iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IX **Month of publication:** September 2017

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue IX, September 2017- Available at www.ijraset.com

Comparison on Distance Measures of Clustering Techniques for Finding Similarity in Articles

Deeksha[1], Shashank Sahu[2]
[1,2] Department of Computer Science & Engineering Ajay Kumar Garg Engineering College, Ghaziabad, India

Abstract: Clustering separates articles into groups using some similarity criteria. The purpose of clustering is to cluster articles such that articles in a same group are more comparable to each other. This paper discusses numerous clustering techniques such as Hierarchical, K-means, K-medoids and Fuzzy C-means (FCM) clustering. Here research work is to evaluate numerous distance measures of clustering techniques and find out the suitable distance measure. The research work is begins with selecting various categories of articles. Each category contains 50 articles of various sizes. These articles are taken from various news channel websites. Then most common search words are selected for each category, respectively. Now for each category, calculate frequencies of selected words in the article then calculate the distance between frequencies of words using various distance measures of clustering techniques. For an experiment Matlab is used and the results show that in Hierarchical clustering Euclidean distance measure, in K-means clustering Correlation distance measure, in K-medoids clustering City block distance measure and in the FCM Chebychev distance measure provides better results than other distance measures.
***Keywords**: Clustering, Hierarchical clustering, K-means clustering, K-medoids clustering, Fuzzy C-means clustering, Distance measure.*

I. INTRODUCTION

The gathering process usually referred to as clustering, attempts to partition articles with the goal that those appointed to the same group share common characteristics, while those appointed to other groups are conceptually dissimilar. Presenting the information by fewer bunches inevitably loses certain fine facts, however accomplishes simplification. It models, information by its clusters [1].
Now a day's internet is utilized by most of the people to get information from it. The main source to get the information from the internet is blog's and news channels' websites, huge information is available on numerous news channels websites and user wants only some selected and up-to-date news, according to their needs that's why user go through different news channels websites so rather than use different news channel's website user can get news in one place. There is some way via which people can get the preferred news from the internet such as RSS (rich site summary), RSS is a strategy for delivering [2] consistently evolving web contents. But the issue in research is to find out similarity in the articles with the goal that RSS supply best outcomes.
Here the research work is to compare several distance measures of clustering techniques and discover the best one among them. There are two sorts of clustering: Hard Clustering and Soft Clustering. Hierarchical, K-means and K-medoids clustering techniques goes under hard clustering and Fuzzy C-means clustering goes under fuzzy or soft clustering. In hard clustering, articles are assigned only in one group while in fuzzy clustering; articles can allocate more than one group [3][4].

A. Hierarchical Clustering

Hierarchical clustering (HC) separates the articles into a hierarchy of bunches. This hierarchy of groups is made either bottom-top or top-bottom manner [5]. The bottom-top approach is known as agglomerative HC and top-bottom approach is known as divisive HC.
Agglomerative HC (AHC) starts with assigning articles to their own group and then comparable groups are iteratively combined until the preferred group hierarchy is obtained [5] while Divisive HC (DHC) is the reverse of AHC. In DHC, all articles belong to a single group and repeatedly split into sub groups until all articles are belong to their own singleton group [6].
This paper uses 7 distance measures of Hierarchical clustering i.e. 'City block', 'Euclidean', 'Cosine', 'Hamming', 'Correlation', 'Jaccard' and 'Chebychev'. The functional description of these distance measures are as follows -

1) *City block Distance:* The City block distance [7] between two objects h_x & h_y is calculated as follows-

$$D_{xy} = \sum_{j=1}^{n} \left| h_{xj} - h_{yj} \right|$$

2) *Euclidean Distance:* The Euclidean distance [7, 8] between two objects h_x and h_y is calculated as follows -

$$D_{xy}^2 = (h_x - h_y)(h_x - h_y)'$$

3) *Cosine Distance :* The Cosine Distance [7, 8] is calculated as follows –

$$D_{xy} = 1 - \frac{h_x h_y'}{\sqrt{(h_x h_x')(h_y h_y')}}$$

4) Hamming Distance : The Hamming distance [8] between two objects h_x and h_y is calculated as follows –

$$D_{xy} = (\#(h_{xj} \neq h_{yj}) / n)$$

5) Correlation Distance The correlation distance [8] between two objects h_x and h_y is calculated as follows –

$$D_{xy} = 1 - \frac{(h_x - \bar{h}_x)(h_y - \bar{h}_y)'}{\sqrt{(h_x - \bar{h}_x)(h_x - \bar{h}_x)'}\sqrt{(h_y - \bar{h}_y)(h_y - \bar{h}_y)'}}$$

