Engineering Universe for scientific Research and management
ISSN (Online): 2319-3069
Vol. XIII Issue XI
November 2021

A SURVEY ON IMPACT OF ENHANCED K MEANS ALGORITHM

Kanishka Sisodia[1], Chinmay Bhatt[2], Varsha Namdeo[3]
CSE, SRK University, Bhopal, India[1, 2, 3]
kanishkasisodia@gmail.com[1], chinmay20june@gmail.com[2], varsha_namdeo@yahoo.com[3]

ABSTRACT: Traditional K-means clustering algorithms have the drawback of getting stuck at local optima that depend on the random values of initial centroids. Optimization algorithms have their advantages in guiding iterative computation to search for global optima while avoiding local optima. The algorithms help speed up the clustering process by converging into a global optimum early with multiple search agents in action. Inspired by nature, some contemporary optimization algorithms which include Ant, Bat, Cuckoo, Firefly, and Wolf search algorithms mimic the swarming behavior allowing them to cooperatively steer towards an optimal objective within a reasonable time. It is known that these so-called nature-inspired optimization algorithms have their own characteristics as well as pros and cons in different applications. When these algorithms are combined with K-means clustering mechanism for the sake of enhancing its clustering quality by avoiding local optima and finding global optima, the new hybrids are anticipated to produce unprecedented performance. In this paper, we report the results of our evaluation experiments on the integration of nature-inspired optimization methods into K-means algorithms. In addition to the standard evaluation metrics in evaluating clustering quality, the extended K-means algorithms that are empowered by nature-inspired optimization methods are applied on image segmentation as a case study of application scenario.

Keywords: K-means clustering, nearest neighbor searching, clusters and data mining.

1. INTRODUCTION

The advancements in computing, along with the rapid growth and availability of data repositories, have often emphasized the task of gaining meaningful insights from the data. This encourages taking appropriate measures based on knowledge discovery approaches. Machine Learning (ML) can be broadly classified into two: supervised and unsupervised learning. In the case of supervised learning, a function is learned that maps a given input to an output based on the available input–output pairs. These learning algorithms thus require the availability of labeled data with the desired output value [1]. The availability of labeled data represents an ideal scenario; however, such datasets are often expensive and challenging to obtain. For instance, in the intrusion detection domain, zero-day attacks are rare instances and obtaining labels for them is expensive. Hence, when the labels of the datasets are unavailable, unsupervised learning approaches are typically used [2]. Under such a learning framework, the algorithm has no prior knowledge of the true labels of the dataset and tries to draw inferences from the dataset itself. A more recent and popular set of algorithms referred to as semi-supervised learning consists of algorithms that lie in between supervised and unsupervised learning. Such a learning framework makes use of labeled and unlabeled data for better inference. Existing research on semi-supervised learning corroborates that adding a skimpy amount of labeled data together with a large amount of unlabeled data produces considerable improvements in the accuracy of the predictive model. This manuscript primarily deals with different variations of the k-means algorithm, which falls under the family of unsupervised learning. Therefore, this paper will focus only on unsupervised learning algorithms.

Clustering algorithms exploit the underlying structure of the data distribution and define rules for grouping the data with similar characteristics [3]. This process results in the partition of a given dataset according to the clustering criteria without any prior knowledge about the dataset. In an ideal clustering scenario, each cluster consists of similar data instances that are quite dissimilar from the instances in other clusters. Such a dissimilarity measure relies on the underlying data and objective of the algorithm. Clustering is central to many data-driven applications and is considered an interesting and important task in machine learning. It is also studied in statistics, pattern recognition, computational geometry, bioinformatics, optimization, image processing, and in a variety of other fields [4–7]. A plethora of clustering techniques has been invented in the last decade, which is being applied in a wide range of application domains [8]. Table 1 shows the recent applications on k-means clustering [9] in different application domains.

