European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 09, 2020

A UNIFIED DEEP LEARNING FRAMEWORK FOR TEXT DATA MINING USING DEEP ADAPTIVE FUZZY CLUSTERING

S. Praveen[1], Dr. R. Priya, MCA, M.phil, Ph.D[2],
Research Scholar[1], Associate Professor and Head,
Department of Computer Science [1,2]
Sree Narayana Guru College, Coimbatore,

Abstract

Text clustering is an important method for effectively organising, summarising, and navigating text information. The purpose of the clustering is to distinguish and classify the similarity among the text instance as label. However, in the absence of labels, the text data to be clustered cannot be used to train the text representation model based on deep learning as it contains high dimensional data with complex latent distributions. To address this problem, a new unified deep learning framework for text clustering based on deep representation learning is proposed using the deep adaptive fuzzy clustering in this paper to provide soft partition of data. Initially reconstruction of original data into feature space carried out using the word embedding process of deep learning. Word embedding process is a learnt representation of the text or sentence towards clustering into vector containing words, characters and N-grams of words. Further clustering of feature vector is carried out with max pooling layer to determine the inter-cluster seperability and intra-cluster compactness. Moreover learning of the feature space is processed with gradient descent. Moreover tuning of feature vector is fine tuned on basis of Discriminant information using hyper parameter optimization with fewer epochs. Finally representation learning and soft clustering has been achieved using deep adaptive fuzzy clustering and quantum annealing based optimization has been employed .The results demonstrate that the clustering approach more stable and accurate than the traditional FCM clustering algorithm on employing k fold validation for evaluation. The Experimental results demonstrates the proposed technique outperforms of state of arts approaches in terms of set based measures like Precision, Recall and F measure and rank based measures like Mean Average Precision and Cumulative Gain.

Keywords: Text Clustering, Deep learning, quantum annealing, Fuzzy Clustering, Learning representives

1. Introduction

Data Clustering is one of the important research issues in the field of text mining on large volume of data, where the mostly documents are classified with a machine learning models using supervised learning and unsupervised[1]. Especially fuzzy clustering is one of the most widely employed unsupervised machine learning model due to its flexible assignments of data to clusters with high scalability. Fuzzy Clustering uses concept of fuzzy sets and fuzzy logic on abundant fuzzy relationship[2]. Many Cases Fuzzy clustering is capable of handling continuous variable problem and equidistant problem by expanding the value of fuzzifier in the perspective of possibility theory. In addition, fuzzy clustering[3] leads to uncertainty and leads to large distance between the data instances in the cluster. With rise of data driven decision making models and exploration of data from various streams, Deep learning becoming increasingly important as their ability to uncover meaningful information and perform well out-of sample is put to the test in real world scenarios.

Deep learning discovers a good representation of data by neural networks. Jointly optimizing the deep neural network with an unsupervised clustering algorithm becomes an active research field. Motivated by the success of deep learning, this paper proposes a unified deep learning framework for text data mining using Deep Adaptive Fuzzy Clustering by representing the data in a feature space produced by the neural network. On perspective of representation learning, various constraints and strategies has been included on various layer of the neural network especially in dense layer and flat layer. Initially, Input layer, reconstruct the original data into the feature vector which further eases the learning of latent features. Latent features in form of feature vector is further processed on the deep clustering architecture using fuzzy relationship to avoid the problem of overfitting using max pooling layer in order to obtain the desired distribution. Distribution has been produced with minimized cluster compactness and maximized cluster seperability. Moreover tuning of feature vector with hyper parameter optimization produces few epochs. Affinity degree of data representation can be easily achieved. Quantum annealing based optimization can also be utilized to regularize the deep clustering model from the perspective of requiring new feature to retain the affinity in the original space.

The remainder of the paper is organized as follows: Section 2 discusses the related works in evaluation of machine learning methods and its impacts against the performing clustering of high dimensional data, Section 3 briefly discusses the proposed technique named deep adaptive fuzzy clustering and its hyper parameter optimization using quantum healing and Section 4 presents the experimental results on a number of data sets and performance measures on various metrics. Section 5 discusses conclusions and future work.