Where $$\bar{h}_x = \frac{1}{n}\sum_j h_{xj} \text{ and } \bar{h}_y = \frac{1}{n}\sum_j h_{yj}$$

6) *Jaccard Distance:* The Jaccard distance [8, 9] between two objects h_x and h_y is calculated as follows –

$$D_{xy} = \frac{\#[(h_{xj} \neq h_{yj}) \cap ((h_{xj} \neq 0) \cup (h_{tj} \neq 0))]}{\#[(h_{xj} \neq 0) \cup (h_{yj} \neq 0)]}$$

7) *Chebychev Distance :* The Chebychev distance [8] between two objects h_x and h_y is calculated as follows –

$$D_{xy} = \max_j \{ \left| h_{xj} - h_{yj} \right| \}$$

B. K-means Clustering

K-means [10] is a partitioning procedure and it's based on minimizing squared error. K-means starts with deciding k number of groups and after that compute k group centroids. These centroids are primary centroids and must be computed precisely because diverse position demonstrates totally distinctive results. Next step is to consider all articles and allot all articles to its nearest centroids. When all articles are allotted or no articles are left, then re-figure the new k group centroids and re-allocate articles to its nearest centroids. This strategy is repeated until the point that no more centroids change their arrangement [11].

This paper uses four distance measures of K-means clustering i.e. 'City block', 'Correlation', 'Cosine' and 'Sqeuclidean'. The functional description of City block, Cosine and Correlation is same as discuss above in hierarchical clustering. Functional description of Sqeuclidean as follows –

1) *Sqeuclidean Distance :* The Sqeuclidean distance [12] between two objects h_x and h_y as follows –

D(x, y) = (x-y)(x-y)'

C. K-medoids Clustering

K-medoids [10] is also a partitioning procedure and based on minimizing squard error. Aversion to the K-means clustering, K-medoids picks real objects as medoids. K-medoids clustering starts by arbitrarily choosing k articles as medoids which represent k groups and non-selected articles are allotted to the group whose medoids nearer to them. Now again re-figure new medoids which signify groups in improved manner and again allot the articles to the groups whose medoids are nearer to them. This strategy is repeated until the point that no more medoids change their arrangement [13].

In this paper, we have shown five distance measures of K-medoids clustering i.e. 'Sqeuclidean', 'Euclidean', 'Cosine', 'Correlation' and 'City block' and functional description of these distance measures are same as discussed above in Hierarchical and K-means clustering.

D. Fuzzy C-means Clustering

FCM [14] is a fuzzy clustering technique which is based on minimization of the objective function. The objective function is given below-

$$J_m(U, c; X) = \sum_{j=1}^{n}\sum_{k=1}^{c}(u_{k,j})^m d^2(x_j, c_k)$$

Where n is a number of articles, c denotes the cluster center, u_{kj} is the degree of membership, m is the fuzzifier and d (x_j, c_k) is the distance between j^{th} article and center of k^{th} cluster.

FCM starts with a randomly initialize membership matrix U [15] i.e. U= [u_{ij}] and then computes the group centers which is calculated as –

$$c_k = \frac{\sum_{i=1}^{N} u_{ik}^m * x_i}{\sum_{i=1}^{N} u_{ik}^m}$$

After computing group center next step is to update membership matrix U as follows-

$$u_{ij} = \frac{1}{\sum_{t=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_t\|}\right)^{\frac{2}{m-1}}}$$

If $\max_{ij}\{|u_{ij}^{t+1} - u_{ij}^{t}|\} < \varepsilon$ then stop otherwise repeat procedure.

For FCM, we use three distance measures that are city block, Euclidean and Chebychev and functional description of these distance measures are same as discussed above in HC and K-means clustering.

The rest of the paper arranged as follows- section II signifies related work of the clustering techniques using numerous distance measuring methods, section III represents experiments and results. At the end of the paper conclusion of our work is presented.