This survey studies the problems of and solutions to partition-based clustering, and more specifically the widely used k-means algorithm [9], which has been listed among the top 10 clustering algorithms for data analysis [10]. Due to their popularity and ease of use, k-means clustering methods are being used in conjunction with deep learning for tasks such as image segmentation and handwriting recognition [11]. A more recent work in [58] used a fully connected deep convolution neural network along with k-means and performed pixel matching between a segmented image and a convoluted image. This overcomes the problem that exists when useful information from images is lost by repeated convolution of the images [10]. Although the k-means clustering algorithm itself performs well with compact and hyper-spherical clusters, we are interested in highlighting its limitations and suggesting solutions. The primary focus is given to two unavoidable problems of the k-means algorithm: (i) assignment of centroids and number of clusters and (ii) ability to handle different types of data. Although researchers have proposed variants (Figure 1) of k-means algorithms to overcome these impediments, they are however domain specific and do not generalize well. For example, a k-means algorithm which can handle categorical data might perform poorly because of the initialization process used.

2. RELATED WORK

Neha Aggarwal et al., [12] presented a new clustering method based on K-means that removes the limitation of k-means technique of having more computational complexity. The Enhanced K-Means Algorithm and Basic K-Means algorithm have been implemented on the same dataset and the Enhanced K-Means is proved to be efficient than Basic K-Means.

Ahamed Shafeeq B M et al., [13] presents a modified K-means algorithm with the intension of improving cluster quality and to fix the optimal number of cluster. The K-means algorithm takes number of clusters (K) as input from the user. But in the practical scenario, it is very difficult to fix the number of clusters in advance. The proposed method works for both the cases i.e. for known number of clusters in advance as well as unknown number of clusters. The user has the flexibility either to fix the number of clusters or input the minimum number of clusters required.

Anand M. Baswade et al., [14] presents a new k-means type algorithm called W-k-Means that can automatically weight variables based on the importance of the variables in clustering. By knowing the weight of variables we can select those variables which are good for better clustering results.

Amar Singh et al., [15] presented a cluster based ranking scheme i.e. K-means clustering over weighted page rank algorithm. Weighted Page Rank algorithm is the modified form of page rank algorithm which works on two factors, in links and out links and gives the ranks to the websites accordingly. In the previous work done, page rank algorithm is used which results to take more execution time. But we use weighted page rank algorithm with k-means which results to take less execution time as compare to the previous work done and also leads to more relevant as compared to page rank algorithm.

Harpreet Kaur et al., [16] presented an improved variant of standard k-means which performs the image compression with less running and provides more efficiency.

Victor Chukwudi Osamor et al., [17] presents a novel pearson correlation-based metric matrices k-means

(MMk-means). Experimental results show that it has a better run-time than the Traditional k-means and other variants of k-means algorithm. Azhar Rauf et al., [18] propose a new method of K-mean clustering in which we calculate initial centroids instead of random selection, due to which the number of iterations is reduced and elapsed time is improved.

Kajal C. Agrawal et al., [19] presents an improved k-means method. The main idea of algorithm is to set two simple data structures to retain the labels of cluster and the distance of all the date objects to the nearest cluster during the each iteration, that can be used in next iteration.

M.N. Vrahatis et al., [20] present an improvement of the k-means clustering algorithm, aiming at a better time complexity and partitioning accuracy. Our approach reduces the number of patterns that need to be examined for similarity, in each iteration, using a windowing technique.

Mingyu Yao et al., [21] presented an improved k-means clustering method based on isolated point problem found in experiments, and proved that the algorithm is correct and effective by experiments.

Youguo Li et al., [22] combine the largest minimum distance algorithm and the traditional K-Means algorithm to propose an improved K-Means clustering algorithm. This improved algorithm can make up the shortcomings for the traditional K-Means algorithm to determine the initial focal point.