2. Related work

There exist many machine learning techniques for text clustering of high dimensional data are analysed in terms of design and implementation on various aspects. Each of these techniques follows some sort of effectiveness on the evaluation of the model and among those techniques, few performs nearly equivalent to the proposed model which is described as follows

2.1. Text Clustering using K means

K-means algorithm is an automatic partitioning of a number of sentences into a finite set of cluster in quick and easy operation, and is applied to the cluster analysis of text[4]. It tends to terminate iterative process quickly to only obtain partial optimal results and fluctuate the clustering result because of random selection of the initial iterative center point. Initially the conversion of text files into a numerical form can be performed using the Bag of-Words (BoW) approach. Each data object can finally be represented as a point in a finite-dimensional space, where the dimension corresponds to the number of unique tokens. Prior to the actual cluster analysis, a measure must also be defined to determine the similarity or dissimilarity between the objects. To measure dissimilarity, metrics such as Euclidean distance are used.

2.2. Text Clustering using Fuzzy C means Clustering

Fuzzy clustering allows for degrees of membership to which a data or text belongs to different cluster. FCM clustering[5] is one of well-know unsupervised clustering techniques, in this each document is an m-dimensional vector of m elements or m features. Since the m features in each document can have different units, in general, each of the features to a unified scale before clustering had normalized. In a geometric sense, each data is a point in m-dimensional feature space, and the set of document is a point set with n elements. However FCM algorithm requires the user to pre-define the number of clusters and different values of clusters corresponds to different fuzzy partitions. In this system, the FCM algorithm is executed to various values of clusters and results are evaluated by a cluster validity function, PBM index for internal performance measures and F-measure for external performance measure.

3. Proposed Model

In this section, we describe a unified deep learning framework for text data mining using deep adaptive fuzzy clustering on inclusion hyper parameter optimization and word embedding of the deep learning architecture as follows

3.1.Preliminaries

- **Dissimilarity Matrix**

Data to be clustered is presented in form of dissimilarity matrix which further partitions the data into subset. This matrix determines latent pervasive and clusters-specific factors estimated from the data. To deal with data with imperfect labels, likelihood values for each input data which denote the degree of membership toward the normal and abnormal has to be computed. The subset of the dataset is represented as feature vector. Dissimilarity matrix to streaming data is given as

$$D_N = [d_{i\,j}]$$

Where $d_{i\,j}$ represents dissimilarity between the data instance or objects. In addition, proposed model is capable of computing the topic difference effective based on determination of concept and semantic similarity. In addition, it identifies the bigrams available in the subset.

- **Data Sampling**

Random Projection is applied to dissimilarity Matrix to obtain reduced data Y_{ij}. Sampling begins by finding the distinguished objects in reduced data Y_{ij} which are furthest from each other. Sampling starts at a random point[6], and then chooses as the second sample which is furthest from the initial point with respect to a chosen measure of distance on the set being sampled. The third object selected maximizes the distance from both of the first two points. This process continues until maximum number of samples are chosen Then, each object in reduced data is grouped with its nearest distinguished instance. This stage divides the entire dataset into k groups.

- **Distance Matrix**

Samples in form of reduced feature subset Q are used to build the distance matrix.Reduced samples can be drastically different from each other due to the random nature of the mapping from large samples to reduced samples, the distance matrices will be diverse. Therefore, the $Q_{n \times n}$ distance matrices are aggregated to obtain a more reliable distance matrix $D_{d,i}$

- **Normalization**

Each distance matrix is computed from randomly projected samples, the distance of each data point from the remaining data points may have a different range in different distance matrices[7]. Therefore, the distance of each data point from the remaining datapoints is normalized to a unit scale in each distance matrix. The rows (or columns) of each $D_{d,i}$ are normalized such that the i j-th entry of $D_{d,i}$ is in [0,1].

