



## **DESIGN OF SIMILARITY MEASURE UNDER MULTI VIEW POINT CLUSTERING**

**M.Naveen<sup>1</sup>, A.Mallikarjuna Reddy<sup>2</sup>**

<sup>1</sup>M.Tech Student, Dept of CSE, Anurag Group of Institutions (formerly CVSR College of Engineering),  
Hyderabad, T.S, India

<sup>2</sup>Assistant Professor, Dept of CSE, Anurag Group of Institutions (formerly CVSR College of Engineering),  
Hyderabad, T.S, India

### **ABSTRACT:**

Clustering is one of the prominent methods mainly used for the classification of the data in a well oriented fashion respectively. Here the classification takes place by the analysis of the multi variant based strategy under which it is related to the application of the algorithm based on the design oriented strategy of the objects plays a crucial role respectively. Here the coupling of the objects oriented datasets takes place based on the well effective scenario of the similarity of the datasets in a well oriented fashion respectively. Here the classification of the data can be done in either of the ways that is done both in the form of the explicit followed by the implicit in a well stipulated fashion respectively. Here a new mechanism is designed under which it is related to the design oriented well effective strategy under which it is related to the design based strategy under which it is related to the specification so far proper implementation of the data with respect to the similarity of the multi viewpoint based data classification followed by the methods of the clustering oriented measures plays a crucial role in its implementation based strategy in a well effective manner respectively. There is a lot of variation take place in the system in terms of the measures related to the score of the similarity or the dissimilarity under the viewpoint of the single strategy is a major concern respectively. There is proper utilization of the well implemented mechanism under which it is related to the design based strategy where there is an accurate analysis of the system which is related to the classification in an implicit

the explicit basis and the accuracy is maintained in terms of the classification is a major concern respectively. Here the consideration of the viewpoint under the strategy of the multi phenomena basis where the assessment of the information followed by the score of the similarity is a major concern respectively. Experiments have been conducted on the present method under which the test bed is conducted with respect to the large number of the dataset in a well oriented fashion under which there is an accurate analysis of the system in terms of the improvement in the performance followed by the outcome of the entire system in a well oriented fashion respectively.

**KEYWORDS:** *Data classification, clustering, Morphological operation, clustering of the document, measure of the similarity, mining of the text.*

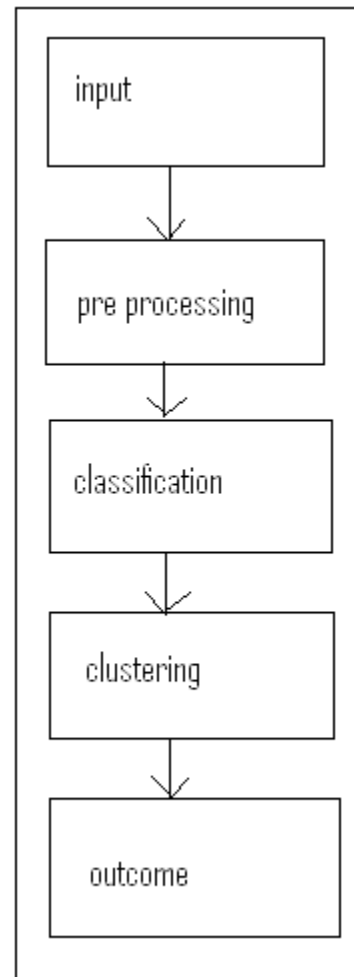
## 1. INTRODUCTION:

There is an advancement that takes place in the classification of the data previously it is done by the help of the retrieval of the data by the help of some basic fundamental principle or the properties involvement that take place in the system then after the segmentation followed by the morphological and the extension of this particular phenomena is a clustering is a major concern respectively [1]. Here there is a lot of research that takes place in the mining of the data under the well effective scenario of the clustering is a major concern respectively. Here depending on the structures of the intrinsic based characteristics the classification of the data is a major concern

with respect to the efficient integration of the clustering strategy where the similarity of the features the classification of the data done by the similarity concern respectively. Here the publishing of the algorithms plays a crucial role relative to the clustering respectively [9][10]. IN the research oriented strategy under which there is a lot of innovation that takes place in the system in terms of the technique of the clustering under the differential strategy and the well defined approaches point of view respectively. Here the survey takes place in the system where there is an accurate classification of the data with respect to the well effective scenario of the various clustering based approaches and some of them are K means, cleans, modified c means

followed by the well effective pillar k means plays a crucial role in its implications respectively[7][8]. Similarity of the multi viewpoint strategy followed by the similarity score and correlation respectively. Here depending on the application oriented perspective under which there are a lot of varieties in the clustering based approaches and then depending on the scenario of its implacability there is a minute variation in the approach of the process of the data respectively. Here the variation of the system in terms of the K means followed by the C means is the accurate analysis of the system takes place followed by the well effective process of the data take place in the C means clustering as of the comparison to that of the K means clustering is a major concern respectively. Whereas in the modified C means clustering there is a reduction of the complexity takes place in the system respectively.