Khaled Alsabti et al., [23] presented a novel algorithm for performing k-means clustering. Our experimental results demonstrated that our scheme can improve the direct k-means algorithm by an order to two orders of magnitude in the total number of distance calculations and the overall time of computation.

Dharmendra S. Modha et al., [24] presented a framework for integrating multiple, heterogeneous feature spaces in the k-means clustering algorithm. Our methodology adaptively selects, in an unsupervised fashion, the relative weights assigned to various feature spaces with the objective of simultaneously attaining good separation along all the feature spaces.

Siddheswar Ray et al., [25] present a simple validity measure based on the intra-cluster and inter-cluster distance measures which allows the number of clusters to be determined automatically. The basic procedure involves producing all the segmented images for 2 clusters up to K max clusters, where K max represents an upper limit on the number of clusters.

Kiri Wagstaff et al., [26] developed a general method for incorporating background knowledge in the form of instance level constraints into the k-means clustering algorithm. In experiments with random constraints on six datasets, we have shown significant improvements in accuracy.

P.S. Bradley et al., [27] extend K-means to insure that every cluster contains at least a given number of points. Using a cluster assignment step with constraints, solvable by linear programming or network simplex methods, can guarantee a sufficient population within each cluster.

Charles Elkan [28] shows how to accelerate k-means algorithm dramatically, while still always computing exactly the same result as the standard algorithm. The accelerated algorithm avoids unnecessary distance calculations by applying the triangle inequality in two different ways, and by keeping track of lower and upper bounds for distances between points and centers.

Madhu Yedla et al., [29] proposed a new method for finding the better initial centroids and to provide an efficient way of assigning the data points to suitable clusters with reduced time complexity. According to our experimental results, the proposed algorithm has the more accuracy with less computational time comparatively original k-means clustering algorithm.

Vance Faber [30] presents the continuous k-means clustering algorithm which is faster than the standard version and thus extends the size of the datasets that can be clustered. It differs from the standard version in how the initial reference points are chosen and how data points are selected for the updating process. In the standard algorithm the initial reference points are chosen more or less arbitrarily. In the continuous algorithm reference points are chosen as a random sample from the whole population of data points.

Sanpawat Kantabutra et al., [31] improves the k-means clustering algorithm by a factor of O(K/2), where K is the number of desired clusters, by applying theories of parallel computing to the algorithm. In addition to time improvement, the parallel version of K-means algorithm also enables the algorithm to run on larger collective memory of multiple machines when the memory of a single machine is insufficient to solve a problem.

K. A. Abdul Nazeer et al., [32] presents an enhanced k-means algorithm which combines a systematic method for finding initial centroids and an efficient way for assigning data points to clusters. This method ensures the entire process of clustering in O(n) time without sacrificing the accuracy of clusters.

Dan Pelleg et al., [33] presented a new K-means based algorithm that incorporates model selection. By adopting and extending algorithmic improvements to K-means, it is efficient to the extent that running it once is cheaper than looping over K with the fixedmodel algorithm. It uses statistically-based criteria to make local decisions that maximize the model's posterior probabilities.

Kohei Arai et al., [34] proposed Hierarchical K-means algorithm. It utilizes all the clustering results of K-means in certain times, even though some of them reach the local optima. Then, we transform the all centroids of clustering result by combining with Hierarchical algorithm in order to determine the initial centroids for K-means. This algorithm is better used for the complex clustering cases with large numbers of data set and many dimensional attributes.

Madhuri A. Dalal et al., [35] introduced an improved algorithm to start the k-means with good starting points (means). The good initial starting points allow k-means to converge to a better local minimum; also the numbers of iteration over the full dataset are being decreased.

Bondu Venkateswarlu et al., [36] give the improvement to the original k-means algorithm by improving the initial centroids with distribution of data. Results and discussions show that improved K-means algorithm produces accurate clusters in less computation time to find the donors information.