II. RELATED WORKS

Jafar & Sivakumar [4] provides comparative studies of FCM and K-means using Chebyshev, Chi-square and σ-distance measuring methods and these clustering algorithms are evaluated on four data sets via cluster validity indices .Partition coefficient and partition entropy indices are used to evaluate clustering algorithms. The results of this paper show, FCM algorithm using Chi-square distance measure provides preferable outcome over K-means algorithm for all four data sets while Chebyshev distance measuring method show maximum partition coefficient and minimum partition entropy than other distance measuring methods for several data sets.

Shruti Kapil et. al. [11] evaluates the performance of K-means clustering and compared numerous distance metrics. This paper utilizes two distance metrics: Euclidean and Manhattan distance. These distance metrics are compared and observed that the Euclidean distance perform better than Manhattan.

Usha et. al. [16] compared numerous distance measuring methods of Hierarchical and K-means clustering for finding out the similarities between articles. This paper used five categories which comprise 28 articles that are taken from different news channels and concluded that City block distance measuring method in Hierarchical and Correlation distance measuring method in K-means shows better outcomes.

Saratha Sathasivam et. al. [17] analyize distance measuring methods of FCM. These distance measuring methods are Euclidean and City block and observed that Euclidean shows better performance in term of quality of cluster and execution time. Archana Singh et. al. [18] implemented K-means clustering using Euclidean (basic K-means), Manhattan and Minkowski distance metrics and found that Euclidean shows better performance than others distance metrics.

III. EXPERIMENTS AND RESULTS

In this paper, all clustering techniques with numerous distance measures are implemented in Matlab. Here research work is to find out which distance measure is best and to compare these distance measures of clustering techniques, three evaluation measures are used. These evaluation measures are Cophenetic Correlation Coefficient, Silhouette Index and Xie and Beni Index.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue IX, September 2017- Available at www.ijraset.com

To implement our work we select nine categories such as 'Agriculture', 'Business', 'Crime', 'Education', 'Election', 'Entertainment', 'Health', 'Game' and 'Weather'. All these categories contain 50 articles of varying size (100-500 words in each article). These articles are taken from various news channel websites such as CNBC, Business News and Star Cricket News etc. Then different numbers of most common search words of each selected category has been chosen, i.e. 22, 35, 32, 49, 25, 30, 31, 35 and 27 words are selected against each category respectively.

After selecting the categories and most common search words for the experiment next work is to apply various clustering techniques using numerous distance measures. Firstly, for each category, we developed a program that automatically counts the frequencies of selected words in the articles. Once frequencies of the words are calculated then distance between them are calculated by using numerous distance measuring methods of Hierarchical clustering, K-means clustering, K-medoids clustering and FCM clustering.

There are some basic steps to perform our research work are as follows-

Steps 1 – Calculate frequencies of the selected words in the articles and a matrix of frequencies of words are generated. Steps 2- Apply various clustering techniques using numerous distance measures on matrix.

Step 3- Evaluate and compare numerous distance measures of clustering techniques using evaluation measures.

There are three evaluation measures such as Cophenetic Correlation Coefficient, Silhouette Index and Xie and Beni Index is used. To evaluate and compare distance measures of HC we use the cophenetic correlation coefficient .Nearer the value of the cophenetic correlation coefficient [19] [20] is to 1, more precisely clustering is done.

To evaluate and compare several distance measures of K-means and K-medoids clustering, Silhouette Index is used. Silhouette index value [21] is varying from -1 to 1. When the value is more than 0.5 means clustering is done more precisely and to compare distance measures of FCM clustering, Xie and Beni Index is used. Minimum the value of Xie and Beni index [22][23], more precisely clustering is performed.

A. Results

This section discusses the results that are obtained after applied several distance measuring methods of clustering techniques .Table I shows distance measures that are used in HC such as 'City block', 'Euclidean', 'Cosine', 'Hamming', 'Correlation', 'Jaccard' and 'Chebychev'. Table II shows distance measuring methods that are used in K-means clustering. These distance measures are: 'City block', 'Correlation', 'Cosine' and 'Sqeuclidean'. Table III shows distance measures of K-medoids clustering i.e. 'Sqeuclidean', 'Euclidean', 'Cosine', 'Correlation' and 'City block'. Table IV shows distance measures of FCM clustering that are 'Euclidean', 'City block' and 'Chebychev'. The row of the tables I, II, III and IV signifies categories, while column of the tables signifies distance measures. The values in table I represent the cophenetic correlation coefficient and the mean cophenetic correlation coefficient values are 0.8572, 0.8729, 0.5878, 0.7853, 0.5783, 0.4721 and 0.8370.The outcomes show that Euclidean distance measuring methods showing higher mean cophenetic correlation coefficient value which is closer to 1 so the Euclidean distance measure is better than other distance measures of HC. Fig 1 shows a chart diagram on mean values of distance measures of HC. In the diagram, horizontal axis signifies distance measures while vertical axes signify the range of the cophenetic correlation coefficient. Figure clearly shows that the Euclidean distance method showing higher value of the cophenetic correlation coefficient so the Euclidean Distance measure is better than other distance measures of HC.

