ISSN No. 0976-5697

Volume 3, No. 5, Sept-Oct 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Fuzzy Similarity Statistics Self-Constructing Clustering Algorithm for Text Classification with Geodesic Distance

Mrs. C. Rathika and Miss. V.Karthika*
M.Sc., M.PhilAsst. Prof, Department of Computer Science,
*M.ScResearch scholar, M.Phil in Computer Science,
Sri Ramakrishna College of Arts and Science for Women, Coimbatore-44, India
rathikaprakash@gmail.com,vairakarthika@gmail.com*

Abstract:In text classification, Feature clustering is a powerful method to reduce the dimensionality of feature vectors. In existing method a fuzzy similarity-based self-constructing feature clustering algorithm used for text classification. By this algorithm the derived membership functions match closely and describe properly the real distribution of the training data. The proposed method is to present a fuzzy statistical similarity measure instead of using the fuzzy set. The proposed method used fuzzy mean deviation and develops a fuzzy statistical similarity measure (FSS) in evaluating the similarity between the feature vectors. It can merge cluster centers to extract land-cover information by FSS. Fuzzy statistics is a subject based on the combination of fuzzy set theory and statistical methods. Fuzzy set theory is the basis in studying membership relationships from the fuzziness of the phenomena. The similarity is a negative value therefore a small value is equivalent to a large similarity. By introducing fuzzy mean deviation into similarity measure, it can exploit fuzzy sets in decision making. The FSS can take into account the difference of the same band between texts. The proposed method used the geodesic distance to measure distance between clusters. Geodesic distance is used to calculate a shortest path between the distance points for clustering methods. It reflects true embedded manifold, various cluster prototypes also can be used the distance measure. The proposed approaches are able to handle the data on a low dimensional manifold of the feature space. It does the clustering in the original feature space. The Proposed method shows the method can run better accuracy compared with the existing method.

Keywords:Fuzzy similarity statistics, feature clustering, feature extraction, feature reduction, text classification, Geodesic distance

I. INTRODUCTION

Feature clustering is a powerful method to reduce the dimensionality of the feature vectors for text classification. In text classification, the dimensionality of the feature vector is very high. For example, 20 Newsgroups have 20,000 more than documents and 20,000 more than features. In high dimensionality the classification algorithms can be difficult. So to avoid the difficulty, feature reduction approaches are applied before classification [6]. Feature reduction method has been proposed two types of the major approaches (i) Feature Selection and (ii) Feature Extraction. Feature selection is particularly important for data sets with large numbers of features. Feature selection has been widely studied in the context of supervised learning where the ultimate goal is to select features that can achieve the highest accuracy on unseen data.

Feature selection has received comparatively very little attention in unsupervised learning or clustering. Feature selection for classification is a well-researched problem, aimed at reducing the dimensionality and noise in data sets. In General,feature extraction approaches are more effective than the feature selection techniques but it is more computationally expensive. Many scalable and efficient extraction methods was developed for high-dimensional documents data sets. Classical feature extraction methods aim to convert the representation of the original high-dimensional data set into a lower-dimensional data set by a projecting process through algebraic transformations. For example, Principal Component Analysis [11], Linear Discriminant Analysis [12], and Orthogonal Centroid algorithm [13] perform the projection by linear transformations, while Locally Linear Embedding [14], ISOMAP [15], and LaplacianEigenmaps [16] do feature extraction by nonlinear transformations. In practice, linear algorithms are in wider use due to their efficiency. Several scalable online linear feature extraction algorithms [7], [17], [18], [19] have been proposed to improve the computational complexity. However, the complexity of these approaches is still high. Feature clustering [8], [10], [3], [9], [20] is one of effective techniques for feature reduction in text classification.In existing method has been proposed the fuzzy self-constructing algorithm based feature clustering for text classification. The idea of feature clustering is to group the original features into clusters with a high degree of pair wise semantic relatedness. Each cluster is treated as a single new feature, and, thus, feature dimensionality can be drastically reduced. A fuzzy self-constructing feature clustering algorithm is an incremental clustering approach to reduce the dimensionality of the features in text classification. Features that are similar to each other are grouped into the same cluster. Each cluster is characterized by a membership function with statistical mean and deviation.In the proposed method have been proposed a novel clustering method called fuzzy statistics- based affinity propagation which is based on a fuzzy statistical similarity measure (FSS) to extract the feature reduction.