Er. Nikhil Chaturvedi et al., [37] proposed a new K means clustering algorithm in which we calculate the initial centroids systemically instead of random assigned due to which accuracy and time improved.

Deepika Khurana et al., [38] presents a new approach to k-Means clustering by providing a solution to initial selection of cluster centroids and a dynamic approach based on silhouette validity index. Instead of running the algorithm for different values of k, the user need to give only initial value of k as ko as input and algorithm itself determines the right number of clusters for a given dataset. The experimental results demonstrate that our proposed scheme improves the initial center selection and overall computation time.

Abhishek Patel et al., [39] presents the new approach for the k mean algorithm eliminates the deficiency of exiting k mean. It first calculates the initial centroids k as per requirements of users and then gives better, effective and stable cluster. It also takes less execution time because it eliminates unnecessary distance computation by using previous iteration.

P. Ashok et al., [40] presents the K-Means algorithm which is implemented by three distance functions and to identify the optimal distance function for clustering methods. The proposed K-Means algorithm is compared with K-Means, Static Weighted K-Means (SWK-Means) and Dynamic Weighted K-Means (DWK-Means) algorithm by using Davis Bouldin index, Execution Time and Iteration count methods.

Haitao Xu et al., [41] presents an improved k-means algorithm based on optimized simulated annealing is used to segment the stations of Hangzhou Public Bicycle System. The optimized simulated annealing (SA) algorithm is used to assign k-means initial cluster centers.

Mrs. S. Sujatha et al., [42] uses the partitioned data along the data axis with the highest variance for assigning the initial centriod for K-Means clustering.

Tapas Kanungo et al., [43] present a simple and efficient implementation of Lloyd's k-means clustering algorithm, which we call the filtering algorithm. This algorithm is easy to implement, requiring a kd-tree as the only major data structure. We establish the practical efficiency of the filtering algorithm in two ways. First, we present a data-sensitive analysis of the algorithm's running time, which shows that the algorithm runs faster as the separation between clusters increases. Second, we present a number of empirical studies both on synthetically generated data and on real data sets from applications in color quantization, data compression, and image segmentation.

Chunfei Zhang et al., [44] elaborate the method of improving the K-means clustering algorithm based on improve the initial focal point and determine the K value. Simulation experiments prove that the improved clustering algorithm is not only more stable in clustering process, at the same time, improved clustering algorithm to reduce or even avoid the impact of the noise data in the dataset object to ensure that the final clustering result is more accurate and effective.

D. Napoleon et al., [45] proposes a method for making the K-means algorithm more effective and efficient; so as to get better clustering with reduced complexity. In this research, the most representative algorithms K-Means and the Enhanced K-means were examined and analyzed based on their basic approach. The best algorithm was found out based on their performance using Normal Distribution data points. The elapsed time taken by proposed enhanced k-means is less than k-means algorithm.

Nazeer, K. A. A. et al., [46] proposes an improvement on the classic k-means algorithm to produce more accurate clusters. The proposed algorithm comprises of an O (n logn) heuristic method, based on sorting and partitioning the input data, for finding the initial centroids in accordance with the data distribution.

S. Deelers et al., [47] proposes a novel initialization algorithm of cluster centers for K-means algorithm. The algorithm was based on the data partitioning algorithm used for color quantization. A given data set was partitioned into k clusters in such a way that the sum of the total clustering errors for all clusters was reduced as much as possible while inter distances between clusters are maintained to be as large as possible.

Neha Aggarwal et al., [48] present the way to find the initial centres for the k-means so that every time the K-Means Algorithm produces same result for the same dataset.

M.V.B.T Santhi et al., [49] proposed a new method for finding the better initial centroids and to provide an efficient way of assigning the data points to suitable clusters with reduced time complexity.

Rajeev Kumar et al., [50] proposed an improved version of k-means algorithm which offers to provide a remedy of the aforesaid problem. This algorithm employs two data structures viz. red-black tree and min-heap. These data structures are readily available in the modern programming languages. While red black tree is available in the form of map in C++ and Tree Map in Java, min-heap is available in the form of priority queue in the C++ standard template library.

