Enhancement of Fuzzy Possibilistic C-Means Algorithm using EM Algorithm (EMFPCM)

R.Shanthi
Research Scholar, Dept. of Computer Science,
Dr.SNS Rajalakshmi College of Arts & Science,
Coimbatore-49, INDIA

R.Suganya
Assistant Professor, Dept. of Computer Science,
Dr.SNS Rajalakshmi College of Arts & Science,
Coimbatore-49, INDIA

ABSTRACT

The major difficulties that arise in several fields, comprising pattern recognition, machine learning and statistics, is clustering. The basic data clustering problem might be defined as finding out groups in data or grouping related objects together. A cluster is a group of objects which are similar to each other within a cluster and are dissimilar to the objects of other clusters. The similarity is typically calculated on the basis of distance between two objects or clusters. Two or more objects present inside a cluster and only if those objects are close to each other based on the distance between them. In order to provide better clustering approaches that fits for all applications and to improve the efficiency of data clustering, this paper proposes a effective clustering techniques called Enhancement of Fuzzy Possibilistic C-Means Algorithm using EM Algorithm (EMFPCM). Thus with the help of EMFPCM, noise is reduced, provides more accuracy and thus provides better result in predicting the user behavior. The algorithm was implemented and the experiment result proves that this method is very effective in predicting user behavior. This approach is suitable for applications in business, such as to design personalized web service. The performance of the proposed approaches is evaluated on the UCI machine repository datasets such as Iris, Wine, Lung Cancer and Lymphograma. The parameters used for the evaluation is Clustering accuracy, Mean Squared Error (MSE), Execution Time and Convergence behavior.

Keywords

Feature Selection, C-Mean Clustering, EM Algorithms, Mat lab, Unsupervised Learning

1. INTRODUCTION

Data analysis is considered as a very important science in the real world. Cluster analysis is a technique for classifying data; it is a method for finding clusters of a data set with most similarity in the same cluster and most dissimilarity between different clusters. The conventional clustering methods put each point of the data set to exactly one cluster. Since 1965, Zadeh proposed fuzzy sets in order to come closer of the physical world [1]–[2]–[3]. Zadeh introduced the idea of partial memberships described by membership functions. An efficient technique in reducing these difficulties is web mining [4]. Web mining is that data mining technique is applied to web data to the discovery of the interesting usage patterns and implicit information [5].

Web usage mining includes three most important steps:

- Data Preprocessing
- Knowledge Extraction
- Analysis of Extracted Results

After the preprocessing step, knowledge extraction is done by using clustering algorithm to extract the interesting patterns based on the user behavior. Clustering is one of the most common techniques used for data analysis and classification [6]. It is described as the method of grouping N item sets into individual clusters based on comparison or distance function [7]. A high-quality clustering technique possibly will yield clusters thus have more inter cluster and less intra cluster distance. The main purpose of clustering is to increase the resemblance of the data points within each cluster and also increase the dissimilarity across clusters. Thus with the help of clustering technique, it is easy to extract and analyze the log data from the web sites by using the technology of web data mining and behavior analysis. Through this way, it is very simple to find out the information interrelated to the user's behavior, which is important for personalizing the web sites. User's behavior on web sites specifies the character of the user, possibly will help in the personalization of web sites. The process of web site user's behavior is essentially composed of the data pre-processing, the discovery of website user's behavior model and the analysis of user's behavior. The user's behavior analysis is a series of activities from identifying users' objective to evaluating the final result.

The previous approaches like FCM, EMFCM and FPCM are the clustering approaches. Were the FCM is an iterative algorithm. The aim of FCM is to find cluster centers (centroid) that minimize a dissimilarity function. This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. In the second proposed algorithm is called Expectation maximization fuzzy c-means clustering (EMFCM). EMFCM algorithms is very help full to increase the performance of machine learning, all data mining approaches, image processing, network security etc. Now the EMFPCM algorithm is used in this paper to predict the user behavior in more accurate way.