3.2. Deep Adaptive Fuzzy Clustering

In this deep adaptive fuzzy clustering model that simultaneously discovers a deep representation and conducts the fuzzy clustering on the normalized representation using sampling technique. Further sampled data is processed on the deep representative learning against various constraints,

- **Input Layer: Vector of feature values**

Reduced Sample occurred in the previous process is represented in form vector containing the index for the each instance in the subset. It employs the bag of words to create vector. Vector is an interval representation by using maximum and minimum values of term frequency instance of sampled data. Vector is represented in key and value form. The dimension of reduced sample and vector has been represented below

Reduced Sample data x={x1, x2, x3....xn} $x_i \varepsilon R^d$

Vector Representation of Reduced Sample Data V = {v1, v2, v3...vn} $v_i \varepsilon R^c$

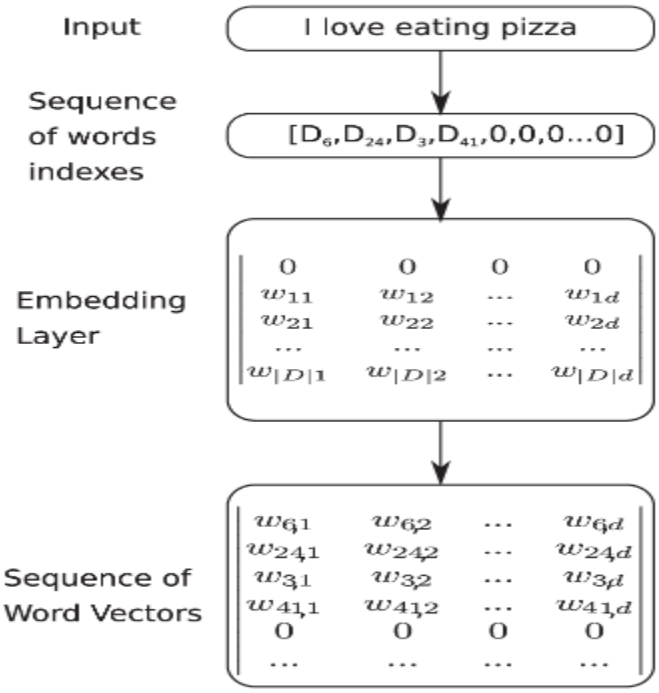
Weight of the Vector is computed as W= {w1, w2, w3...wc}

Weight of the vector is computed to determine the over fitting of the data instance. Over fitting in the subset can be eliminated on formulation of variance. It also uses the updating rule to eliminate.

Variance of the sample in subset is computed as

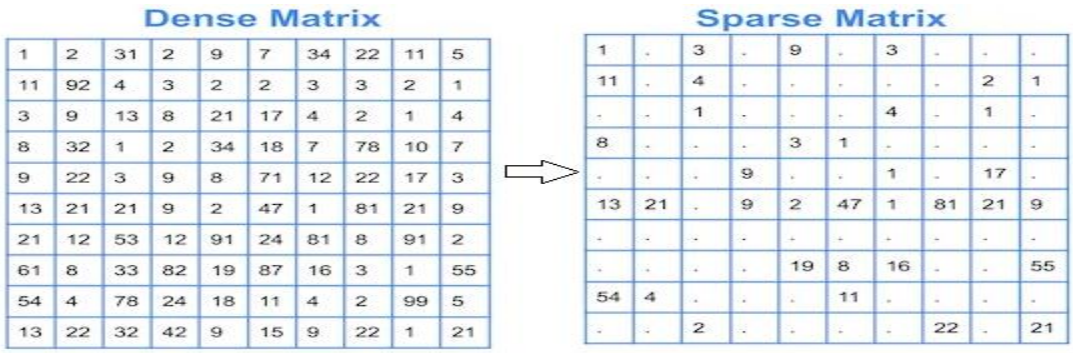
$$S_v = \sum_{k=0}^{n} \binom{d}{c} x^k v^{n-k} \frac{nx}{1!} + \binom{n}{k} x^k v^{n-k} \frac{n(n-1)x^2}{2!}$$

Variance determines the correlation between adjacent words and understanding complex semantics of instance in a subset. It discover the discriminating structure of a vector space, a Locality Preserving Indexing computes close inputs should have similar multi-valued data. Data flow of the work vector established is as follows