M. Sakthi et al., [51] provides a new technique to modify K-means clustering which can result in better performance. For initialization, this paper uses an improved version of Hopfield Artificial Neural Network (HANN) algorithm. Also, the Genetic Algorithm (GA) is in combined with K-Means algorithm.

D. Mariammal et al., [52] presents a heuristic approach for performing K-means clustering on multidimensional data based on attribute (column) with maximum range is proposed. It also proposes an improvement on the classic k-means algorithm to produce more accurate clusters.

R.Ranga Raj et al., [53] present the enhanced clustering method which improves the accuracy and efficiency. In the enhanced method data points are assigned to the clusters.

Gomathi. D et al., [54] less similarity based clustering method is proposed for finding the better initial centroids and to provide an efficient way of assigning the data points to suitable clusters with reduced time complexity.

Nidhi Gupta et al., [55] propose a new clustering algorithm that can remove the disadvantages of K-means algorithm. In Proposed algorithm we do not need to specify the value of K i.e. the number of cluster required.

Pallavi Purohit et al., [56] the new approach for the k-mean algorithm eliminates the deficiency of exiting k mean. It first calculates the initial centroids k as per requirements of users and then gives better, effective and good cluster without scarifying Accuracy. It generates stable clusters to improve accuracy. It also reduces the mean square error and improves the quality of clustering.

Fahim A M et al., [57] proposed an efficient method for assigning data-points to clusters. Fahim's approach makes use of two distance functions for this purpose- one similar to the k-means algorithm and another one based on a heuristics to reduce the number of distance calculations.

3. LIMITATIONS

K-means clustering has some of the limitations which need to get overcome. Several people got multiple limitations while working on their research with K-means algorithm. Some of the common limitations are discussed below

- **Outliers**

It has been observed by several researchers that, when the data contains outliers there will be a variation in the result that means no stable result from different executions on the same data. Outliers are such objects they present in dataset but do not result in the clusters formed. Outliers can also increase the sum of squared error within clusters. Hence it is very important to remove outliers from the dataset. Outliers can be removed by applying preprocessing techniques on original dataset.

- **Number of clusters**

Determining the number of clusters in advance is always been a challenging task for K-means clustering approach. It is beneficial to determine the correct number of clusters in the beginning. It has been observed that sometimes the numbers of clusters are assigned according to the number of classes present in the dataset. Still it is an issue that on what basis the number of clusters should be assigned.

- **Empty clusters**

If no points are allocated to a cluster during the assignment step, then the empty clusters occurs. It was an earlier problem with the traditional K-means clustering algorithm.

- **Non globular shapes and sizes**

With the K-means clustering algorithm if the clusters are of different size, different densities and non globular shapes, then the results are not optimal. There is always an issue with the convex shapes of clusters formed.

4. CONCLUSION

This paper presents an overview of the K-means clustering algorithm & various enhanced variations done on K-means clustering algorithm. K-means is the basic algorithm used for discovering clusters within a dataset. The initial point selection effects on the results of the algorithm, both in the number of clusters found and their centroids. Methods to enhance the k-means clustering algorithm are discussed in this paper. With the help of these methods efficiency, accuracy, performance and computational time are improved. Some enhanced variations improve the efficiency and accuracy of algorithm. Some methods improve clustering quality and better clustering results. Basically in all the methods the main aim is to reduce the number of iterations which will decrease the computational time.

Studies shows that K-means algorithm in clustering is widely used technique. Various enhancements done on K-mean are collected in this paper, so by using these enhancements one can build a new hybrid algorithm which will be more efficient, accurate and less time consuming than the previous work.