TABLE I
COMPARISONS OF DIFFERENT DISTANCE MEASURES OF HIERARCHICAL CLUSTERING IN ALL NINE CATEGORIES

Categories/Distance Measures	*City block*	*Euclidean*	*Cosine*	*Hamming*	*Correlation*	*Jaccard*	*Chebychev*
Agriculture	0.9159	0.9140	0.6282	0.8972	0.6348	0.5384	0.9207
Business	0.8540	0.9387	0.5479	0.7543	0.4991	0.5470	0.9009
Crime	0.8422	0.8840	0.4987	0.8014	0.5366	0.3844	0.8488
Education	0.7970	0.7655	0.4861	0.7913	0.4930	0.3881	0.6537
Election	0.8802	0.9114	0.7067	0.7350	0.6838	0.4481	0.8887
Entertainment	0.8526	0.9145	0.5024	0.6507	0.4741	0.5699	0.8807
Game	0.7766	0.7429	0.7314	0.7810	0.6596	0.4183	0.7381
Health	0.8847	0.8742	0.5573	0.8578	0.5639	0.4436	08437
Weather	0.9118	0.9105	0.6312	0.7988	0.6596	0.5114	0.8577
Mean Value	0.8572	0.8729	0.5878	0.7853	0.5783	0.4721	0.8370

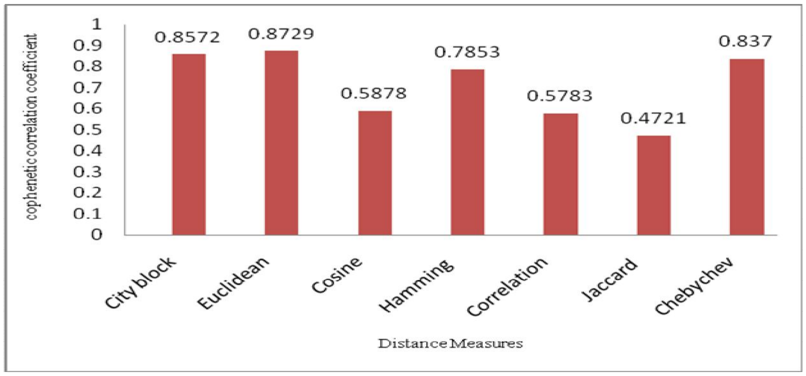


FIG 1: MEAN VALUES OF DSTANCE MEASURES OF HIERARCHICAL CLUSTERING

Table II represents a comparison between distance measures of K-means clustering. The values of the table II signifies Silhouette values and the mean silhouette values are 0.5268, 0.5401, 0.5192 and 0.5357. The outcomes of the table II show that Correlation distance measures showing higher mean silhouette value so Correlation distance measuring method is better than other distance measures methods of K-means clustering. Fig 2 shows a chart diagram on mean values of distance measures of K-means clustering. The horizontal axis of fig 2 signifies distance measures while the vertical axis signifies range of Silhouette index. Fig 2 clearly shows that Correlation distance is better than other distance measures.