FSS allows the algorithm to assign proper memberships to uncertain data. It can avoid poor solutions caused by unlucky initializations and hard decisions and can save a significant amount of execution time for getting optimal results. The obtained classification accuracy, and general accuracy indices related to the proposed FSS clustering method are always higher than yielded by the other considered algorithms. The proposed method has been proposed the Geodesic distancesto measure the distance between clusters. Geodesic distance is to find the shortest path and to calculate the distance. In the Geodesic distance used for the location based information. So the proposed method can run faster and obtain better extracted features than other methods.The remainder of this paper is organized as follows: Section 2 gives a related works about feature reduction. Section 3 presents self constructing feature clustering algorithm. Section 4 presents the proposed work fuzzy statistics and geodesic distance. Experimental results are presented in Section 5. Finally, we conclude this work in Section 6.

II. RELATED WORKS

In machine learning, dimension reduction is the process of reducing the number of random variables under consideration, and can be divided into feature selection and feature extraction.Feature selection approaches try to find a subset of the original variables (also called features or attributes). Feature clustering is an efficient approach for feature reduction [10], [9], which groups all features into some clusters, where features in a cluster are similar to each other. The feature clustering methods proposed in [8], [3], are “hard” clustering methods, where each word of the original features belongs to exactly one word cluster. Feature extraction transforms the data in the high-dimensional space to a space of fewer dimensions. The data transformation may be linear, as principal component analysis (PCA), but many nonlinear dimensionality reduction techniques also exist **[7]**.

E.F. Combarro says Linear Measures for Feature Selection in Text Categorization to select relevant features by means of a family of linear filtering measures which are simpler than the usual measures applied for the purpose. Text Categorization **[4]** which consists of automatically assigning documents to a set of categories, usually involves the management of a huge number of features. Most of them are irrelevant and others introduce noise which could mislead the classifiers. Thus, feature reduction is often performed in order to increase the efficiency and effectiveness of the classification. In the proposed work, to select relevant features by means of a family of linear filtering measures which are simpler than the usual measures applied for the purpose. It carries out experiments over two different corpora and finds that the proposed measures perform better than the existing ones.

Jung-Yi Jiang proposedSelf Constructing Feature clustering algorithm is an incremental, self-constructing learning approach **[1].** Word patterns are considered one by one. The user does not need to have any idea about the number of clusters in advance. No clusters exist at the beginning, and clusters can be created if necessary. For each word pattern, the similarity of this word pattern to each existing cluster is calculated to decide whether it is combined into an existing cluster or a new cluster is created. Once a new cluster is created, the corresponding membership function should be initialized. When the word pattern is combined into an existing cluster, the membership function of that cluster should be updated accordingly.

Huan Liu (2005) proposedToward Integrating Feature Selection Algorithms for Classification and Clustering, it introduces concepts and algorithms of feature selection, surveys existing feature selection algorithms for classification and clustering, groups and compares different algorithms with a categorizing framework based on search strategies, evaluation criteria, and data mining tasks, reveals unattempted combinations, and provides guidelines in selecting feature selection algorithms. In the work concludes work by identifying trends and challenges of feature selection research and development**[5].**

L.D. Baker and **A. McCallum (1998)** say Distributional Clustering to document classification**[2].** The proposed approach clusters words into groups based on the distribution of class labels associated with each word. Unlike some other unsupervised dimensionality reduction techniques such as Latent Semantic Indexing. To compress the feature space much more aggressively, while still maintaining high document classification accuracy.

I.S. Dhillon (2003) proposed Divisive Information-Theoretic Feature Clustering Algorithm for Text Classification a new information theoretic divisive algorithm for feature/word clustering and apply it to text classification**[3].** Existing techniques for such “distributional clustering” of words are agglomerative in nature and result in (i) sub-optimal word clusters and (ii) high computational cost. A divisive algorithm that directly minimizes this objective function, converging to a local minimum and the algorithm minimizes the within-cluster Jensen-Shannon divergence, and simultaneously maximizes the between-cluster Jensen-Shannon divergence.

III. EXISTING WORK

There are many algorithm are used for feature clustering method. In the existing work has a fuzzy similarity-based self-constructing feature clustering algorithm, which is an incremental feature clustering approach to reduce the number of features for the text classification task. Most of the clustering method proposed the desired number of extracted features, has to be specified in advance. This gives a burden to the user, since trial-and-error has to be done until the appropriate number of extracted features is found and when calculating similarities, the variance of the underlying cluster is not considered. Intuitively, the distribution of the data in a cluster is an important factor in the calculation of similarity. Then all words in a cluster have the same degree of contribution to the resulting extracted feature. These issues are reduced by this approach. The words in the feature vector of a document set are represented as distributions, and processed one after another. Each cluster is characterized by a membership function with statistical mean and deviation.

