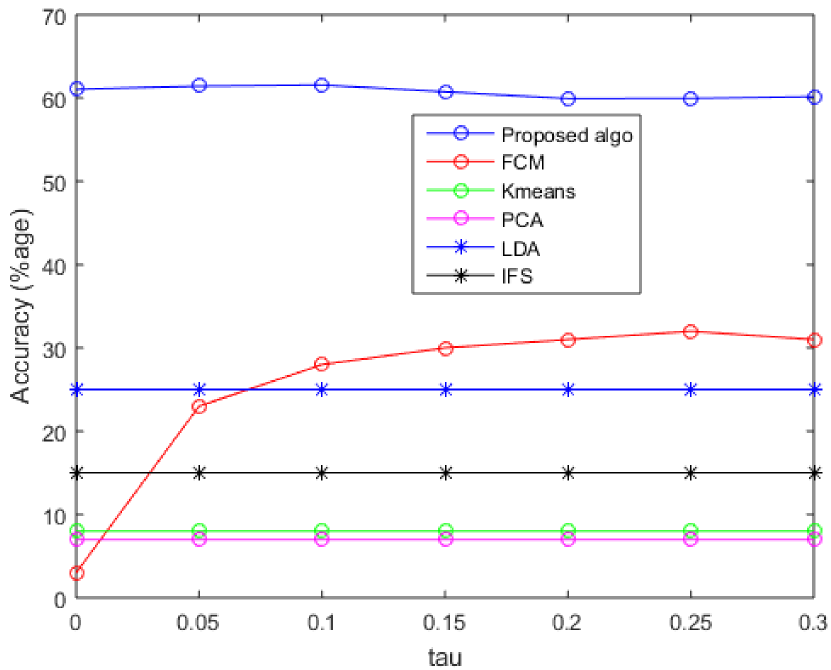
Figure 4.4 illustrate the performance variation of the various algorithms with respect to parametersτ. Moreover, all the parameters are experimentally tested on YaleB, FEI, and ORL databases.

For rank index parameter- In figure 4.4 the proposed algorithm performs quite well as compared to other algorithms. As the value of τ increase, the proposed algorithm provides stable outcome for all 3 datasets. While in figure 4.4 (a) the value of FCM and K-means algorithm performs poorly when value of τ is lower than 0.05. But as the value of τ increases, these algorithms get steady. When τ is too small the algorithm performs poorly. It is due to the reason that, when the rank of dataset is set too low, then the structure of data can't be recovered properly.

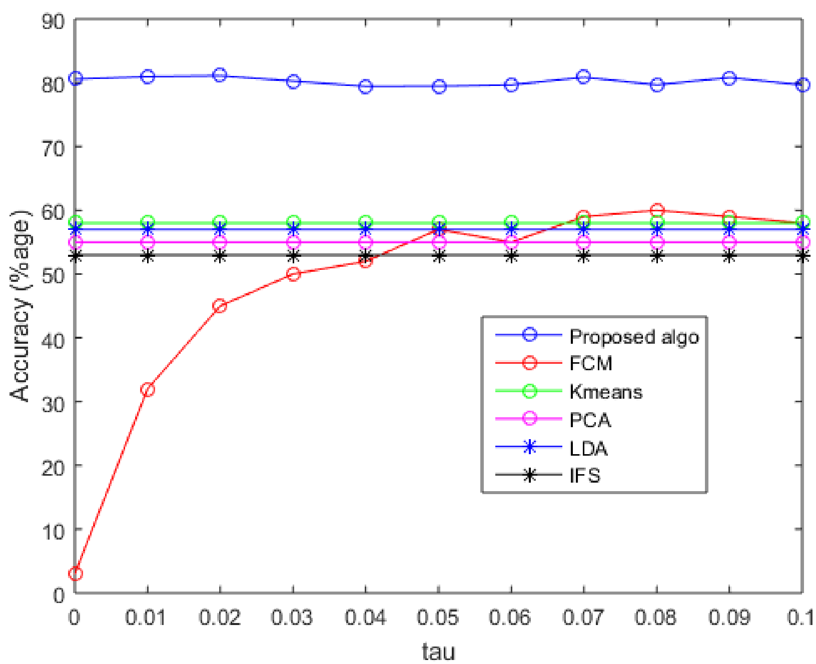
In figure 4.4 (b), proposed algorithm attains accuracy of 61% in YaleB dataset whereas FEI dataset provides highest accuracy of proposed algorithm of 90% with the same τ parameter. For convenience, set the value of $\tau = 0.10$ for proposed algorithm in testing for FEI and ORL database as shown in figure 4.4(c) and set the value $\tau = 3.0$ in the YaleB dataset as shown in figure 4.4 (b).



(a)



(b)



(c)

Figure 4.4: Illustration of performance of various methods versus parameter τ on three datasets. (a) FEI Dataset, (b) YaleB dataset, (c) ORL dataset

Table 4.2: Clustering performance for τ parameter on FEI, ORL and Yale-B datasets.

Datasets	ACCURARCY (%)					
	Algorithms					
	k-means	PCA	IFS	LDA	FCM	Proposed
FEI	02.00	54.00	52.00	60.00	05.00	90.29
ORL	58.00	55.00	53.00	57.00	03.00	80.01
YALE-B	08.00	07.00	15.00	25.00	32.00	61.05

4.2.2 Comparisons-

In this section, the comparison is made on the basis of the performance of proposed algorithm with various existing algorithms such as K-means, PCA, LDA, IFS, FCM for the three image datasets on the basis of clustering accuracy, sparseness measure and time complexity. Time complexity is discussed later in this section.

- Accuracy- The clustering results on the dataset FEI, ORL and Yale-B are shown in table 4.2. From table 4.2, it is observed that the proposed algorithm shows the finest performance and outperforms other algorithms by 10 -15% on average. The proposed algorithm gives higher accuracy rate in FEI dataset for as shown in table 4.2 as compared to other two datasets. In summary, the proposed algorithm is effective and out-performs other above-mentioned algorithms.

Table 4.3: Sparseness measure of various algorithms

Datasets	SPARSENESS					
	Algorithms					
	k-means	PCA	IFS	LDA	FCM	Proposed
FEI	0.980	0.460	0.480	0.400	0.950	0.097
ORL	0.420	0.450	0.470	0.430	0.970	0.194
YALE-B	0.920	0.930	0.850	0.750	0.970	0.389

- Sparseness measure- As it could be seen in table 4.3, the proposed algorithms have the lowest sparseness value among all the other algorithms. For FEI dataset the sparseness value of k-means and FCM algorithm is quite competitive.

The proposed algorithm provides lowest sparseness measure in FEI dataset in comparison to other two datasets that is of 0.097. Whereas, in Yale B dataset it provide the highest value as compared to other two dataset that is value of 0.389. The sparseness measure of various algorithms is shown in figure 4.5.

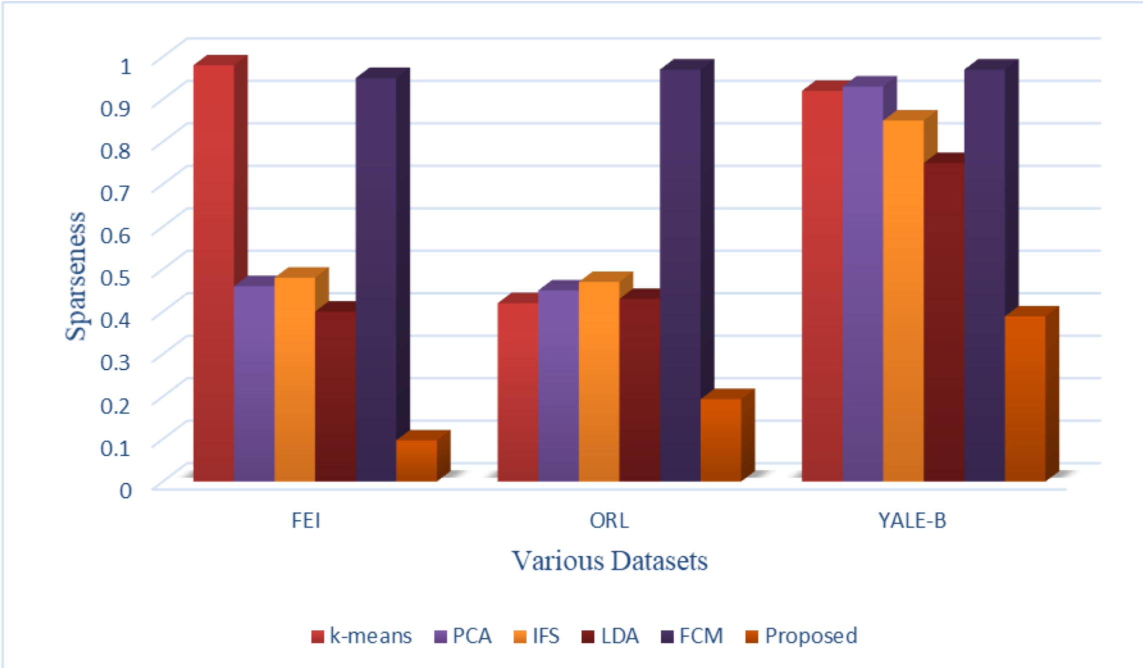


Figure 4.5: Sparseness measure of various algorithms

4.3 TIME COMPLEXITY

Time Complexity is used to define the total amount of time required by an algorithm or program to run. It is generally analyses by the amount of elementary steps accomplished by algorithm to complete the execution. A constant factor is considered to be the difference among the elementary steps number performed and the total time taken by the algorithm. The worst-case scenario time complexity is considered commonly in which the execution time of program is maximum. The average case scenario time complexity is less common, and it is the average time taken by program on provided size of inputs. The time complexity in these two cases is usually represent as the function of input size [51]. The big O notation for instance $\mathcal{O}(\mathrm{n}), \mathcal{O}(n^a), \mathcal{O}(2^n),\ \mathcal{O}(n\,log\,n)$ are the commonly used expression for time complexity of algorithms. It is considered as asymptotic notation for time complexity.

4.3.1 Various Notations for Time Complexity

There are five generally used notations for time complexity represented below-

- Big O- It represents the same as or the fewer than (expression) iterations.
- Big Theta- It represents the same as (expression) iterations.

- Big Omega- It represents the same as or more than (expression) iterations.
- Little Omega- It represents more than (expression) iterations.
- Little O- It represents fewer than (expression) iterations.

The complexities of algorithms are categorized with respect to the kind of function emerged in big O notation. As exemplar, $\mathcal{O}(n)$ time complexity-based algorithm is considered as linear-time algorithm and $\mathcal{O}(n^a)$ time complexity algorithm is an algorithm for polynomial time for constant $a > 1$.

4.3.2 Explanation of Notations of Time Complexity

The O, Omega and Theta notation are explained below in details.

- O Notation-In this the set of functions would grow either at the same rate or at the rate slower than the expression. The O notation symbolizes the worst-case of time complexity for an algorithm. It also specifies the maximum time requirement of algorithm for each value of input.
- Omega Notation-This set of functions grow either at the same rate or at the faster rate than the expression. It symbolizes the best-case of the time complexity for an algorithm. It also specifies the minimum time requirement of algorithm for each value of input.
- Theta Notation-This contains of both O notation functions and Omega notation functions. It symbolizes the average-case of the time complexity for an algorithm. It also specifies the average algorithm bound.

4.3.3 Common Time Complexities

There are various types of time complexities. Some of the most common time complexities are explained below.

- Constant Time- In this the running time does not need to be independent of size of problem but the in case of upper bound, it has to be independently bounded to problem size. In this $t(n)$ or $\mathcal{O}(1)$ time is independently bounded to the input size and it is consider as a linear time operation which takes $\mathcal{O}(n)$ time.
- Logarithmic Time- If $t(n) = \mathcal{O}(log\, n)$, only then the algorithm would take logarithmic time. The algorithms based on logarithmic time are generally used in binary search or in binary tree

operations. It provide great efficiency as in this the number of operations ratio to input size decreases, and when value of $n$increase it tends to become zero.

- Polylogarithmic Time- If $t(n) = \mathcal{O}((\log n)^k)$for constant k value, only then algorithm is considered to execute polylogarithmic time [52].
- Sub-linear Time- If $t(n) = o(n)$, then an algorithm would execute sub linear time. It uses non-classical processing, parallel processing or pledged assumptions of the structure of the input [53].
- Linear Time- If time complexity is $\mathcal{O}(n)$, then algorithm is considered to execute linear time. When algorithm has to read the complete input values, this is considered to be the best time complexity possible.
- Quasilinear Time- If time complexity $t(n) = \mathcal{O}(n \log^k n)$, then algorithm is considered to execute quasilinear time for non-negative k constant. These time complexity algorithms are $\mathcal{O}(n^{1-\varepsilon})$ for all $\varepsilon > 0$ constant. Therefore, these algorithms provide faster execution in comparison to polynomial time complexity algorithm [54].
- Sub quadratic Time- If time complexity $t(n) = o(n^2)$, then algorithm is considered to execute sub-quadratic time. The comparison dependent sorting algorithms are quadratic time complexity for instance insertion sort, whereas advanced algorithms are sub-quadratic time complexity for instance shell sort.
- Polynomial Time- If time complexity $t(n) = \mathcal{O}(n^k)$, then algorithm is considered to execute polynomial time for non-negative k constant and its running time is upper bounded with polynomial expression in input size [51].
- Super polynomial Time- If time complexity $t(n) = \omega(n^c)$, then algorithm is considered to execute super-polynomial time. In this running time is not bounded to any polynomial. Here n is the input parameter and c is the constant.

4.3.4 Time Complexity of the Proposed Algorithm

The comparison of time complexity of proposed algorithm is done with other algorithms like k-means, PCA, LDA, IFS and FCM algorithms.

The FCM clustering algorithm used a dataset and number of clusters is to be pre-specified and then it returns the membership value and cluster centers for every data-point. It work on early presumption of cluster center and after that for each cluster it allocate membership value to each data-point. It is an iterative process, dependent on minimization of cost function which describes the distance among the cluster center and data point. The iterative process updates the cluster center and the membership value iteratively and it ends when maximum iteration is achieved [56]. So, the time complexity of FCM clustering algorithm is given as $\mathcal{O}(ndc^2i)$ where, n is the number of datapoints, d is the number of dimensions$c$ is the number of clusters and i is the number of iterations.

The k-means is a partitioning iterative process that minimizes the cluster sum. The k-means algorithm the objective function is Squared Euclidean Distance (SED). In this algorithm columns represents the variables and the rows represent the data-points. The k-means clustering algorithm give back $n \times 1$ vector which includes the datapoint cluster indices [56]. So, the time complexity of the k-means algorithm can be given as $\mathcal{O}\ (ncdi)$ where, n represent the number of datapoints, d represent the number of dimensions,c represent the number of clusters and $i$represent the number of iterations.

The Principal Component Analysis (PCA) algorithm calculates the eigen-values and is an unsupervised dimension reduction technique. In PCA algorithm consists of two basic components that is co-variance matrix computation and eigen value decomposition. The time complexity of PCA algorithm can be given as $\mathcal{O}(n\,f^2 + f^3)$, as the time complexity of individual parts of PCA can be given as eigen value decomposition is $\mathcal{O}(\,f^3)$ co-variance matrix computation is $\mathcal{O}(\,n\,f^2)$. Here n is the number of observation or samples and$f$ is the number of features.

The Linear Discriminant Analysis (LDA) is a linear transformation scheme used for dimension reduction commonly. The LDA is supervised algorithm which would indicate axes that increase the separation among various classes. The time complexity of LDA algorithms can be given as $\mathcal{O}\ (nf^2)$ in case where $n > f$. otherwise the time complexity would be $\mathcal{O}(f^3)$. Here n represent number of samples or observations and f represent number of features.

Table 4.4: Time complexity of various algorithms

Algorithm	Time Complexity
K-means	$\mathcal{O}\ (ncdi)$
FCM	$\mathcal{O}(ndc^2i)$
PCA	$\mathcal{O}(n\,f^2 + f^3)$
LDA	$\mathcal{O}\ (nf^2)$
IFS	$\mathcal{O}\ (n^4)$
Proposed Algorithm	$\mathcal{O}\ (n^4c^2f^2i)$

The Infinite Feature Selection (IFS) technique is a filter-type algorithm which execute ranking step in an un-supervised way and then tracked by simple cross validation tactic for choosing the finest features. This approach estimates the significance of given feature, whereas taking in account every possible feature subset and this is considered as most interesting characteristic of this approach. The time complexity of IFS algorithm can be given as $O\ (n^4)$, where n is the number of samples or observations. It became unfeasible when number of features is very large.

The proposed algorithms have been proven to achieve efficient accuracy, but its major drawback is its time complexity. As proposed FCM algorithm comprises of LDA and IFS scheme which tends to contribute in its increased complexity. In proposed algorithm, LDA technique is used to extract the features from the data set and IFS is used to rank the features most useful features from the database and neglect those features which do not contribute much in the algorithm. The time complexity of proposed algorithm can be given as $O\ (n^4 c^2 f^2 \text{i})$, where $n$is the number of samples, $f$is the number of features, c is the number of clusters and$i$is the number of iterations. The time complexity of various algorithms is shown in table 4.4.

Table 4.5 shows the comparison of elapsed time between various algorithms. Proposed algorithm tends to have competitive elapsed time. The elapsed time is calculated on three data set- FEI, ORL, YALE-B datasets as shown in figure 4.6. For FEI dataset, the elapsed time of proposed algorithm is 65.255 sec which is less in comparison to FCM, k-means and PCA algorithm. The elapsed time of proposed algorithm is 149.449 sec for ORL dataset, which is less in comparison to FCM, k-means, IFS and PCA algorithm. For ORL dataset the elapsed time of proposed algorithm is 210.298 sec, which is less in comparison to FCM, k-means, LDA algorithm.

Table 4.5: Elapsed Time of Various Algorithms

Datasets	Algorithms					
	PCA	K-means	FCM	LDA	IFS	Proposed
FEI	77.297245	71.289431	73.678031	53.335347	49.396463	65.255263
ORL	151.902386	213.895243	252.114300	130.528031	160.975288	149.449250
YALE-B	173.629031	215.801542	242.640810	225.010921	194.373685	210.298548

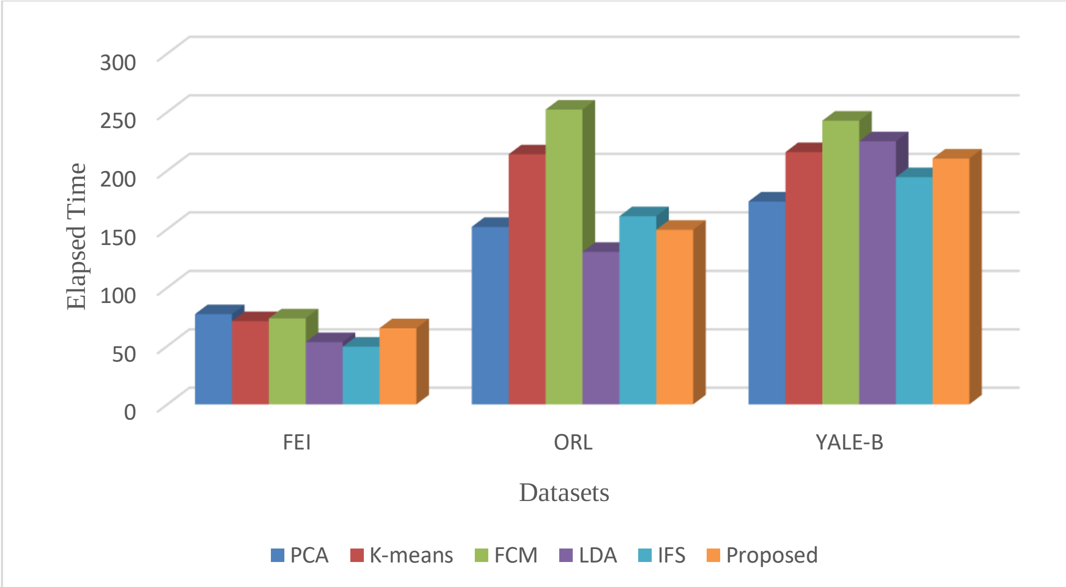


Figure 4.6: Computational time of various algorithms

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION

Clustering has long been used for feature construction. The idea is to replace a group of similar variables by a cluster centroid, which becomes a feature. The most popular algorithms include K-means and FCM clustering. Clustering is usually associated with the idea of unsupervised learning. It can be useful to introduce some supervision in the clustering procedure to obtain more discriminant features.

The proposed image clustering method in this research work is based on fuzzy c-means (FCM) clustering approach. FCM [41] is a representative algorithm of fuzzy clustering which is based on K-means concepts to partition dataset into clusters. The FCM algorithm is a soft clustering method in which the objects are assigned to the clusters with a degree of belief. Hence, an object may belong to more than one cluster with different degrees of belief. It attempts to find the most characteristic point in each cluster, named as the center of one cluster; then it computes the membership degree for each object in the clusters. The fuzzy c-means algorithm diminishes intra-cluster variance as well.

The proposed algorithm includes infinite feature selection [39] and linear discriminant analysis [38]. In proposed method, the particular area is selected and then the best features are chosen from that datasets and neglect those features that won't make any contribution in performance of algorithms thus providing more efficient results. Experimental results are provided in paper which demonstrates the efficiency of proposed algorithm on three dataset- FEI, ORL and Yale-B datasets. All the programming of proposed algorithms is carried out in Matlab software.

6.2 FUTURE SCOPE

The Fuzzy c-means clustering is an efficient approach for clustering-based application. In this the work has been done on improving FCM clustering algorithm efficiency, but the proposed algorithm faces time complexity problem. The proposed algorithm has achieved high accuracy and competitive time complexity in comparison to existing algorithms. So, future work would focus on reducing its time complexity. Moreover, try to improve its capability and efficiency as an image clustering approach and work on larger datasets.

REFERENCES

[1] Samuel, Arthur (1959). Some Studies in Machine Learning Using the Game of Checker, *IBM Journal of Research and Developmen*t, 3(3), 210–229.

[2] Mannila, Heikki (1996). Data mining: machine learning, statistics, and databases, *Int'l Conf. Scientific and Statistical Database Management. IEEE Computer Society*.

[3] Bishop, C. M. (2006). Pattern Recognition and Machine Learnin*g*, *Springer*, *ISBN* 0-387-31073-8.

[4] Machine Learning: What it is and why it matter. Available at http://www.sas.com/it_it/insights/analytics/machine-learning.html

[5] Kotsiantis, S. B., Zaharakis, I., &Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160, 3-24.

[6] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons , 2003.

[7] A. D. Gordon. *Classification*, Chapman & Hall/CRC ,1999.

[8] Judd, D., McKinley, P. K., & Jain, A. K. (1996, August). Large-scale parallel data clustering, *In Pattern Recognition, 1996, Proceedings of the 13th International Conference on* vol. 4, 488-493. IEEE.

[9] T. M. Mitchell, *Machine Learning*, McGraw Hill, 2005.

[10] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques, *Emerging artificial intelligence applications in computer engineering*, 160, 3-24.

[11] R.O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, John Wiley & Sons, Inc , 2001.

[12] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm, *Computers & Geosciences*, 10(2-3), 191-203.

[13] Hung, M. C., & Yang, D. L. (2001). An efficient fuzzy c-means clustering algorithm. *In Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, 225-232. IEEE.

[14] Howland, P., & Park, H. (2004). Generalizing discriminant analysis using the generalized singular value decomposition, *IEEE transactions on pattern analysis and machine intelligence*, 26(8), 995-1006.

[15] Wehrens, R., Buydens, L. M., Fraley, C., & Raftery, A. E. (2004). Model-based clustering for image segmentation and large datasets via sampling. *Journal of Classification*, 21(2), 231-253.

[16] Ye, J., Janardan, R., & Li, Q. (2005). Two-dimensional linear discriminant analysis. In *Advances in neural information processing systems*, 1569-1576.

[17] Li, T., Zhu, S., &Ogihara, M. (2006). Using discriminant analysis for multi-class classification: an experimental investigation, *Knowledge and information systems*, 10(4), 453-472.

[18] Ding, C., & Li, T. (2007, June). Adaptive dimension reduction using discriminant analysis and k-means clustering. In *Proceedings of the 24th international conference on Machine learning*, 521-528. ACM.

[19] Kannan, S. R., Ramathilagam, S., & Chung, P. C. (2012). Effective fuzzy c-means clustering algorithms for data clustering problems, *Expert Systems with Applications*, 39(7), 6292-6300.

[20] Zanaty, E. A. (2012). Determining the number of clusters for kernelized fuzzy C-means algorithms for automatic medical image segmentation, *Egyptian Informatics Journal*, 13(1), 39-58.

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

[22] Krinidis, S., &Chatzis, V. (2010). A robust fuzzy local information c-means clustering algorithm, *IEEE transactions on image processing*, 19(5), 1328-1337.

[23] Fahad, A., Alshatri, N., Tari, Z., *et al*. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis, *IEEE transactions on emerging topics in computing*, 2(3), 267-279.

[24] Abdu, K. K., Faiza, B., & Alias, R. A. (2014) A review and analysis of fuzzy-c means clustering techniques,*International Journal of Scientific & Engineering Research*, 5(11), 1072-1077.

[25] Arora, S., & Chana, I. (2014, September). A survey of clustering techniques for big data analysis. In *Confluence the Next Generation Information Technology Summit (Confluence), 2014 5th International Conference,* 59-65. IEEE.

[26] Rajapriyadharshini, R. (2015, March). SAR image denoising via clustering based Linear Discriminant Analysis, In *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on*, 1-6. IEEE.

[27] Memon, K. H., & Lee, D. H. (2016). Generalized fuzzy c-means clustering algorithm with local information, *IET Image Processing*, 11(1), 1-12.

[28] Saxena, A., Prasad, M., Gupta, A, *et al*. (2017). A review of clustering techniques and developments, *Neurocomputing*, 267, 664-681.

[29] Chen, J., Yang, C., Xu, G., & Ning, L. (2018, April). Image Segmentation Method Using Fuzzy C Mean Clustering Based on Multi-Objective Optimization, In J*ournal of Physics: Conference Series* (Vol. 1004, No. 1, pp. 012035). IOP Publishing.

[30] O. A. Mohamed Jafar and R. Sivakumar (2013). A Comparative Study of Hard and Fuzzy Data Clustering Algorithms with Cluster Validity Indices, *emerging research in computing, information, communication and applications ERCICA*.

[31] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review, *ACM computing surveys (CSUR)*, 31(3), 264-323.

[32] Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age, *ACM Computing Surveys (CSUR)*, 40(2), 5.

[33] Banerjee, B., Bovolo, F., Bhattacharya, A., *et al.* (2015). A new self-training-based unsupervised satellite image classification technique using cluster ensemble strategy, *IEEE Geoscience and Remote Sensing Letters*, 12(4), 741-745.

[34] Tao, D., Jin, L., Liu, W., & Li, X. (2013). Hessian regularized support vector machines for mobile image annotation on the cloud, *IEEE Trans. Multimedia*, 15(4), 833-844.

[35] Liu, G., Zhang, Y., & Wang, A. (2015). Incorporating Adaptive Local Information in to Fuzzy Clustering for Image Segmentation, *IEEE Trans. Image Processing*, 24(11), 3990-4000.

[36] Chen, L., Xu, D., Tsang, I. W. H., & Li, X. (2014). Spectral embedded hashing for scalable image retrieval, *IEEE transactions on cybernetics*, 44(7), 1180-1190.

[37] Li, T., Zhu, S., &Ogihara, M. (2006). Using discriminant analysis for multi-class classification: an experimental investigation. *Knowledge and information systems*, 10(4), 453-472.

[38] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, USA, 2nd edition, 2001.

[39] Roffo, G., Melzi, S. and Cristani, M., (2015). Infinite feature selection, *In Proceedings of the IEEE International Conference on Computer Vision*, 4202-4210.

[40] Roffo, G., Melzi, S., Castellani, U., &Vinciarelli, A. (2017). Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach. *arXiv preprint arXiv:*1707.07538.

[41] Saxena, A., Prasad, M., Gupta, A., *et al*. (2017). A review of clustering techniques and developments, *Neurocomputing*, 267, 664-681.

[42] Ghosh, S., & Dubey, S. K. (2013). Comparative analysis of k-means and fuzzy c-means algorithms, *International Journal of Advanced Computer Science and Applications*, 4(4).

[43] Sisodia, D., Singh, L., Sisodia, S., & Saxena, K. (2012). Clustering techniques: a brief survey of different clustering algorithms. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, 1(3), 82-87.

[44] Chattopadhyay, S., Pratihar, D. K., & Sarkar, S. C. D. (2012). A comparative study of fuzzy c-means algorithm and entropy-based fuzzy clustering algorithms, *Computing and Informatics*, 30(4), 701-720.

[45] Havens, T. C., Bezdek, J. C., Leckie, C., *et al*. (2012). Fuzzy c-means algorithms for very large data, *IEEE Transactions on Fuzzy Systems*, 20(6), 1130-1146.

[46] Nguyen Xuan Vinh, Julien Epps and James Bailey (2009). Information Theoretic Measures for Clustering Comparison: Is a Correction for Chance Necessary, *ICML 09: Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. 1073–1080.

[47] Yale dataset, available at http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html

[48] FEI dataset available at http://fei.edu.br/~cet/facedatabase.html

[49] ORL dataset available at http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[50] Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints, *Journal of machine learning research*, *5*(Nov), 1457-1469.

[51] Sipser, M. (2006). *Introduction to the Theory of Computation. Course Technology Inc*. ISBN 0-619-21764-2. Biolardi Yoshogi 13509035.

[52] Bradford, Phillip G.; Rawlins, Gregory J. E.; Shannon, Gregory E. (1998). Efficient Matrix Chain Ordering in Polylog Time, *SIAM Journal on Computing*. *Philadelphia: Society for Industrial and Applied Mathematics*, 27 (2), 466–490.

[53] Kumar, R., & Rubinfeld, R. (2003). Algorithms column: sublinear time algorithms, *ACM SIGACT News*, 34(4), 57-67.

[54] Naik, A. V., Regan, K. W., & Sivakumar, D. (1994, February). Quasilinear time complexity theory. In *Annual Symposium on Theoretical Aspects of Computer Science*, Springer, Berlin, Heidelberg, 97-108.

[55] Almeida, R. J., & Sousa, J. M. C. (2006, September). Comparison of fuzzy clustering algorithms for classification, In *Evolving Fuzzy Systems, 2006 International Symposium on*, 112-117. IEEE.

[56] Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms, *IEEE Transactions on neural networks*, 16(3), 645-678.

[57] El-Zaghmouri, B. M., Abu-Zanona, M. A., & Islamic, I. M. I. S. (2012). Fuzzy c-mean clustering algorithm modification and adaptation for applications. *World of Computer Science and Information Technology Journal (WCSIT)*, 2(1), 42-45.

LIST OF PUBLICATIONS

[1] Singroha P. and Mishra A. (2018), Fuzzy c-means clustering using LDA and Infinite feature selection, *Journals of system engineering and electronics*, (Communicated- Science Citation Indexed).

[2] Singroha P. and Mishra A. (2018), Non-negative matrix factorization- a survey, *International journal on electrical engineering and informatics*, (Communicated- Scopus Indexed).