TABLE III

COMPARISONS OF DIFFERENT DISTANCE MEASURES OF K-MEANS CLUSTERING IN ALL NINE CATEGORIES

Categories/Distance Measures	City block	Correlation	Cosine	Sqeuclidean
Agriculture	0.5212	0.5594	0.5289	0.5587
Business	0.5391	0.5456	0.5167	0.5132
Crime	0.5289	0.5458	0.5015	0.5358
Education	0.5733	0.5623	0.5347	0.5054
Election	0.5125	0.5033	0.5534	0.5500
Entertainment	0.5071	0.5538	0.5019	0.5502
Game	0.5139	0.5253	0.5055	0.5485
Health	0.5217	0.5352	0.5207	0.5051
Weather	0.5234	0.5305	0.5095	0.5545
Mean Value	0.5268	0.5401	0.5192	0.5357

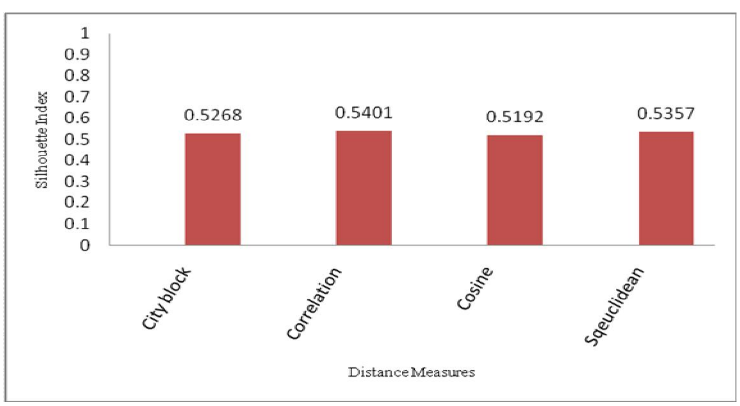


FIG 2: MEAN VALUES OF DISTANCE MEASURES OF K-MEANS CLUSTERING

Table III represents a comparison between distance measures of K-medoids clustering . The values of the table III signifies Silhouette values and mean values of silhouette index are 0.5724, 0.5125, 0.5140, 0.5139 and 0.5165. The outcomes of the table III show that Sqeuclidean distance measures is better than other distance measures methods of K-medoids clustering because the mean value of silhouette index in Sqeuclidean is higher than other distance measures . Fig 3 shows chart diagram on mean values of distance measures of K-medoids clustering. The horizontal axis of fig 3 signifies distance measures while the vertical axis signifies the range of Silhouette index. Fig 3 clearly shows that Seuclidean distance is better than other distance measures.

TABLE IIIII

COMPARISONS OF DIFFERENT DISTANCE MEASURES OF K-MEDOIDS CLUSTERING IN ALL NINE CATEGORIES

Categories/Distance Measures	Sqeuclidean	Euclidean	Cosine	Correlation	City block
Agriculture	0.7447	0.5263	0.5040	0.5258	0.5019
Business	0.5333	0.5174	0.5189	0.5009	0.5019
Crime	0.5053	0.5005	0.5211	0.5054	0.5113
Education	0.5065	0.5013	0.5179	0.5029	0.5092
Election	0.5032	0.5157	0.5249	0.5103	0.5190
Entertainment	0.5070	0.5050	0.5076	0.5175	0.5534
Game	0.5051	0.5308	0.5227	0.5009	0.5260
Health	0.5281	0.5050	0.5045	0.5518	0.5124
Weather	0.8185	0.5108	0.5043	0.5094	0.5137
Mean Value	0.5724	0.5125	0.5140	0.5139	0.5165