REFERENCES

1. Mohri, M.; Rostamizadeh, A.; Talwalkar, A. Foundations of Machine Learning; MIT Press: Cambridge, MA,USA, 2012.

2. Bishop, C.M. Neural Networks for Pattern Recognition; Oxford University Press: Oxford, UK, 1995.

3. Jain, A.K.;Murty,M.N.; Flynn, P.J. Data clustering: A review. ACMComput. Surv. 1999, 31, 264–323.

4. Ahmed, M.; Choudhury, V.; Uddin, S. Anomaly detection on big data in financial markets. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Sydney, Australia, 31 July–3 August 2017; pp. 998–1001.

5. Ahmed, M. An unsupervised approach of knowledge discovery from big data in social network. EAI Endorsed Trans. Scalable Inf. Syst. 2017, 4, 9.

6. Ahmed, M. Collective anomaly detection techniques for network traffic Analysis. Ann. Data Sci. 2018, 5, 497–512.

7. Tondini, S.; Castellan, C.; Medina, M.A.; Pavesi, L. Automatic initialization methods for photonic components on a silicon-based optical switch. Appl. Sci. 2019, 9, 1843.

8. Ahmed, M.; Mahmood, A.N.; Islam, M.R. A survey of anomaly detection techniques in financial domain Future Gener. Comput. Syst. 2016, 55, 278–288.

9. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 1 January 1967;pp. 281–297.

10. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Philip, S.Y.; et al. Top 10 algorithms in data Mining. Knowl. Inf. Syst. 2008, 14, 1–37. [CrossRef]

11. Tian, K.; Zhou, S.; Guan, J. Deepcluster: A general clustering framework based on deep learning. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Skopje, Macedonia, 18–22 September 2017; pp. 809–825.

[12] Neha Aggarwal, Kirti Aggarwal, Kanika Gupta, "Comparative Analysis of k-means and Enhanced K-means clustering algorithm for data mining", International Journal of Scientific & Engineering Research, Vol. 3, Issue 3, August-2012.

[13] Ahamed Shafeeq B M, Hareesha K S,"Dynamic Clustering of Data with Modified K-Means Algorithm", International Conference on Information and Computer Networks, Vol.27, 2012.

[14] Anand M. Baswade, Kalpana D. Joshi, Prakash S. Nalwade,"A Comparative Study of K-Means and Weighted K-Means For Clustering", International Journal of Engineering Research & Technology, Vol. 1, Issue 10, December-2012.

[15] Amar Singh, Navot Kaur,"To Improve the Convergence Rate of K-Means Clustering Over K-Means with Weighted Page Rank Algorithm", International journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 8, August 2012.

[16] Harpreet Kaur, Jaspreet Kaur Sahiwal,"Image Compression With Improved K-Means Algorithm for Performance Enhancement", International Journal of Computer Science and Management Research, Vol. 2, Issue 6, June 2013.

[17] Osamor VC, Adebiyi EF, Oyelade JO, Doumbia S "Reducing the Time Requirement of k-Means Algorithm", PLoS ONE, Vol. 7, Issue 12, 2012.

[18] Azhar Rauf, Sheeba, Saeed Mahfooz, Shah Khusro, Huma Javed,"Enhanced K-Mean Clustering Algorithm to Reduce Number of Iterations and Time Complexity", Middle-East Journal of Scientific Research, pp. 959-963, 2012.

[19] Kajal C. Agrawal, Meghana Nagori,"Clusters of Ayurvedic Medicines Using Improved K-means Algorithm", International Conf. on Advances in Computer Science and Electronics Engineering, 2013.

[20] M. N. Vrahatis, B. Boutsinas, P. Alevizos, G. Pavlides, "The New k-Windows Algorithm for Improving the k-Means Clustering Algorithm", Journal of Complexity 18, pp. 375-391, 2002.

[21] Mingyu Yao, Dechang Pi, Xiangxiang Cong,"Chinese text clustering algorithm based k-means", International Conference on Medical Physics and Biomedical Engineering, pp. 301-307, 2012.

[22] Youguo Li, Haiyan Wu,"A Clustering Method Based on K-Means Algorithm", International Conference on Solid State Devices and Materials Science, pp. 1104-1109, 2012.

[23] Khaled Alsabti, Sanjay Ranka, Vineet Singh,"An Efficient K-Means Clustering Algorithm", SURFACE, 1997.

[24] Dharmendra S. Modha, W. Scott Spangler,"Feature Weighting in k-Means Clustering", Kluwer Academic Publishers, pp. 217-237, 2003.

[25] Siddheswar Ray, Rose H. Turi,"Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation", 1998.

[26] Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schroedl,"Constrained K-means Clustering with Background Knowledge", Proceedings of the Eighteenth International Conference on Machine Learning, pp. 577-584, 2001.

[27] P. S. Bradley, K. P. Bennett, A. Demiriz,"Constrained K-Means Clustering", May 2000.

[28] Charles Elkan,"Using the Triangle Inequality to Accelerate K-means", Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003.

[29] Madhu Yedla, Srinivasa Rao Pathakota, T M Srinivasa, " Enhancing K-Means Clustering Algorithm with ImprovedInitial Center", International Journal of Computer Science and Information Technologies, Vol. 1, pp. 121-125, 2010.

[30] Vance Faber,"Clustering and the Continuous k-Means Algorithm", Los Alamos Science,1994.

[31] Sanpawat Kantabutra, Alva L. Couch,"Parallel K-means Clustering Algorithm on NOWs", Nectec Technical Journal, Vol. 1, Jan-Feb 2000.

[32] K. A. Abdul Nazeer, M. P. Sebastian,"Improving the Accuracy and Efficiency of the K-Means Clustering Algorithm", Proceedings of the World Congress on Engineering, July 1-3, 2009, London, U.K.

[33] Dan Palleg, Andrew Moore,"X-means: extending k means with efficient estimation of the number of clusters", Proceedings of the Seventeenth International Conference on Machine Learning, Palo Alto, CA, July 2000.

[34] Kohei Arai, Ali Ridho Barakbah,"Hierarchical K-means: an algorithm for centroids initialization for K-means", Reports of the Faculty of Science and Engineering, Vol. 36, 2007.

[35] Madhuri A. Dalal, Nareshkumar D. Harale, Umesh L.Kulkarni,"An Iterative Improved k-means Clustering", ACEEE International Journal on Network Security, Vol. 02, July 2011.

[36] Bondu Venkateswarlu, Prof G.S.V.Prasad Raju, "Mine Blood Donors Information through Improved K-Means Clustering", International Journal of Computational Science and Information Technology , Vol. 1, August 2013.

[37] Er. Nikhil Chaturvedi, Er. Anand Rajavat,"An Improvement in K-Mean Clustering Algorithm Using Better Time and Accuracy", International Journal of Programming Languages and Applications, Vol. 3, October 2013.

[38] Deepika Khurana, Dr. M.P.S Bhatia,"Dynamic Approach to K-means Clustering Algorithm", International Journal of Computer Engineering & Technology, Vol. 4, Issue 3, pp. 204-219, May-June 2013.

[39] Abhishek Patel, Purnima Singh,"New Approach for K-mean and K-medoids Algorithm", International Journal of Computer Applications Technology and Research, Vol. 2, Issue 1, 2013.

[40] P.Ashok, Dr.G.M Kadhar Nawaz, E.Elayaraja, V.Vadivel, "Improved Performance of Unsupervised Method by Renovated K-Means", IJASCSE, Vol. 2, Issue 1, 2013.

[41] Haitao Xu, Jing Ying, Fei Lin, Yubo Yuan,"Station Segmentation with an Improved K-Means Algorithm for Hangzhou Public Bicycle System", Academy Publisher, 2013.

[42] Mrs. S. Sujatha, Mrs. A. Shanthi Sona,"New Fast K-Means Clustering Algorithm using Modified Centroid Selection Method", International Journal of Engineering Research & Technology, Vol. 2, Issue 2, Feb 2013.

[43] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu Christine D. Piatko, Ruth Silverman, Angela Y. Wu," An Efficient K-Means Clustering Algorithm: Analysis and Implementation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, July 2002.

[44] Chunfei Zhang, Zhiyi Fang,"An Improved K-means Clustering Algorithm", Journal of Information & Computational Science, 2013.

[45] D.Napoleon, P. Ganga Lakshmi,"An Enhanced k-means algorithm to improve the Efficiency Using Normal Distribution Data Points", International Journal on Computer Science and Engineering, Vol. 2, 2010.

[46] Nazeer, K. A. A., Kumar, S. D. M., Sebastian, M. P.,"Enhancing the K-means Clustering Algorithm by Using a O(n logn) Heuristic Method for Finding Better Initial Applications of Information Technology, Feb. 2011.

[47] S. Deelers, S. Auwatanamongkol," Enhancing K-Means Algorithm with Initial Cluster Centers Derived from Data Partitioning along the Data Axis with the Highest Variance", International Journal of Electrical and Computer Engineering,2007.

[48] Neha Aggarwal, Kirti Aggarwal,"A Mid-Point based k-mean Clustering Algorithm for Data mining", International Journal on Computer Science and Engineering, Vol. 4, June 2012.

[49] M.V.B.T.Santhi, V.R.N.S.S.V.Sai Leela, P.U.Anitha, D.Nagamalleswari,"Enhancing K-Means Clustering Algorithm", IJCST, Vol. 2, Issue 4, Oct. - Dec. 2011.

[50] Rajeev Kumar, Rajeshwar Puran, Joydip Dhar,"Enhanced K-Means Clustering Algorithm Using

Red Black Tree and Min-Heap", International Journal of Innovation, Management and Technology, Vol. 2, Feb 2011.

[51] M.Sakthi, Antony Selvadoss Thanamani,"An Enhanced K Means Clustering using Improved Hopfield Artificial Neural Network and Genetic Algorithm", International Journal of Recent Technology and Engineering, Vol. 2, Issue 3, July 2013.

[52] D Mariammal, M Gowthami, N Sindhujaa,"New Algorithm to get the initial centroids of clusters on multidimensional data", International Journal of Research in Engineering & Advanced Technology, Vol. 1, Issue 1, March 2013.

[53] R.Ranga Raj, Dr. M.Punithavalli,"Evaluation of Enhanced K-Mean Algorithm to the Student Dataset", International Journal on Advanced Networking and Applications, Vol. 4, Issue 2, 2012.

[54] Gomathi.D, Tajunisha.N,"Improving the accuracy and efficiency of k-means algorithm by using less similarity based clustering technique for better initial centroids", International Journal of Computer Science and Management Research, Vol 1, Issue 4, November 2012.

[55] Nidhi Gupta, R.L.Ujjwal,"An Efficient Incremental Clustering Algorithm", World of Computer Science and Information Technology Journal, Vol. 3, 2013.

[56] Pallavi Purohit, Ritesh Joshi,"A New Efficient Approach towards k-means Clustering Algorithm", International Journal of Computer Applications, March 2013.

[57] Fahim A.M, Salem A. M, Torkey A, Ramadan M. A, "An Efficient enhanced k-means clustering algorithm", Journal of Zhejiang University, 10(7), pp. 1626–1633, 2006.

[58.] He, B.; Qiao, F.; Chen, W.; Wen, Y. Fully convolution neural network combined with k-means clustering algorithm for image segmentation. In Proceedings of the Tenth International Conference on Digital Image Processing (ICDIP 2018), Shanghai, China, 11–14 May 2018; Volume 10806, pp. 1–7.