## BLOCK DIAGRAM



**Fig 1: Shows the block diagram of the present method respectively**

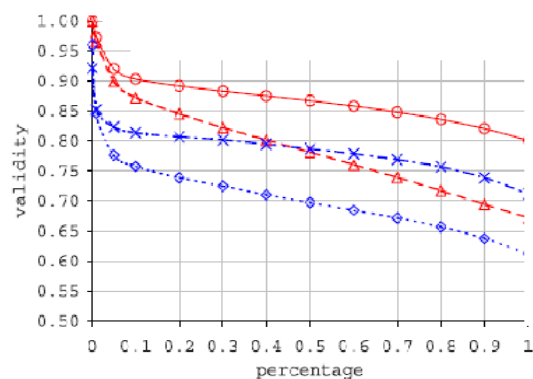
## 2. METHODOLOGY:

In this paper a new technique is proposed and is shown by the above figure in the form of the block diagram and is explained in an elaborative fashion respectively. Here the algorithm is implemented based on the study of the

characteristics of the data both with respect to the features of the intrinsic followed by the extrinsic based features in a well oriented fashion respectively. Here the application and the process of the system takes place by the clustering based aspect under the effective strategy oriented integration of the analysis of the multivariate strategy respectively [4][5]. Here the classification of the data under the effect of the classification of the document by the help of multivariate clustering based approach in well oriented fashion respectively. Here the a light is depicted in the brief representation and includes that is illustrated as follows. Here the representation of the system initially the complete documents are taken in to the system then the next step is the pre processing under which there is an effective removal of the noise that is the elimination of the unwanted noise takes place in this system respectively. Then the features are applied that is depending on the desired properties the classification of the data plays a crucial role in its application based strategy under which the feature extracted as per the requirement of the developer plays a crucial role in its application based strategy in a well oriented fashion respectively [2].

Finally the clustering technique is applied to the system of the data that is well oriented in terms of the documents this is the latest advancement of the feature extraction techniques and then the system is well effective by the properties of the multivariate feature respectively. Here we finally conclude that the present method is effective and efficient in term of the improvement in the performance followed by the outcome of the entire system in well stipulated fashion respectively.

### 3. EXPECTED RESULTS:



**Fig 2: Shows the graphical representation of the present method respectively**

Here a technique is proposed by the help of the clustering based on the features includes with the relativity of the multivariate strategy plays a crucial role for the effective extraction of the data in a well oriented fashion respectively. Here the

comparative analysis is made between the present methods to that of the several previous methods in well oriented fashion respectively. Here the comparison of the system is shown by the above figure in the form of the graphical representation where there is an evaluation of the system in terms of the performance followed by the outcome of the entire system in a well oriented fashion respectively.

#### 4. CONCLUSION:

In this paper a new technique is proposed by the help of the design oriented characteristics of the multi variant features of the clustering where the classification of the data that is here the documents are classified by the help of the above set features and the feature are taken by the help of the multi variant features respectively. Here the design of the present method is ineffective and inaccurate in terms of the performance and also in terms of the reduction of the complexity is a major concern respectively.

#### REFERENCES

[1] A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, "Clustering on the Unit Hypersphere Using Von Mises-Fisher Distributions," *J. Machine Learning Research*, vol. 6, pp. 1345-1382, Sept. 2005.

[2] W. Xu, X. Liu, and Y. Gong, "Document Clustering Based on Non-Negative Matrix Factorization," *Proc. 26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*, pp. 267-273, 2003.

[3] I.S. Dhillon, S. Mallela, and D.S. Modha, "Information-Theoretic Co-Clustering," *Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD)*, pp. 89-98, 2003.

[4] C.D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*. Cambridge Univ. Press, 2009.

[5] C. Ding, X. He, H. Zha, M. Gu, and H. Simon, "A Min-Max Cut Algorithm for Graph Partitioning and Data Clustering," *IEEE Int'l Conf. Data Mining (ICDM)*, pp. 107-114, 2001.

[6] H. Zha, X. He, C. Ding, H. Simon, and M. Gu, "Spectral Relaxation for K-Means Clustering," *Proc. Neural Info. Processing Systems (NIPS)*, pp. 1057-1064, 2001.

[7] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 22, no. 8, pp. 888-905, Aug. 2000.

[8] I.S. Dhillon, "Co-Clustering Documents and Words Using Bipartite Spectral Graph Partitioning," *Proc. Seventh ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD)*, pp. 269-274, 2001.

[9] Y. Gong and W. Xu, *Machine Learning for Multimedia Content Analysis*. Springer-Verlag, 2007.

[10] Y. Zhao and G. Karypis, "Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering," *Machine Learning*, vol. 55, no. 3, pp. 311-331, June 2004.