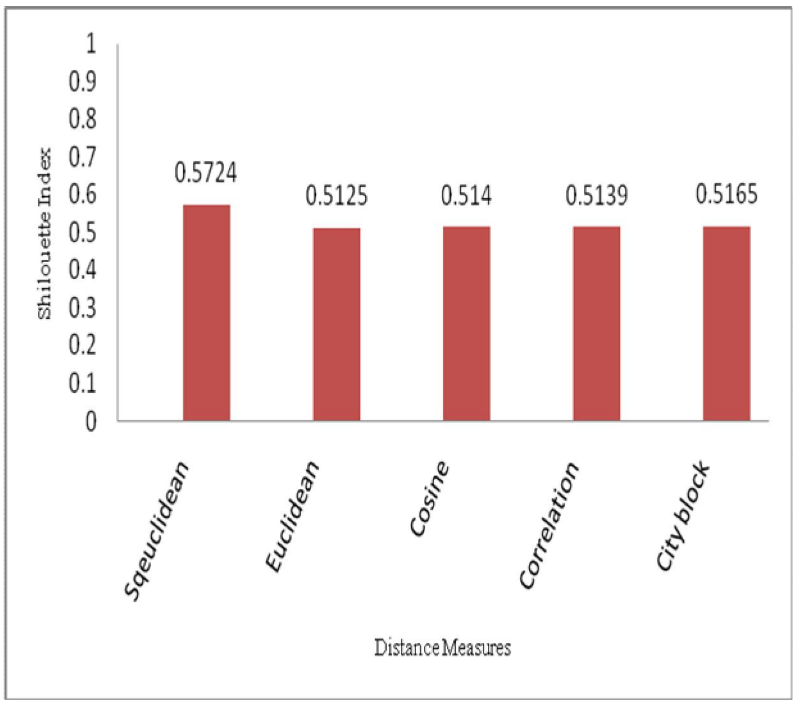


FIG 3: MEAN VALUES OF DISTANCE MEASURES OF K-MEDOIDS CLUSTERING

Table IV represents a comparison between distance measures of FCM clustering . The values of the table IV signifies Xie and Beni index values and the mean values of Xie and Beni index are 4.1654e+06 , 2.5131e+12 and 1.3061e+06. The outcomes of the table V shows that Chebychev distance measures showing the minimum mean value of Xie and Beni index, which is better than other distance measures methods of FCM clustering.Fig 4 shows chart diagram on mean values of distance measures of FCM clustering. The horizontal axis of fig 4 signifies distance measures while the vertical axis shows the values of Xie and Beni index. Fig 4 clearly shows that Chebychev distance showing minimum value which is better than other distance measures.

TABLE IV
COMPARISONS OF DIFFERENT DISTANCE MEASURES OF FCM CLUSTERING IN ALL NINE CATEGORIES

Categories/Distance Measures	Euclidean	City block	Chebychev
Agriculture	0.1641	0.1893	0.0989
Business	3.1076e+03	1.5208e+12	2.7066e+04
Crime	1.1167	2.8173e+07	0.5602
Education	7.0606e+06	1.2820e+12	4.8094
Election	8.7804e+03	2.0863e+07	1.6292
Entertainment	2.7112e+07	1.1123e+09	1.1122e+07
Game	3.3043e+06	7.1288e+07	4.3048e+05
Health	23.5543	1.9814e+13	1.7566e+05
Weather	2.4978	9.5624e+07	1.6477
Mean Value	4.1654e+06	2.5131e+12	1.3061e+06