This system derived membership functions match closely and describe properly the real distribution of the training data. Here, trial-and-error for determining the appropriate number of extracted features can then be avoided. The author proposed self-constructing learning approach. Word patterns are considered one by one. The user does not need to have any idea about the number of clusters in advance. No clusters exist at the beginning, and clusters can be created if necessary. For each word pattern, the similarity of this word pattern to each existing cluster is calculated to decide whether it is combined into an existing cluster or a new cluster is created. Cluster points are calculated by the Euclidean distance. Euclidean distance calculated by this formula,

$$\mathbf{d}(\mathbf{p},\mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}.$$

Once a new cluster is created, the corresponding membership function should be initialized.On the contrary, when the word pattern is combined into an existing cluster, the membership function of that cluster should be updated accordingly.By this algorithm, the derived membership functions match closely with and describe properly the real distribution of the training data.

IV. PROPOSED WORK

The proposed work has a novel clustering method called fuzzy statistical similarity which exhibits a fast execution speed and finds clusters with small error, particularly for large datasets. Fuzzy statistics is a subject based on the combination of fuzzy set theory and statistical methods. Fuzzy set theory is the basis in studying membership relationships from the fuzziness of the phenomena.FSS can get objective estimates of how closely two feature vectors resemble each other. This method simultaneously considers all data points to be equally suitable as initial feature, thus reducing the dependence of the final clustering from the initialization. The proposed fuzzy mean deviation and then develop a fuzzy statistical similarity measure in evaluating the similarity between data sets to increase the system performance.

A document set D of n documents d1, d2; . . . ; dn, together with the feature vector W of m words w1; w2; . . . ; wm and p classes c1, c2; . . . ; cp, We construct one word pattern for each word in W. For word wi, its word pattern xi is defined,

xi =<xi1; xi2; . . . ; xip>

$= P(c_1|w_i), P(c_2|w_i), \ldots, P(c_p|w_i)$, →**(1)**

Where,

$$P(c_j|w_i) = \frac{\sum_{q=1}^{n} d_{qi} \times \delta_{qj}}{\sum_{q=1}^{n} d_{qi}},$$ →**(2)**

For $1 \le j \le p$. Note that dq i indicates the number of occurrences of wi in document dq,

The goal is to group the words in W into clusters, based on these word patterns. A cluster contains a certain number of word patterns, and is characterized by the product of p one-dimensional Gaussian functions. Gaussian functions are adopted because of their superiority over other functions in performance.

Let G be a cluster containing q word patterns

$x_1, x_2, \ldots, x_q$.Let $x_j = < x_{j1}, x_{j2}, \ldots, x_{jp} >$, $1 \le j \le q$,the mean $m = < m_1, m_2, \ldots, m_p >$ and the deviation of the $\sigma = < \sigma_1, \sigma_2, \ldots, \sigma_p >$ defined as,

$$m_i = \frac{\sum_{j=1}^{q} x_{ji}}{|G|}$$

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{q} (x_{ji} - m_{ji})^2}{|G|}}$$ →**(3)**

For $1 \le i \le p$, where |G| denotes the size of G, i.e., the number of word patterns contained in G. The fuzzy similarity of a word pattern x =<x1; x2; . . . ; xp> to cluster G is defined by the following membership function,

$$\mu_G(\mathbf{x}) = \prod_{i=1}^{p} \exp\left[-\left(\frac{x_i - m_i}{\sigma_i}\right)^2\right].$$ →**(4)**

A word pattern close to the mean of a cluster is regarded to be very similar to this cluster. Each word pattern, the similarity of this word pattern to each existing cluster is calculated to decide whether it is combined into an existing cluster or a new cluster is created. Once a new cluster is created, the corresponding membership function should be initialized. When the word pattern is combined into an existing cluster, the membership function of that cluster should be updated accordingly, The similarity of xi to each existing clusters, i.e.,(4)

$$\mu_{G_j}(\mathbf{x}_i) = \prod_{q=1}^{p} \exp\left[-\left(\frac{x_{iq} - m_{jq}}{\sigma_{jq}}\right)^2\right]$$ →**(5)**

Fuzzy similarity statistics are also calculated after the membership calculation

FSS = $s(i, k) = disi(k)$ $i = 1, 2, \ldots, n$;

$k = 1, 2, \ldots, m$; $i \ne k$ →**(6)**

and the preference is calculated using

$s(i, i) = \min - \text{CTS}(\max - \min)$, $i = 1, 2, \ldots, n$

Fuzzy set, fuzzy mean distance is the amount of difference between individual feature vectors and clustering vector. Mean distance deviation and membership function is a measure of difference for interval and ratio variables between the distance value and the mean and the membership degree depends on distance deviation in the space. Cluster threshold scalar (CTS) is used to get the expected number of clusters through setting the appropriate value. The similarity is a negative value; a small value is equivalent to a large similarity.

We say that xi passes the similarity test on cluster Gj if

$$\mu_{G_j}(\mathbf{x}_i) \ge \rho,$$ →**(7)**

$0 \le \rho \le 1$,is a predefined threshold.

xi is not similar enough to any existing cluster and a new cluster Gh, h = k + 1, is created with

mh =xi; $\sigma_h = \sigma_0$ →**(8)**

Where σ_0=<σ0; . . . ; σ0 > is a user-defined constant vector. The deviation of a new cluster is 0, since it contains only one member. We cannot use zero deviation in the calculation of fuzzy similarities. Therefore, we initialize the deviation of a newly created cluster by σ0, as indicated in (6). Of course, the number of clusters is increased by 1 and the size of cluster Gh, Sh, should be initialized, i.e.,

k =k +1; Sh =1: →**(9)**

If there are existing clusters on which xi has passed the similarity test, let cluster Gt be the cluster with the largest membership degree, i.e.,

$$t = arg \max_{1 \le j \le k} (\mu_{G_j}(\mathbf{x}_i)).$$ →**(10)**

The regard xi to be most similar to cluster Gt and mt and σt of cluster Gt should be modified to include xi as its member. The modification to cluster Gt is described as follows:

$$m_{tj} = \frac{S_t \times m_{tj} + x_{ij}}{S_t + 1},$$ →**(11)**

Equations can be derived easily from (3). Note that k is not changed in this case. Suppose x1 is most similar to cluster G1, and the size of cluster G1 is 4 and the initial deviation σ 0 is 0.1. Then cluster G1 is modified as follows:

St= St + 1 →**(12)**

Steps for proposed Algorithm

Step: 1 Initialization:

no of original word patterns: m

number classes: p

Threshold: ρ

Initial deviation: σ0

Initial of clusters: k = 0

Step: 2 Input:

xi =<xi1; xi2; . . . ; xip>, 1 ≤ i≤ m

Step: 3 Output:

Clusters G1, G2; . . .Gk

Step: 4 *Procedure* Self-Constructing-Clustering-Algorithm

Step: 5 *for* each word pattern xi, 1 ≤ i ≤ m

Step: 6 *temp* _ W ={Gj |μGj, (xi) I ≥ρ,1 ≤ j≤k;

Step: 7 *if (temp_* W==ø)

Step: 8 FSS = *s*(*i*, *k*) = *disi*(*k*)

Step: 9 A new cluster Gh, h = k+ 1, is created by (3)

Step: 10*else* let Gt € temp _W be the cluster to which xi is closest by (10);

Step: 12Incorporate x i into Gt by (11)**-**(12);

endif;

end for;

*return*with the created *K* clusters;

endprocedure

This system consists of some statistical characteristics which are based on some fuzzy characteristics. Fuzzy set, fuzzy mean distance is the amount of difference between individual feature vectors and clustering vector. Mean distance deviation and membership function is a measure of difference for interval and ratio variables between the distance value and the mean and the membership degree depends on distance deviation in the space. Cluster threshold scalar (CTS) is used to get the expected number of clusters through setting the appropriate value. The similarity is a negative value; a small value is equivalent to a large similarity. By introducing fuzzy mean deviation into similarity measure, it can exploit fuzzy sets in decision making. Based on the above operations the system can avoid poor solutions caused by unlucky initializations and hard decisions and can save a significant amount of execution time for getting optimal results. Geodesic distance is measure the data points between the clusters. Following steps are calculated the geodesic distance measure.

Step: 1 Construct a neighborhood graph:

Define a graph G over all N data points by connecting *xi* and *xj*if their distance *dE*(*xi*, *xj*) is closer thanεor if *xi* is one of the *k* nearest neighbors of *xi* or vice versa. Set the edge lengths equal to *dE* (*xi*, *xj*).

Step:2 Compute the shortest paths:

Initialize *dG* (*xi*, *xj*) = *dE* (*xi*, *xj*) if *xi* and *xi* are linked by an edge, otherwise *dG* (*xi*, *xj*) = ∞. For each *k* = 1, · · ·,*N*, replace all entries *dG*(*xi*, *xj*) by min(*dG*(*xi*, *xj*), *dG*(*xi*, *xk*) +*dG*(*xk*, *xj*)). DG = $(dG(xi, nj))^{n}_{j} = 1$contains the shortest-path distances between all point pairs in *G*.

The method considers that geodesic distance can be fairly approximated by Euclidean distance for a neighborhood of points; while for faraway points geodesic distance is estimated by the shortest path through neighbor points. Geodesic distance measure in a fuzzy clustering method and have achieved better results than using Euclidean distance.

V. PERFORMANCE RESULTS

In this section, we present experimental results to show the effectiveness of our fuzzy Statistics similarity based feature clustering method with geodesic distances.Compared with existing approach, the results are more efficiency in the proposed method. Based on the **accuracy**, the two tables are calculated as follows: The Table Irepresent the data that are used for accuracy of sensitivity of the algorithm, specificity of the algorithm and the frequency of the algorithm.

Table 1: Performance comparison of the approaches

Approaches	**Precision value (In values)**	**Recall value (In values)**	**F_measure value (In values)**
Fuzzy similarity Self-Constructing algorithm with Euclidean distance (existing method)	59.99	61.292	78.115
Fuzzy Similarity Statistics**(FSS)** (proposed method)	69.436	67.766	80.665
FSS with **Geodesic Distance** (proposed method)	**76.831**	**73.151**	**86.033**

The Table II represents the data that are used for algorithm accuracy and time.

Table 2: Time Comparison of the approaches

Approaches	**Accuracy (%)**	**Time (in millisec)**
Information Gain (existing method)	71.566	10,975
Fuzzy similarity Self-Constructing Algorithm with Euclidean distance (existing method)	76.568	9,047
Fuzzy Similarity Statistics **(FSS)** (proposed method)	84.256	6,624
FSS with **Geodesic Distance** (proposed method)	**92.678**	**5,558**

The experimental results demonstrate that fuzzy statistics similarity with geodesic distances has higher accuracy compared with the existing method.

Fig (a) shows performance of precision value, in Fuzzy Statistics Similarity performs much better existing method. Precision value is the sensitivity of the algorithm to be calculated. Precision is the fraction of retrieved documents that are relevant to the search:

Precision value calculated by this formula,

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

To compare the existing work, the precision value is more efficient than the existing algorithm.

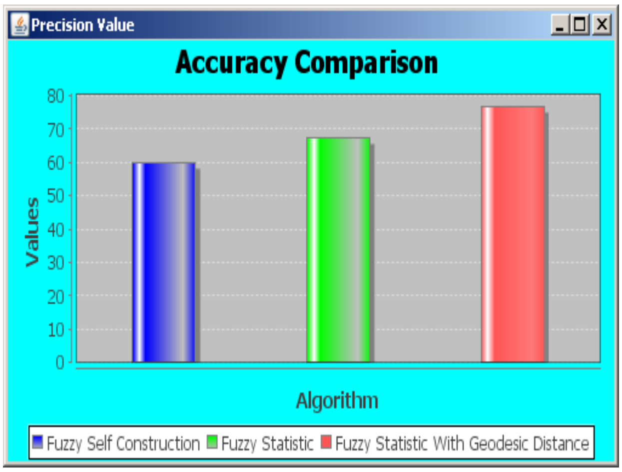


Figure (a) precision value Accuracy

Fig (b) shows performance of Recall value, in Fuzzy Statistics Similarity performs much better existing method. Recall value is the productivity of the algorithm to be calculated.

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

Recall value calculated by this formula,

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{total relevant documents}\}|}$$

Fig (b) To compare the existing work, the Recall value accuracy is more efficient method to retrieved documents than the existing algorithm.

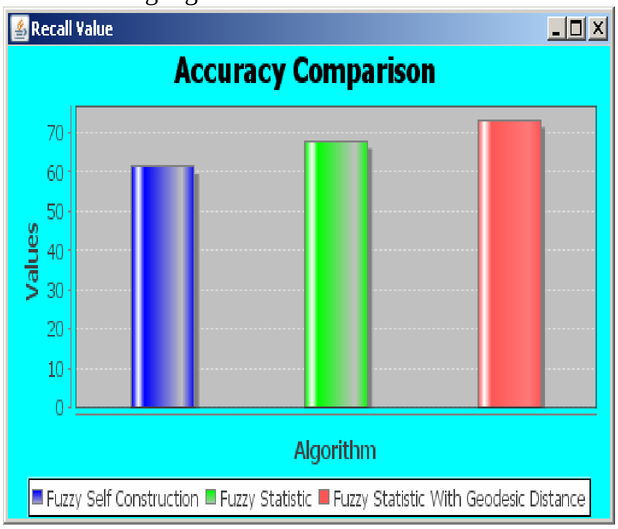


Figure (b) Recall value Accuracy

Fig (c) shows performance of f_measure value, in Fuzzy Statistics Similarity performs much better existing method. F_measure value is the frequency of the algorithm to be calculated. A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

F_measure value calculated by this formula,

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Fig (c) To compare the existing work, thef_measurevalue accuracy is more efficient method to retrieved documents than the existing algorithm.

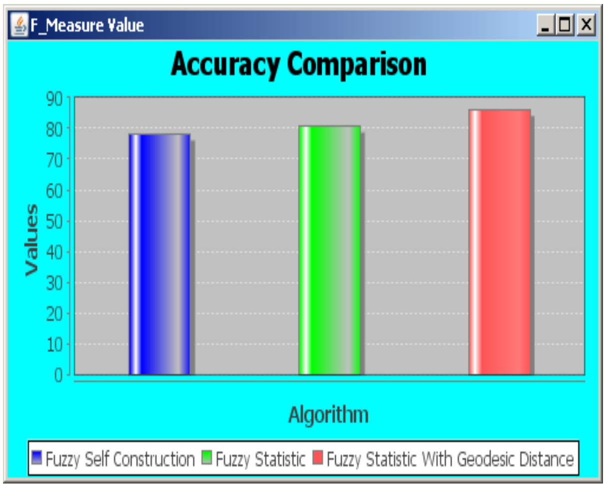
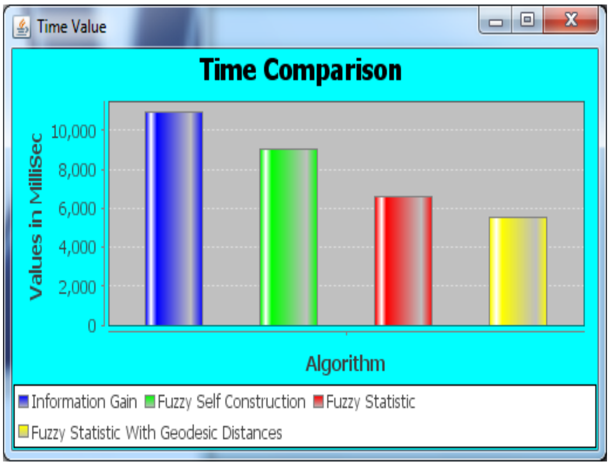


Figure (c) f-measure Accuracy

Fig (d) shows performance of time, in Fuzzy Statistics Similarity performs much better than existing method. Time is calculated upon the location based cluster.

The result shows the less time of proposed method compared between the existing methods.



Figure(d) Time Accuracy

The proposed Fuzzy statistics Similarity with Geodesic distance method is too compared with existing method. It reduces the time, improving the accuracy of the algorithm.

VI. CONCLUSION AND FUTURE WORK

It concludes Fuzzy Statistics Similarity (FSS) allows the algorithm to assign proper memberships to uncertain data and can get an accurate and objective estimate of how closely two attributes resemble each other. Geodesic distance is also used to measure the distance, which reflects the true embedded manifold. Each cluster ischaracterized by a membership function with statisticalmean and deviation. To calculate the Fuzzy statistics after the membership function. All the Words have been fed in, a desired number of clusters are formed automatically. Then have one extracted feature for each cluster. Geodesic distance, it measure distances between the cluster data point. Our method is good for text classification problems due to the suitability of the distributional word clustering method and the distribution of data than the real groups within the data. Similarity-based clustering is one of the techniques have developed in our machine learning research. In the proposed work, it applies this clustering technique to text categorization problems. In future, it also applying it to other problems, such as image segmentation, data sampling, fuzzy modeling, web mining, etc

VII. REFERENCES

[1]. Jung-Yi Jiang, Ren-JiaLiou, and Shie-Jue Lee, "A Fuzzy Self-Constructing Feature Clustering Algorithm for Text Classification" Member, IEEE transactions on knowledge and data engineering, vol. 23, no. 3, march 2011

[2]. L.D. Baker and A. McCallum, "Distributional Clustering of Words for Text Classification," Proc. ACM SIGIR, pp. 96-103, 1998.

[3]. I.S. Dhillon, S. Mallela, and R. Kumar, "A Divisive Information- Theoretic Feature Clustering Algorithm for Text Classification," J. Machine Learning Research, vol. 3, pp. 1265-1287, 2003

[4]. E.F. Combarro, E. Montan e´s, I. Dı´az, J. Ranilla, and R. Mones, "Introducing a Family of Linear Measures for Feature Selection inText Categorization," IEEE Trans. Knowledge and Data Eng., vol. 17,no. 9, pp. 1223-1232, Sept. 2005.

[5]. L.S. HuanLiu, Toward Integrating Feature SelectionAlgorithms for Classification and Clustering, Senior Member,IEEE

[6]. B.Y. Ricardo and R.N. Berthier,Modern Information Retrieval.Addison Wesley Longman, 1999.

[7]. J. Yan, B. Zhang, N. Liu, S. Yan Q. Cheng, W. Fan, Q. Yang, W. Xi,and Z. Chen, "Effective and Efficient for Large- Scale and Streaming Data Preprocessing," IEEE transaction knowledge and Data Eng., vol. 18, 3 pp. 320-333, Mar. 2006.

[8]. L.D. Baker and A. McCallum, "Distributional Clustering of Words for Text Classification," Proc. ACM SIGIR, pp. 96-103, 1998.

[9]. N. Slonim and N. Tishby, "The Power of WordClusters for Text Classification," Proc. 23rd European Colloquium on Information Retrieval Research (ECIR), 2001.

[10]. R. Bekkerman, R. El-Yaniv, N. Tishby, and Y. Winter, "Distributional Word Clusters versus Words for Text Categorization," J. Machine Learning Research, vol. 3, pp. 1183-1208, 2003.

[11]. I.T. Jolliffe, Principal Component Analysis. pringer-Verlag, 1986.

[12]. A.M. Martinez and A.C. Kak, "PCA versus LDA," IEEE Trans.Pattern Analysis and Machine Intelligence, vol. 23, no. 2 pp. 228-233, Feb. 2001.

[13]. H. Park, M. Jeon, and J. Rosen, "Lower DimensionalRepresentation of Text Data Based on Centroids and Least Squares," BIT Numerical Math, vol. 43, pp. 427-448, 2003.

[14]. S.T. Roweis and L.K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science, vol. 290, pp. 2323-2326, 2000.

[15]. J.B. Tenenbaum, V. de Silva, and J.C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science, vol. 290, pp. 2319- 2323, 2000.

[16]. M. Belkin and P. Niyogi, "LaplacianEigenmaps and Spectral Techniques for Embeddingand Clustering," Advances in Neural Information Processing Systems, vol. 14, pp. 585-591, The MIT Press 2002.

[17]. K. Hiraoka, K. Hidai, M. Hamahira, H. Mizoguchi, T. Mishima, and S. Yoshizawa, "Successive Learningof Linear Discriminant Analysis: Sanger-Type Algorithm," Proc. IEEE CS Int'l Conf.Pattern Recognition, pp. 2664-2667, 2000.

[18]. J. Weng, Y. Zhang, and W.S. Hwang, "Candid Covariance-Free Incremental Principal Component Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pp. 1034-1040, Aug. 2003.

[19]. J. Yan, B.Y. Zhang, S.C. Yan, Z. Chen, W.G. Fan, Q. Yang, W.Y. Ma, and Q.S. Cheng, "IMMC: Incremental Maximum Margin Criterion," Proc. 10th ACM SIGKDD, pp. 725-730, 2004.

[20]. D. Ienco and R. Meo, "Exploration and Reduction of the Feature Space by Hierarchical Clustering," Proc. SIAM Conf. Data Mining, pp. 577-587, 2008