2. RELATED WORK

Bradley et al., [8] put forth a technique for the purpose of refining initial points for clustering algorithms, in particular K-

Means clustering algorithm. They presented a quick and competent algorithm in order to refine an initial starting point a normal class of clustering algorithms. The iterative techniques that are more sensitive to initial preliminary conditions were used in most of the clustering algorithms like K-Means, and EM normally converges to one local minima. They implemented this iterative technique for refining the initial condition which permits the algorithm to converge to an enhanced local minimum value. The refined initial point is used to evaluate the performance of K-Means algorithm in clustering the given data set.

Jiabin Deng et al., [9] proposed an improved fuzzy clustering-text clustering method based on the fuzzy C-Means clustering algorithm and the edit distance algorithm. The author used the feature evaluation to reduce the dimensionality of high-dimensional text vector. Because the clustering results of the traditional fuzzy C-Means clustering algorithm lack the stability, the author introduced the high-power sample point set, the field radius and weight. Due to the boundary value attribution of the traditional fuzzy C-Means clustering algorithm, the author recommended the edit distance algorithm.

Celikyilmaz et al., [10] proposed a new fuzzy system modeling approach based on improved fuzzy functions to model systems with continuous output variable. The new modeling approach introduces three features: i) an Improved Fuzzy Clustering (IFC) algorithm, ii) a new structure identification algorithm, and iii) a nonparametric inference engine. The IFC algorithm yields simultaneous estimates of parameters of c-regression models, together with fuzzy c-partitioning of the data, to calculate improved membership values with a new membership function. The structure identification of the new approach utilizes IFC, instead of standard fuzzy C-Means clustering algorithm, to fuzzy partition the data, and it uses improved membership values as additional input variables along with the original scalar input variables for two different choices of regression methods: least squares estimation or support vector regression, to determine ldquofuzzy functionsrdquo for each cluster. With novel IFC, one could learn the system behavior more accurately compared to other FSM models. The nonparametric inference engine is a new approach, which uses the alike -nearest neighbor method for reasoning.

In 1997, Pal et al., [11] proposed the fuzzy-possibilistic C-Means (FPCM) technique and algorithm that generated both membership and typicality values when clustering unlabeled data. FPCM constrains the typicality values so that the sum over all data points of typicality's to a cluster is one. For large data sets the row sum constraint produces unrealistic typicality values. In this approach, a new model is presented called Possibilistic-Fuzzy C-Means (PFCM) model. PFCM produces memberships and possibilities concurrently, along with the usual point prototypes or cluster centers for each cluster. PFCM is a

$$0 \prec \sum_{j=1}^{n} u_{ij} \prec 1, \forall i \in \{1, \dots, c\} \qquad (2)$$

hybridization of FCM and Possibilistic C-Means (PCM) that often avoids various problems of PCM, FCM and FPCM. The noise sensitivity defect of FCM is resolved in PFCM, overcomes the coincident clusters problem of PCM and eliminates the row sum constraints of FPCM. The first-order essential conditions for extreme of the PFCM objective function is driven, and used them as the basis for a standard alternating optimization approach to find local minima of the PFCM objective function. PFCM prototypes are less sensitive to outliers and can avoid coincident clusters, PFCM is a strong candidate for fuzzy rule-based system identification

An EM (Expectation maximization) algorithm is very use full in statically model. The most common algorithm uses an iterative refinement technique. These algorithms are giving the best result in clustering method; it is also referred to as LoyardAlgo particularly in the computer science community. EM algorithms given an initial set of c-means $m_1^{(1)} \dots m_k^{(1)}$, the algorithm proceeds by alternating between two steps [12].

3. IMPLEMENTATION OF FUZZY POSSIBILISTIC C-MEANS ALGORITHM USING EM ALGORITHM (EMFPCM)

3.1 Fuzzy Possibilistic C-Means Algorithm for Clustering (FPCM)

The fuzzified version of the K-Means algorithm is the Fuzzy C-Means (FCM) using EM Algorithm. It is a method of clustering which allows one piece of data to belong to two or more clusters. This method was developed by Dunn in 1973 [13] and Modified by Bezdek in 1981 [14] and this is frequently used in pattern recognition. The algorithm is an iterative clustering method that brings out an optimal c partition by minimizing the weighted within group sum of squared error objective function J_{FCM}:

$$J_{FCM}(V, U, X) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \, d^2(x_j, v_i), 1 < m < +\infty \qquad (1)$$

In the equation $X = \{x_1, x_2, \dots, x_n\} \subseteq R^p$ is the dataset in the p-dimensional vector space, the number of data items is represented as p, c is the number of clusters with $2 \leq c \leq n - 1$. $V = \{v_1, v_2, \dots, v_c\}$ is the c centers or prototypes of the clusters, v_i represents the p-dimension center of the cluster i, and $d^2(x_j, v_i)$ represents a distance measure between object x_j and cluster center v_i. $U = \{u_{ij}\}$ represents a fuzzy partition matrix with $u_{ij} = u_i(x_j)$ isthe degree of membership of x_j in the ith cluster; x_j is the jth of p-dimensional measured data. The fuzzy partition matrix satisfies:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1, \dots, n\} \qquad (3)$$

where m is a weighting exponent parameter on each fuzzy membership and establishes the amount of fuzziness of the resulting classification; it is a fixed number greater than one. Under the constraint of U the objective function J_{FCM} can be minimized. Specifically, taking of J_{FCM} with respect to u_{ij} and v_i and zeroing them respectively, is necessary but not sufficient conditions for J_{FCM} to be at its local extrema will be as the following:

$$u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{d^2(x_j, v_i)}{d^2(x_j, v_k)}\right)^{2/(m-1)}\right]^{-1}, 1 \leq i \leq c, 1 \leq j \tag{4}$$
$$\leq n.$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}, 1 \leq i \leq c. \tag{5}$$

In noisy environment memberships of FCM do not always correspond well to the degree of belonging of the data, and may be inaccurate, because the real data unavoidably involves some noises. To recover this weakness of FCM, relaxed the constrained condition (14) of the fuzzy c-partition to obtain a possibilistic type of membership function and PCM for unsupervised clustering is proposed. The component generated by the PCM corresponds to a dense region in the dataset; each cluster is independent of the other clusters in the PCM strategy. The following formulation is the objective function of the PCM.

$$J_{PCM}(V, U, X) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ik}^m d^2(x_j, v_i)$$
$$+ \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n}(1 - u_{ij})^m \tag{6}$$

where

$$\eta_i = \frac{\sum_{j=1}^{n} u_{ij}^m ||x_j - v_i||^2}{\sum_{j=1}^{n} u_{ij}^m} \tag{7}$$

is the scale parameter at the ith cluster,

$$u_{ij} = \frac{1}{1 + \left[\frac{d^2(x_j, v_i)}{\eta_i}\right]^{\frac{1}{m-1}}} \tag{8}$$

is the possibilistic typicality value of training sample x_j belonging to the cluster i. $m \in [1, \infty]$ is a weighting factor called the possibilistic parameter. Typical of other cluster approaches, the PCM also depends on initialization. In PCM techniques, the clusters do not have a lot of mobility, since each data point is classified as only one cluster at a time rather than all the clusters simultaneously. Consequently, a suitable initialization is required for the algorithms to converge to nearly global minimum.

The characteristics of both Fuzzy and Possibilistic C-Means are combined. Memberships and typicality's are important for the correct feature of data substructure in clustering problem. Thus, an objective function in the FPCM depending on both memberships and typicality's can be represented as below:

$$J_{FPCM}(U, T, V) = \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij}^m + t_{ij}^\eta) d^2(x_j, v_i) \tag{9}$$

with the following constraints:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1, \dots, n\}$$

$$\sum_{j=1}^{n} t_{ij} = 1, \forall i \in \{1, \dots, c\} \tag{10}$$

A solution of the objective function can be obtained through an iterative process where the degrees of membership, typicality and the cluster centers are updated with the equations as follows.

$$u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{d^2(x_j, v_i)}{d^2(x_j, v_k)}\right)^{2/(m-1)}\right]^{-1}, 1 \leq i \leq c, 1$$
$$\leq j \leq n. \tag{11}$$

$$t_{ij} = \left[\sum_{k=1}^{n}\left(\frac{d^2(x_j, v_i)}{d^2(x_j, v_k)}\right)^{2/(\eta-1)}\right]^{-1}, 1 \leq i \leq c, 1 \leq j$$
$$\leq n. \tag{12}$$

$$v_i = \frac{\sum_{k=1}^{n}(u_{ik}^m + t_{ik}^\eta)X_K}{\sum_{k=1}^{n}(u_{ik}^m + t_{ik}^\eta)}, 1 \leq i \leq c. \tag{13}$$

FPCM constructs memberships and possibilities simultaneously, along with the usual point prototypes or cluster centers for each cluster. Hybridization of PCM and FCM is the FPCM using EM Algorithm that often avoids various problems of PCM, FCM and FPCM using EM. FPCM using EM solves the noise sensitivity defect of FCM, overcomes the coincident clusters problem of PCM. But the estimation of centroids is influenced by the noise data. Hence Fuzzy Possibilistic C-Means Algorithm using EM Algorithm (EMFPCM) is proposed.

3.2 EM (Expectation Maximization) Algorithm

An EM (Expectation maximization) algorithm is very use full in statically model. The most common algorithm uses an iterative refinement technique. These algorithms are giving the best result in clustering method; it is also referred to as LoyardAlgo particularly in the computer science community. EM algorithms given an initial set of c-means $m_1^{(1)} \dots m_k^{(1)}$, the algorithm proceeds by alternating between two steps [12]

Assignment step

In these steps Assign each observation to the cluster with the closest means, it is according to Voronoi diagrams for finding new mean [15].

$$s_i^{(l)} = \{x_p : \| x_p - m_i^{(l)} \| \leq \| x_p - m_j^{(l)} \| \; \forall 1 \leq j \leq k \qquad (14)$$

Where each x_p goes into exactly one $s_i^{(l)}$ even if it could go in two of them

Update step by

Calculate the new means to be the centroid of the observations in the cluster.

$$m_i^{(l+1)} = \frac{1}{\left| s_i^{(l)} \right|} \sum_{x_{j \Sigma s_i^{(l)}}} X_j \qquad (15)$$

In this equation (15) find new mean for new cluster. So we can say these algorithms is deemed to have converged when the assignment no longer change. And it gave the best performance of initial means. EM algorithms commonly used initialization methods are random partition.

3.3 Fuzzy Possibilistic C-Means Algorithm using EM Algorithm (EMFPCM)

The proposed algorithm is called Expectation maximization fuzzy Possibilistic c-means clustering (EMFPCM). EMFPCM algorithms is very help full to increase the performance of machine learning, all data mining approaches, image processing, network security etc. Proposed algorithms used the clustering techniques and EM algorithms, which provided the sufficient result for the cluster analysis in maximum mean to calculate the fixed centroid and correct threshold value. Proposed algorithms have these steps.

Step1: In these step we find the membership matrix (U) initialize randomly in equation 4.

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1, \ldots, n\}$$

$$u_{ij} = \left[\sum_{k=1}^{c} \left(\frac{d^2(x_j, v_i)}{d^2(x_j, v_k)} \right)^{2/(m-1)} \right]^{-1}, 1 \leq i \leq c, 1 \leq j \leq n.$$

This equation represent the membership matrix, it has taken the value equal to 1.

Step2: Calculate the centroids (V_i) in equation 5

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}, 1 \leq i \leq c.$$

Centroid is main point of the cluster analysis system, in clustering this value of v_i is depends on the member matrix function and related parameter of X_j.

Steps3: Using dissimilarity function to calculate the dissimilarities between centroid and data points in equation 9 and check threshold value in equation 3.4 then stop if we find the correct threshold value.

$$J_{FPCM}(U, T, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij}^m + t_{ij}^\eta) \, d^2(x_j, v_i)$$

In this steps check the threshold value using membership matrix and Euclidian distance between ith centroid (V_i) and j_{th} data point.

$$if \; \| U(k+1) - U(k) \| < \in$$

In this equation we have to check the value of present classes and the next classes of the membership function. Check the value of membership matrix in next membership matrix in correct threshold value, which have been done by dissimilarities between centroid and data points. If values are satisfied, then we forwarded the next steps.

Step4: If threshold value is not correct we find the new mean (mi) using EM algorithm that has constraints in equation (14) using this equation we can find new mean, which is provided the correct threshold value for the dissimilarity function, So we can say these algorithms is deemed to have converged when the assignment no longer change. And it gave the best performance of initial means. EM algorithms commonly used initialization methods are random partition.

$$m_i^{(l+1)} = \frac{1}{\left| s_i^{(l)} \right|} \sum_{x_{j \Sigma s_i^{(l)}}} X_j$$

Step5: In these steps assign each observation to the cluster with the closest means

$$s_i^{(l)} = \{x_p : \| x_p - m_i^{(l)} \| \leq \| x_p - m_j^{(l)} \| \; \forall 1 \leq j \leq k$$

Hence find new mean for new cluster. So we can say these algorithms is deemed to have converged when the assignment no longer change.

The proposed algorithm using EM Algorithm is developed to obtain better quality of clustering results. The objective function is based by adding new weight of data points in relation to every cluster and modifying the exponent of the distance between a point and a class. Thus with the help of EMFPCM, noise is reduced, provides more accuracy and thus provides better result in predicting the user behavior. The algorithm was implemented and the experiment result proves that this method is very effective in predicting user behavior.

4. EXPERIMENTAL RESULTS

The performance of the proposed approaches is evaluated based on the following parameters clustering accuracy, execution time and convergence behaviour. The proposed approach is evaluated using the UCI machine repository such as,

- Iris Dataset
- Wine Dataset
- Lung cancer Dataset
- Lymphograma Dataset

4.1 Performance Evaluation

To evaluate the performance of proposed EMFPCM the experiments were carried out using the similar experimental setup and parameters

4.1.1 Clustering Accuracy

Clustering accuracy of the proposed EMFPCM is evaluated using the above mentioned datasets. Table 1 shows the comparison of the clustering accuracy of the clustering techniques is done by using four datasets.

Table 1:Clustering Accuracy Comparison of Four Approaches

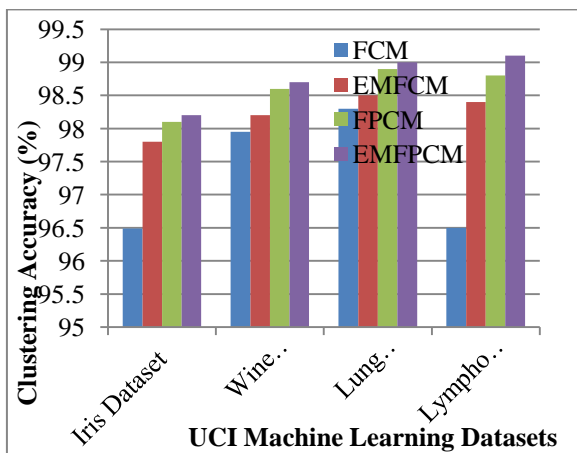| Number of Clusters	FCM	EMFCM	FPCM	EMFPCM
Iris Dataset	96.49	97.80	98.10	98.20
Wine Dataset	97.95	98.20	98.60	98.70
Lung Cancer Dataset	98.30	98.50	98.90	99.00
Lymphograma Dataset	96.50	98.40	98.80	99.10

Figure 1: Clustering Accuracy Comparison of Four Approaches

It is clear from the figure 1 that the clustering accuracy of the proposed EMFPCM is considerably higher than the other methods like FCM, EMFCM, and FPCM in all the four datasets.

4.1.2 Execution Time

Table 2 shows the comparison of execution time for clustering techniques in all four datasets. Execution time value of the proposed EMFPCM is very low since it uses EM Algorithm.

Table 2: Execution Time Comparison of four Approaches

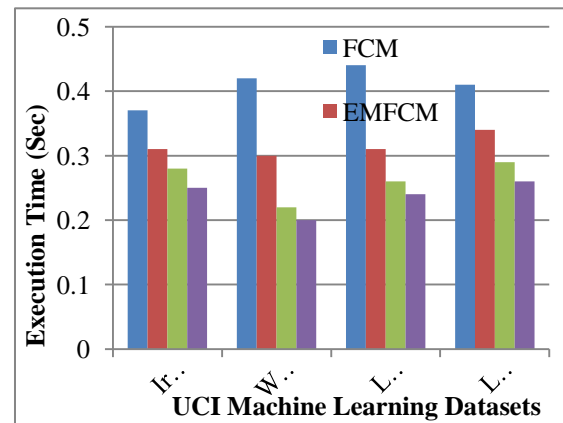| Number of Clusters	FCM	EMFCM	FPCM	EMFPCM
Iris Dataset	0.37	0.31	0.28	0.25
Wine Dataset	0.42	0.30	0.22	0.20
Lung Cancer Dataset	0.44	0.31	0.26	0.24
Lymphograma Dataset	0.41	0.34	0.29	0.26

Figure 2: Execution Time Comparison of four Approaches

It is revealed from the figure 2 that the execution time of the proposed EMFPCM is very low when compared with the other three approaches like FCM, EMFCM, FPCM.

4.1.3 Convergence Behaviour

If the clustering approach is very effective, it should take only less number of iterations for its convergence. The convergence behaviour resulted for the proposed and other approaches using four dataset are provided in table 3.

Table 3: Convergence Behaviour Comparison of four Approaches

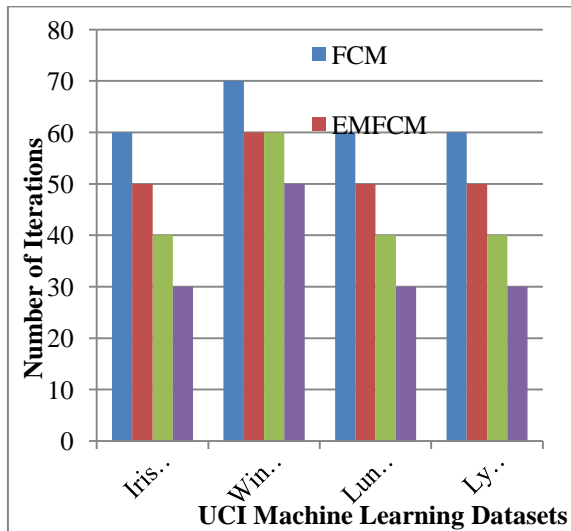| Number of Clusters	FCM	EMFCM	FPCM	EMFPCM
Iris Dataset	60	50	40	30
Wine Dataset	70	60	60	50
Lung Cancer Dataset	60	50	40	30
Lymphograma Dataset	60	50	40	30

Figure 3: Convergence Behaviour of the four Approaches

From the figure 3, it is revealed that EMFPCM approach takes lesser number of iterations for its convergence in all the four datasets whereas the other three approaches take more number of iterations to converge.

It is clear from the experimental results that the performance of the proposed approach of EMFPCM is better in terms of clustering accuracy, mean squared error, execution time and convergence behaviour than the standard approaches like FCM, EMFCM, and FPCM.

5. CONCLUSION

This research focuses on the effective clustering techniques for data clustering. Data clustering has become an important research area in the field of data mining. In order to improve the clustering accuracy, the proposed approach is called as Enhancement of Fuzzy Possibilistic C-Means Algorithm using EM Algorithm (EMFPCM). With the help of EMFPCM, noise is reduced, provides more accuracy and thus provides better result in predicting the user behavior. The algorithm was implemented and the experiment result proves that this method is very effective in predicting user behavior. The performance of the proposed approach is evaluated on UCI machine repository datasets namely iris, wine, lung cancer and lymphograma. It is observed from the experimental results that the proposed EMFPCM method outperforms the other approaches in terms of accuracy, Mean Squared Error,

Execution Time and Convergence Behaviour. Thus, the proposed EMFPCM approach is best suited for the data clustering applications.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] M R. Berthold and D. J. Hand, "Intelligent Data Analysis". *Springer-Verlag*, Berlin, Germany 1999.

[2] M. W. Berry, "Survey of Text Mining", *Springer-Verlag*, New York,NY, USA 2003.

[3] N. R. Pal, K. Pal and J. C. Bezdek, "A mixed c-means clustering model", *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems,* Vol. 1, pp. 11-21, Jul. 1997.

[4] Chu-Hui Lee and Yu-Hsiang Fu, "Web Usage Mining Based on Clustering of Browsing Features", Eighth International Conference on Intelligent Systems Design and Applications, Vol. 1, Pp. 281-286, 2008.

[5] J. Srivastava, R. Cooley, M. Deshpande and PN. Tan, "Web usage mining: discovery and applications of usage patterns from web data," SIGKDD Explorations, Vol. 1, No. 2, pp.12–23, 2000.

[6] A. Vakali, J. Pokorný and T. Dalamagas, "An Overview of Web Data Clustering Practices", EDBT Workshops, pp. 597-606, 2004.

[7] P. Zhang, X. Wang, and P. X. Song, "Clustering categorical data based on distance vectors," The Journal of the American Statistical Association, Vol. 101, No. 473, pp. 355–367, 2006.

[8] P.S. Bradley and U.M. Fayyad, "Refining Initial Points for K-Means Clustering", ACM, Proceedings of the 15th International Conference on Machine Learning, Pp. 91-99, 1998.

[9] Jiabin Deng, JuanLi Hu, Hehua Chi and Juebo Wu, "An Improved Fuzzy Clustering Method for Text Mining", Second International Conference on Networks Security Wireless Communications and Trusted Computing (NSWCTC), Vol. 1, Pp. 65–69, 2010.

[10] Celikyilmaz A and Burhan Turksen I, "Enhanced Fuzzy System Models With Improved Fuzzy Clustering Algorithm", IEEE Transactions on Fuzzy Systems, Vol. 16, No. 3, Pp. 779–794, 2008.

[11] Pal N.R, Pal K, Keller J.M. and Bezdek J.C, "A Possibilistic Fuzzy c-Means Clustering Algorithm", IEEE Transactions on Fuzzy Systems, Vol. 13, No. 4, Pp. 517–530, 2005.

[12] Neal, Radford; Hinton, Geoffrey (1999). Michael I. Jordan. ed."A view of the EM algorithm that justifies incremental, sparse, and other variants". Learning in Graphical Models (Cambridge, MA: MIT Press): 355–368.

[13] J.C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics Vol. 3, Pp. 32-57, 1973.

[14] Bezdek, J, "Pattern Recognition with Fuzzy Objective Function Algorithms", New York: Plenum Press, 1981.

[15] Jain, A. K. And Dubs, R.C. Algorithms for clustering data (N. J. & Cliffs; Prentice Hall) in 1999.