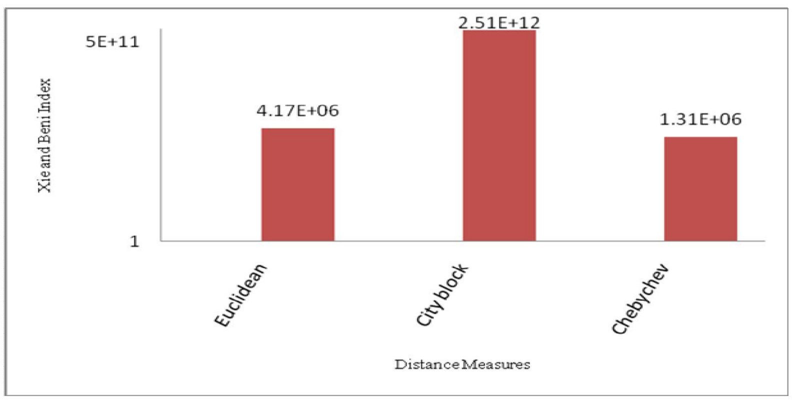


FIG 4: MEAN VALUES OF DISTANCE MEASURES OF FCM CLUSTERING

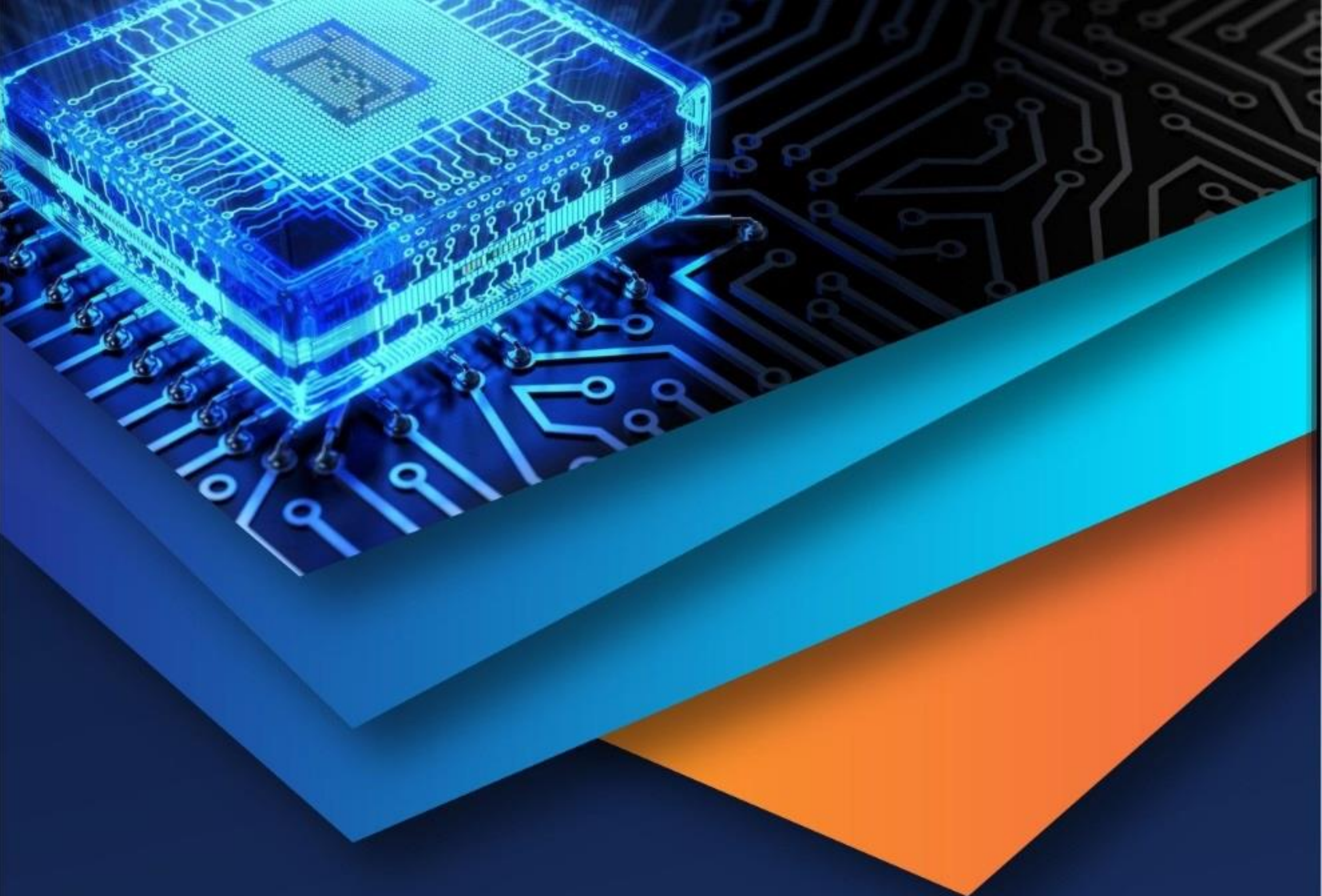
IV. CONCLUSION

This paper uses numerous clustering techniques with several distance measures to find out similarities between articles. For the implementation, we have selected nine categories that are agriculture, business, crime, education, election, entertainment, game, health and weather. These categories contain 50 articles of various sizes. These articles are taken from several news channel websites such as CNBC, Business News and Star Cricket News etc. The results of table II ,III ,IV and V shows that in the HC Euclidean distance , in K-means Correlation distance ,in K-medoids Sqeuclidean distance and in the FCM Chebychev distance measure provides better outcomes than other distance measures. Further, in the future, we will try to extend more meaningful words in our categories for the experiment and try to utilize other clustering techniques as well.