- **Hidden Layer**

It is employed to estimate the non linearity of the data instance. Non linearity is measured on basis of compactness and separation. It compactness determines the inter cluster distance and separation determines the intra cluster distance[8]. Sparse matrix is to reduce the distance between the data instance in subset. Non linearity is computed by sparse matrix for feature vector. Sparse matrix is given by



Sparse matrices are memory efficient data structures that enable us store large matrices with very few non-zero elements. It is used for perform complex matrix computations. It is given by

$$S_m = \sum_{k=0}^{n} (x - \mu) \sum_{k=0}^{n} (x - \mu) v^k a^{n-k}$$

Where μ is the fuzzifier aims to expand the distances from the data to its average point

- **Dense Layer**

It is employed to model the loss function of the fuzzier. Loss function is equivalent to minimize the within-class scatter matrix. Fuzzy membership for separation is defined in this layer along loss function. Model parameters are initialized as the next parameter, making the model suitable for clustering. The proposed fuzzy rule[9] based clustering has been explained taking into account that the observed data samples do not have labels. Model parameter for fuzzy membership is given as

$$V = \sum_{i=1}^{c} u_{ik}^{m} d_{ik}^{2} + \lambda\left(\sum_{i=1}^{c} u_{ik} - 1\right)$$

In this case, the clusters are created based on the existing prototypes and, thus, any prototype represents different clusters.

$$u_{st} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{st}^{2}}{d_{jt}^{2}} \right)^{\frac{1}{m-1}}}$$

It is related to the optimization of membership derived from the normalized fuzzy compactness and separation of the cluster instances. N gram represents different levels of semantics shown by the text data. Figure 1 represents the architecture of the proposed deep adaptive fuzzy clustering model. Further new representations are fine-tuned in accordance to discriminative information and the computing of affinity degree of data representation can be easily achieved using loss function. But the calculation of affinity usually requires some time-consuming steps as cross entrophy model use the Euclidean distance .

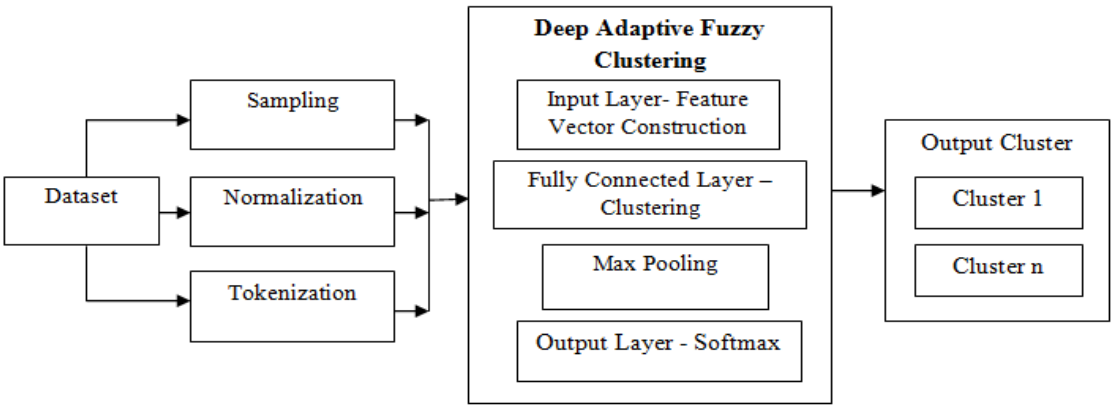


Figure 1: Architecture of the proposed model

In addition, the domain discriminator is added to the model, and the domain discriminator is used to discriminate the input samples. When the domain discriminator is powerful but unable to distinguish the data domain, a common learning model can be learned to solve the

domain adaptation. The domain-adapted parameters[10] are used as the model initialization, and the output feature vector F is clustered.

- **Fully Connected Layer**

The model parameters are continuously updated according to the distribution characteristics of the target data set, making it more suitable for the target domain. In this text clustering are obtained at the same time when the text representation suitable for the task is obtained using fuzzy membership $q_{i,j}$. Hyper Parameter value for fully connected layer through word vector is processed with following parameters. Table 1 provides the values of hyper parameter employed.

Table 1: Hyper parameter value for the fully connected layer

Hyper Parameter	Values
Batch Size	128
Learning Rate	0.01
Size of word vector	100
Number of Epoch	100
Maximum Number of words	20000
Maximum Sequence length	1000
Loss function	Cross entrophy

In fully connected layer, Feature vector has been processed with hyper parameter values on selected word vector and epoch. Further cross entropy loss function has been utilized to manage cluster seperability. Model parameter is updated to generate the cluster with minimum inter cluster distance. Updating of model parameter is given by

$$L(q_{i,j,Mj}) = \left(\sum_{k=0}^{n} x^k v^{n-k} \frac{nx}{1!} + \binom{n}{k} x^k v^{n-k} \frac{n(n-1)x^2}{2!}\right) - \left(\sum_{k=0}^{n} (x-\mu) \sum_{k=0}^{n} (x-\mu) v^k a^{n-k}\right)$$

In other words, it shows how the between-cluster distance effectively affects the within-cluster distance in the expression of cluster Membership [11].

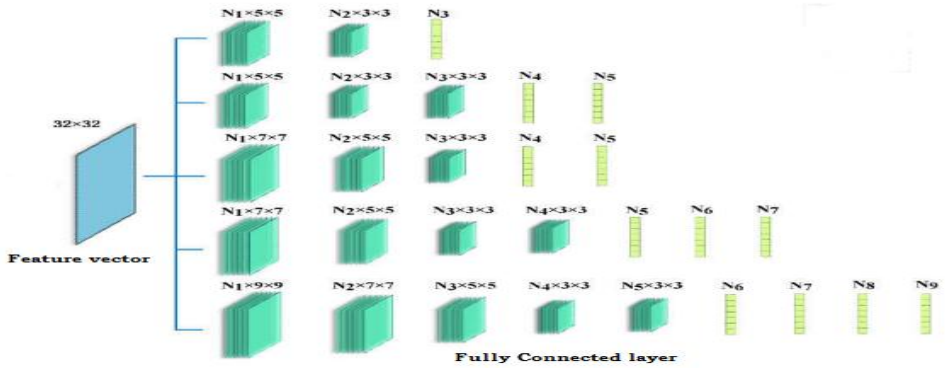


Figure 2: Processing Fully Connected Layer on Feature Vector using hyperparameter

Hyper parameter is important for the balance between the within-cluster distance and the between-cluster distance in the cluster space.

$$Q = \sum_{i=1}^{c} u \sum_{k=1}^{N} u_{ik}^{m} \| \mathbf{x}_k - \mathbf{v}_i \|^2$$

Where U is subject to partitions and prototype of the cluster.,

The model can be used as the initialization of the subset to prevent the gradient from disappearing and being simulated in the training process caused by random initialization of parameters.

- **Output layer:**

Softmax function has enabled as activation function to provide effective probability distribution over output values which is form of cluster. It composed of *K* probabilities proportional to the exponentials of cluster. Difference between current and target output values which are propagated back through network to adjust weights. Soft max function[12] adjusts the weight as affinities between input data should be consistent with the ones between new representations.

$$\text{Softmax } M_z = \sum ||L_{i\_} S_i||^2 D_{ij}$$

Hyper parameter obtained soft max function is optimized using quantum annealing. The particular approach that acquires the discriminative information from the fuzzy membership through clustering in the fully connected layer is to compute the affinity of the data. Affinity of the data can be updated during the training process on the each cluster, and the updating epoch can be larger than 1. As the training epoch increases, the loss will slowly decrease and the accuracy of model will be increased.

$$V = \sum_{i=1}^{c} u_{ik}^{m} d_{ik}^{2} + \lambda \left(\sum_{i=1}^{c} u_{ik} - 1 \right)$$

Where $d_{ik} = \|x_k - v_i\|^2$ and λ –Lagrange multiplier

Algorithm: Deep Adaptive Fuzzy Clustering

Input: Data $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}$

Output: N Cluster

Process

Input Layer: Transform X into Vector

Compute Distance $D_{ij}()$ of the Vector

If D_{ij} of Instance i > Threshold

Generate new group or partition to store instance i

Hidden Layer: Compute Reconstruction Error

$$Q = \sum_{ii=1}^{p} \sum_{k=1}^{c[ii]} \| \hat{\mathbf{v}}(\mathbf{v}_k[ii]) - \mathbf{v}_k[ii] \|_{\mathbf{F}_{ii}}^2$$

mp= Max_pooling(Q)

Fully Connected Layer: Generate Fuzzy Membership for Clustering

$$\text{Employ } L(q_{i,j,Mj}) = \sum_{i=1}^{c} \min(u_{ik}, u_{il})$$

fc= Fully Connected(mp)

Class label= Soft_Max(fc)

Output layer: Softmax using Quantum annealing for Cluster

The training algorithm of the proposed deep fuzzy clustering model is summarized in algorithm 1. The Saliency map[13] has been derived in this section to interpret and rationalize the decision of the trained system.

$$x_{ij} = \frac{1}{K} \sum_{l=0}^{k} x(l)$$

In the above equation, K termed as weight factor and x is considered as feature

4. Experimental results

In section, we describe the experimental results of the proposed framework against the existing approaches on the real-world datasets to evaluate the clustering performance. All the

experimental results, analyses, and conclusions are provided as well. The model is experimented in Dotnet. In that processing of the clustering the text, train and validate the system is highly challenging. In this 60% of data has used to train corresponding text in word embedding and 20% is used to validate the proposed model. On 80% training data, it has been divided as 60% for train the model and 20% to validate the trained model. In this 5 fold cross validation is applied to prevent the data leakage and to improve the accuracy of the training model. The training parameter of the reprehensive learning has been defined in the table 2

Table 2: Training parameters

Parameter	Value
Learning rate	10^{-6}
Loss Function	Categorical cross entropy
Batch size	15
Max epoch	1000

4.1. Dataset Description

We have done extensive experiments on 3 real datasets to measure the clustering performance and each dataset segmented into equal parts for training and testing. In this experiment, training of model consumes 60%, Validation consumes 20% and testing consumes 20%. Detailed properties of the dataset is as follows

- **RCV1 (Reuters Corpus Volume I).**

This data set contains corpus of newswire describing the collection of the news which is mostly frequently used bench mark dataset.

- **Forest covers data set from UCI repository (Forest).**

The data set contains geospatial descriptions of different types of forests. We normalize the data set, and arrange the data so that new classes appear randomly[14].

- **Twitter**

This data set contains 340,000 Twitter messages (tweets) of different trends (classes) in different area and subjects.

4.2. Evaluation

The proposed Framework is evaluated against the following measures after processing architecture of deep representative learning. In this work, model is evaluated using k fold validation to evaluate the performance of clustering on various dataset mentioned. The evaluation of the proposed model depends on the process of activation function, weight, bias and hyper parameter of model. In addition fuzzy membership used criteria to partition the cluster.

- **Precision**

Positive predictive value is the fraction of relevant instances among the each cluster groups. Precision is the number of correct feature divided by the number of all returned feature space.

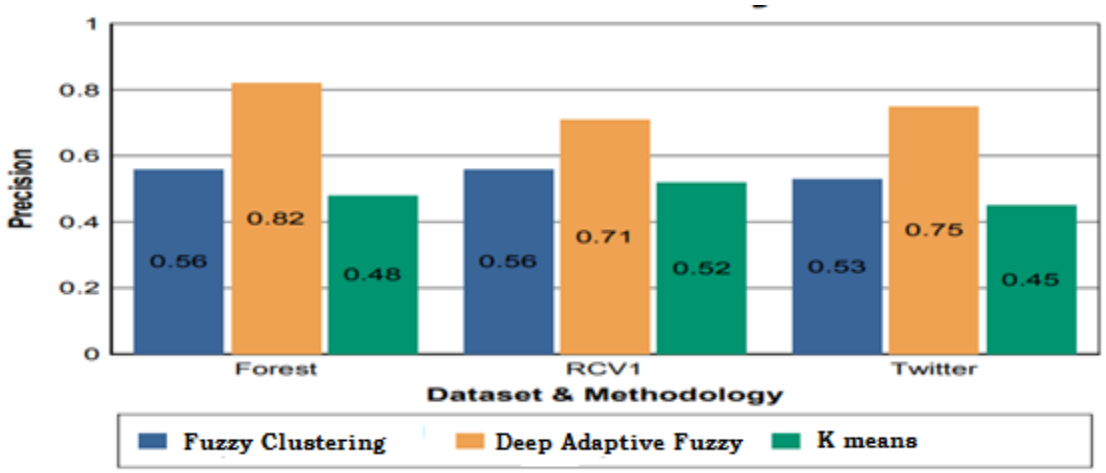


Figure 3: Performance Evaluation of the Precision towards technique against different datasets.

$$\text{Precision} = \frac{\text{True positive}}{True\ positive + False\ Positive}$$

True positive is a number of real positive cases in the data and false negative is number of real negative cases in the data[15]. The precision is evaluated against different dataset is depicted in the figure 3

- **Recall**

  It is the fraction of relevant instances that have been retrieved over the total amount of relevant instances of cluster. The recall is the part of the relevant documents that are successfully classified into the exact classes

$$\text{Recall} = \frac{\text{True positive}}{True\ positive + False\ negative}$$

True positive is a number of real positive cases in the data and false negative is number of real negative cases in the data. The recall is evaluated against different dataset is depicted in the figure 4

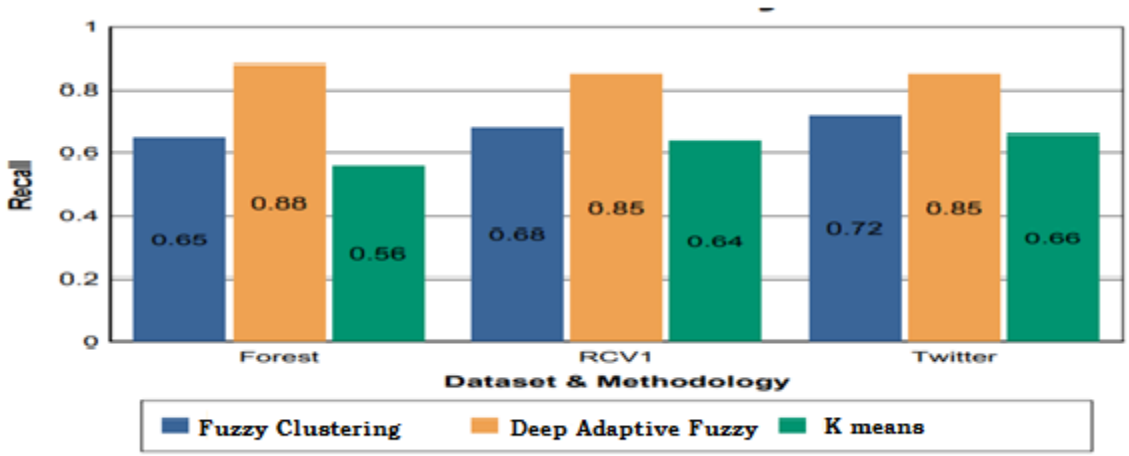


Figure 4: Performance Evaluation of the recall towards technique against different datasets.

- **F measure**

It is the number of correct class predictions to the single document to total number of predictions to whole document

Accuracy is given by

$$\frac{\text{True positive} + \text{True Negative}}{True\ positive + True\ Negative + false\ positive + False\ negative}$$

Although different document may have different impact on novel class detection, they are likely to have the same impact on classification. However, after a certain point, this improvement is diminished because of curse of dimensionality. On the other hand, for the Forest data set, the novel classes are less separable.

Table 3: Performance Evaluation of the Technique to Reuters Corpus Volume I dataset

Technique	**Precision**	**Recall**	**F measure**
K means	0.45	0.66	0.54
Fuzzy Clustering	0.53	0.72	0.61
Deep Adaptive Fuzzy Clustering - Proposed	0.75	0.85	0.80

. The membership function which indicates that the selection of m has impacts on the clustering performance as Fuzzifier m helps to accelerate the convergence speed of the algorithm and enhances the clustering performance.

Table 4: Performance Evaluation of the Technique to twitter dataset

Technique	Precision	Recall	F measure
K Means	0.46	0.72	0.55
Fuzzy Clustering	0.52	0.79	0.61
Deep Adaptive Fuzzy Clustering - Proposed	0.79	0.86	0.80

Hyper-parameter is a very important component of the proposed deep fuzzy clustering models. In this grid search method with the extensive designed range to find the sensitive region of the hyper-parameters has been performed. In addition, the cross-validation has been used to twitter dataset alone to find the best value of the hyper-parameters

Table 5: Performance Evaluation of the Technique to Reuters Forest dataset

Technique	Precision	Recall	F measure	Cumulative Gain	Mean Average Precision
K Means	0.46	0.66	0.64	135	0.64
Fuzzy Clustering	0.54	0.72	0.63	145	0.73
Deep Adaptive Fuzzy Clustering – Proposed	0.77	0.85	0.84	185	092

The evaluation of result is described in the table 3 for Reuters Corpus Volume I dataset, in table 4 for twitter dataset and table 5 for forest dataset. It is observed that the proposed method is always better when compared to clustering results of state of art approaches.

Conclusion

In this work, a unified deep learning framework for text mining using deep adaptive fuzzy clustering to provide soft partition of data has been designed and experiment on k fold validation. Specifically Word embedding process has been implemented in the input layer to generate feature vector simultaneously. Proposed model find the representation model for the clustering the text dataset on regularization of inter-cluster seperability and intra-cluster compactness using hyper parameter optimization using quantum annealing on fully connected layer and saliency maps effectively generates cluster with minimum distance among instances. Discriminant information is effectively handled by soft max function of the model. Finally it is proved that deep learning model

using fuzzy clustering performs consistently better than machine learning models on high dimensional real world datasets.

References

[1] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on. IEEE, , pp. 761–766, 1979.

[2] S. P. Chatzis and T. A. Varvarigou, "A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation," IEEE Transactions on Fuzzy Systems, vol. 16, no. 5, pp. 1351–1361, 2008.

[3] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified mrf energy function for change detection in synthetic aperture radar images," IEEE Transactions on Fuzzy Systems, vol. 22, no. 1, pp. 98– 109, 2014

[4] K.-L. Wu, J. Yu, and M.-S. Yang, "A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests," Pattern Recognition Letters, vol. 26, no. 5, pp. 639–652, 2005.

[5] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

[6] C. Song, F. Liu, Y. Huang, L. Wang, and T. Tan, "Auto-encoder based data clustering," in Iberoamerican Congress on Pattern Recognition. Springer, , pp. 117–124, 2013.

[7] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in Proceedings of the 25th international conference on Machine learning. ACM, , pp. 1096–1103, 2008.

[8] H. Liu, M. Shao, S. Li, and Y. Fu, "Infinite ensemble for image clustering," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, , pp. 1745–1754, 2016.

[9] C.-H. Li, B.-C. Kuo, and C.-T. Lin, "Lda-based clustering algorithm and its application to an unsupervised feature extraction," IEEE Transactions on Fuzzy Systems, vol. 19, no. 1, pp. 152–163, 2011.

[10] L. Bai, J. Liang, C. Dang, and F. Cao, "A novel fuzzy clustering algorithm with between-cluster information for categorical data," Fuzzy Sets and Systems, vol. 215, pp. 55–73, 2013.

[11] Y. Peng, S. Wang, X. Long, and B.-L. Lu, "Discriminative graph regularized extreme learning machine and its application to face recognition," Neurocomputing, vol. 149, pp. 340–353, 2015.

[12] K. G. Dizaji, A. Herandi, C. Deng, W. Cai, and H. Huang, "Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization," in 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, , pp. 5747–5756, 2017.

[13] A. Dosovitskiy, P. Fischer, J. T. Springen berg, M. Riedmiller, and T. Brox, "Discriminative unsupervised feature learning with exemplar convolutional neural networks," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 9, pp. 1734–1747, 2016.

[14] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. IEEE, s, pp. 31–35, 2016.

[15] D. Huang, X. Cai, and C. D. Wang, "Unsupervised feature selection with multi-subspace randomization and collaboration," Knowledge-Based Systems, vol. 182, p. 104856, oct 2019.