REFERENCES

[1] Biswas, Gautam, Jerry B. Weinberg, and Douglas H. Fisher. "ITERATE: A Conceptual Clustering Algorithm for Data Mining." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 28, Issue 2, pp. 219-230, 1998.
[2] Kevin Curran, Sheila Mc Kinney, "Delivering Audio and Video with Rich Site Summary Enclosures", Journal of Multimedia, vol. 1, issue 1, pp.24-31, APR. 2006
[3] Ashish Goel, "A Study of Different Partitioning Clustering Technique", IJSRD - International Journal for Scientific Research & Development, vol. 2, Issue 08, pp.128-130, 2014.
[4] O. A. Mohamed Jafar, R. Sivakumar, "A Comparative Study of Hard and Fuzzy Data Clustering Algorithms with Cluster Validity Indices", Elsevier, pp. 775-782,2013.
[5] Zahra Nazari, Dongshik Kang, M.Reza Asharif,*,"A New Hierarchical Clustering Algorithm", Artificial Intelligence, Robotics, and Human-Computer Interaction, Okinawa, Japan ICIIBMS, IEEE, vol. 575, pp.148-152, 2015.
[6] Dipak P Dabhi, Mihir R Patel, "Extensive Survey on Hierarchical Clustering Methods in Data Mining", International Research Journal of Engineering and Technology (IRJET), vol.03, issue, 11, pp. 659-665, Nov -2016.
[7] http ://www.mathworks.in/help/stats/pdist.html.

[8] Dibya Jyoti Bora, Dr. Anil Kumar Gupta, “Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab”, (IJCSIT) International Journal of Computer Science and Information Technologies, vol. 5, issue 2, pp. 2501-2506, 2014.
[9] Ashish Moon ,T. Raju, “A Survey on Document Clustering with Similarity Measures”, International Journal of Advanced Research in Computer Science and Software Engineering ,vol. 3, issue 11, pp. 599-601,Nov. 2013.
[10] Tagaram Soni Madhulatha, "Comparison between K-Means and K-Medoids Clustering Algorithms", Springer, pp. 472-481, 2011.
[11] Shruti Kapil , Meenu Chawla, "Performance Evaluation of K-means Clustering Algorithm with Various Distance Metrics”, International Conference on Power Electronics, Intelligent Control and Energy Systems, IEEE ,pp. 1-4, 2016.
[12] https://in.mathworks.com/help/stats/kmeans.html.
[13] Preeti Arora, Dr. Deepali, Shipra Varshney, “Analysis of K-Means and K-Medoids Algorithm for Big Data”, Procedia Computer Science, vol. 78 , pp.507 – 512, 2016.
[14] Qi Zhao, Guijuan Li, Shuxia Xing, “FCM Algorithm Based on the Optimization Parameters of Objective Function Point”, 2010 International Conference on Computing, Control and Industrial Engineering, IEEE, vol. 2 , pp. 331-333, 2010.
[15] Bhavana Devi, Vivek Srivastava, "Modified Evolutionary Fuzzy Clustering”, Intl J Engg Sci Adv Research , vol. 2, issue 3, pp 5-12, 2016.
[16] Usha Rani, Shashank Sahu ,"Comparison of Clustering Techniques for Measuring Similarity in Articles",3rd IEEE International Conference on "Computational Intelligence and Communication Technology, pp 1-7, IEEE,2017.
[17] Saratha Sathasivam, Abdu Masanawa Sagir, "Comparative Study of Euclidean and City Block Distances in Fuzzy C-Means Clustering Algorithm”, International Journal of Computational and Electronics Aspects in Engineering (IJCEAE), vol. 1, Issue. 1, Dec 2014, pp 10-12.
[18] Archana Singh, Avantika Yadav, Ajay Rana, "K-means with three different Distance Metrics”, International Journal of Computer Application,s vol. 67, issue 10, April, pp. 13-17, 2013.
[19] Geetanjali, Shashank Sahu, “Similarity of Articles using Hierarchical clustering”, International Journal of Soft Computing and Engineering (IJSCE), vol.-6 Issue-3, pp.50-54, July 2016.
[20] Sinan Saraçli, Nurhan Dogan and ˙Ismet Dogan, “Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation”, Journal of Inequalities and Applications 2013, Springer, vol. 203, issue 1, pp 1-8, 2013.
[21] Peter J. Rousseeuw, ―Silhouettes: a graphical aid to the interpretation and validation of cluster analysis‖, Journal of Computational and Applied Mathematics North-Holland, Elsevier Science Publishers, Elsevier, vol. 20, pp. 53-65 ,1987.
[22] Shaoping Xu,Lingyan Hu, Xiaohui Yang,Xiaoping Liu ,"A Cluster Number Adaptive Fuzzy c-means Algorithm for Image Segmentation",International Journal of Signal Processing, Image Processing and Pattern Recognition, vol.6, issue 5 (2013), pp.191-204.
[23] Xie, X.L., Beni, G.: A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. vol. 13, issue 8, pp .841–847, 1991.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY