

# A State of Art Analysis of Telecommunication Data by k-Means and k-Medoids Clustering Algorithms

T. Velmurugan

PG and Research Department of Computer Science, D. G. Vaishnav College, Chennai, India

Email: velmurugan\_dgvc@yahoo.co.in

**How to cite this paper:** Velmurugan, T. (2018) A State of Art Analysis of Telecommunication Data by k-Means and k-Medoids Clustering Algorithms. *Journal of Computer and Communications*, 6, 190-202.

<https://doi.org/10.4236/jcc.2018.61019>

**Received:** August 26, 2017

**Accepted:** December 26, 2017

**Published:** December 29, 2017

---

## Abstract

Cluster analysis is one of the major data analysis methods widely used for many practical applications in emerging areas of data mining. A good clustering method will produce high quality clusters with high intra-cluster similarity and low inter-cluster similarity. Clustering techniques are applied in different domains to predict future trends of available data and its uses for the real world. This research work is carried out to find the performance of two of the most delegated, partition based clustering algorithms namely k-Means and k-Medoids. A state of art analysis of these two algorithms is implemented and performance is analyzed based on their clustering result quality by means of its execution time and other components. Telecommunication data is the source data for this analysis. The connection oriented broadband data is given as input to find the clustering quality of the algorithms. Distance between the server locations and their connection is considered for clustering. Execution time for each algorithm is analyzed and the results are compared with one another. Results found in comparison study are satisfactory for the chosen application.

## Keywords

k-Means Algorithm, k-Medoids Algorithm, Data Clustering, Time Complexity, Telecommunication Data

---

## 1. Introduction

Data Mining (DM) is a convenient way of extracting patterns, which represents knowledge implicitly stored in large data sets and focuses on issues relating to their feasibility, usefulness, effectiveness and scalability. Data mining approach

and its technology is used to extract the unknown pattern from the large set of data for the business and real time applications. It can be viewed as an essential step in the process of knowledge discovery. Data are normally preprocessed through data cleaning, data integration, data selection, and data transformation and prepared for the mining task. Started as little more than a dry extension of DM techniques, DM is now bringing important contributions in crucial fields of investigations. Among the traditional sciences like astronomy, high energy physics, biology and medicine [1] have always provided a rich source of applications to data miners. An important field of application for data mining techniques is also the World Wide Web. The Web provides the ability to access one of the largest data repositories, which in most cases still remains to be analyzed and understood. Recently, DM techniques are also being applied to social sciences, home land security and counter terrorism. A DM system is therefore composed of a software environment that provides all the functionalities to compose DM applications, and a hardware back-end onto which the DM applications are executed.

Data mining can be performed on various types of databases and information repositories, but the kind of patterns to be found are specified by various data mining functionalities like class/concept description, association, correlation analysis, classification, prediction, cluster analysis etc. Among these, Cluster analysis is one of the major data analysis method widely used for many practical applications in emerging areas [2]. Clustering is the process of finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups. The quality of a clustering result depends on both the similarity measure used by the method and its implementation and also by its ability to discover some or all of the hidden patterns. There is a number of clustering techniques that have been proposed over the years [3]. Different clustering approaches may yield different results. The partitioning based algorithms are frequently used by many researchers for various applications in different domains. This research work compares two of the partitioning based clustering techniques namely k-Means and k-Medoids via its performance based on their execution time.

The remainder of the paper is structured as follows. The next section provides a comprehensive outline of related work via literature survey. Section 3 describes the basic approach and method of both algorithms. An experimental setup of the telecommunication data and the properties of the same data are discussed in Section 4. Section 5 explores the clustering process and obtained results of the algorithms. Finally, Section 6 contains the concluding remarks of the research work.

## **2. Literature Survey**

Nowadays, data clustering has attracted the attention of many researchers in different disciplines. It is an important and useful technique in data analysis. A

large number of clustering algorithms have been put forward and investigated. The main advantage of clustering is that interesting patterns and structures can be found directly from very large data sets with little or none of the background knowledge. The cluster results are not subjective, but implementation dependent. Data Clustering has been addressed by many researchers and many clustering approaches have been explored and studied. A variety of data clustering algorithms are developed and applied for many applications domain in the field of data mining. Clustering techniques have been applied to a wide variety of research problems. Hartigan provides an excellent summary of the many published studies reporting the results of cluster analyses [4]. For example, in the field of medicine, clustering diseases, cures for diseases, or symptoms of diseases can lead to very useful taxonomies. In the field of psychiatry, the correct diagnosis of clusters of symptoms such as paranoia, schizophrenia, etc. is essential for successful therapy. In archeology, researchers have attempted to establish taxonomies of stone tools, funeral objects, etc. by applying cluster analytic techniques. In general, whenever one needs to classify a “mountain” of information into manageable meaningful piles, cluster analysis is of great utility.

Bradley P.S and Fayyad describe refining Initial Points for k-Means Clustering in their paper [5]. They said that the practical approaches to clustering use an iterative procedure (e.g. k-Means, EM) which converges to one of numerous local minima. This paper presents a procedure for computing a refined starting condition from a given initial one that is based on an efficient technique for estimating the modes of a distribution. Recently, Bhukya *et al.*, deals with the performance evaluation of partition based clustering algorithms in grid environment using design of experiments [6]. In their work, they focus mainly on the analysis of k-Means and k-Medoids algorithms. One of the disadvantages of using these algorithms is its unsuitability for larger data sets. To solve this problem Grid environment has been selected. The main objective of the work is to implement the partition based clustering algorithms in the Grid environment on Grid Gain middleware and analyze their performance for large datasets with Design of Experiment (DOE) framework. Finally, they conclude that the k-Means clustering algorithm is faster than k-Medoids when tested with large data sets and the results are found to be satisfactory.

A review of the most common partition algorithms in cluster analysis: a comparative study is discussed in a research work by Susana *et al.*, in [7]. In this work, a simulation study was performed to compare the results obtained from the implementation of the algorithms k-means, k-medians, PAM and CLARA when continuous multivariate information is available. Additionally, a study of simulation is presented to compare partition algorithms qualitative information, comparing the efficiency of the PAM and k-modes algorithms. The efficiency of the algorithms is compared using the Adjusted Rand Index and the correct classification rate. Finally, the algorithms are applied to real databases with predefined classes.

An Enhanced k-means algorithm to improve the Efficiency Using Normal Distribution Data Points is discussed by Napoleon and Ganga Lakshmi in their research work [8]. This paper 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. They found that the elapsed time taken by proposed enhanced k-means is less than k-means algorithm. A work carried out by Benderskaya *et al.*, titled as “Self-organized Clustering and Classification: A Unified Approach via Distributed Chaotic Computing” [9] describes a unified approach to solve clustering and classification problems by means of oscillatory neural networks with chaotic dynamics. The advantages of distributed clusters formation in comparison to centers of clusters estimation are demonstrated. New approach to clustering on-the-fly is proposed.

A Novel Approach to Medical Image Segmentation is presented by Shanmugam *et al.*, in their paper [10]. In this research, a novel approach is used to segment the 2D echo images of various views. A modified k-Means clustering algorithm, called “Fast SQL k-Means” is proposed using the power of SQL in DBMS environment. In k-Means, Euclidean distance computation is the most time consuming process. Since the entire processing is done with database, additional overhead of import and export of data is not required. The 2D echo images are acquired from the local Cardiology Hospital for conducting the experiments.

There are number of research articles utilizing broad band data for the analysis of various types of networks. Also, some of the clustering algorithms are utilized to analyze telecommunication data. One such topic was done by Sung Suk Kim and Sun Ok Yang titled as “Wireless sensor gathering data during long time involving both telecommunication data and clustering algorithms”. In this paper [11], they have proposed bit-vector based information storage method (BV) for the analysis of telecommunication data. The method reduces the amount of overall data by storing only the deviation from the previous measurement as bit-value. They implemented a simulator to evaluate the efficiency of the proposed methods, and the methods turned out to be fairly efficient compared to normal cases.

### 3. Materials and Methods

Clustering is a concept to determine the pattern through map and analysis of available data set according to the need and demand of the business applications. Clustering is belonging to both data analysis and machine learning major domains. Many methodologies have been proposed in order to organize, to summarize or to simplify a dataset into a set of clusters such that data belonging to a same cluster are similar and data from different clusters are dissimilar [12]. Different approaches may yield different results in the clustering techniques. Here, the telecommunication data is analyzed based on the distance between the server

locations and their user points. As stated in previous sections, the k-Means and k-Medoids clustering algorithms are compared and analyzed [13]. In the comparison study, only the time parameter is taken for analysis.

### 3.1. The k-Means Algorithm

The k-Means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume  $k$  clusters) fixed a priori [14]. This algorithm aims at minimizing an objective function, in this case a squared error function.

The objective function

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (1)$$

where  $\|x_i^j - c_j\|^2$  is a chosen distance measure between a data point  $x_i^j$  and the cluster centre  $c_j$ , is an indicator of the distance of the  $n$  data points from their respective cluster centers. The algorithm is composed of the following steps:

Step 1: Place  $k$  points into the space represented by the objects that are being clustered. These points represent initial group centroids.

Step 2: Assign each object to the group that has the closest centroid.

Step 3: When all objects have been assigned, recalculate the positions of the  $k$  centroids.

Step 4: Repeat Steps 2 and 3 until the centroids no longer move.

This produces a separation of the objects into groups from which the metric to be minimized can be calculated. Although it can be proved that the procedure will always terminate, the k-Means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum [15]. This k-Means is a simple algorithm that has been adapted to many problem domains [16].

### 3.2. The k-Medoids Algorithm

The k-Means algorithm is sensitive to outliers since an object with an extremely large value may substantially distort the distribution of data [17]. Instead of taking the mean value of the objects in a cluster as a reference point, a medoid can be used, which is the most centrally located object in a cluster. Thus, the partitioning method can still be performed based on the principle of minimizing the sum of the dissimilarities between each object and its corresponding reference point [18]. This forms the basis of the k-Medoids method. The basic strategy of k-Medoids clustering algorithms is to find  $k$  clusters in  $n$  objects by first arbitrarily finding a representative object (the medoids) for each cluster. Each remaining object is clustered with the medoid to which it is the most similar [19]. The algorithm takes the input parameter  $k$ , the number of clusters to be partitioned among a set of  $n$  objects. A typical k-Medoids algorithm for partitioning based on medoid or central objects is as follows:

Input: k: The number of clusters

D: A data set containing n objects

Output: A set of k clusters that minimizes the sum of the dissimilarities of all the objects to their nearest medoid.

Method: Arbitrarily choose k objects in D as the initial representative objects;

Repeat assign each remaining object to the cluster with the nearest medoid;

Randomly select a non medoid object  $O_{\text{random}}$ ;

Compute the total points S of swapping object  $O_j$  with  $O_{\text{random}}$ ;

if  $S < 0$  then swap  $O_j$  with  $O_{\text{random}}$  to form the new set of k medoid;

Until no change;

It attempts to determine k partitions for n objects. After an initial random selection of k medoids, the algorithm repeatedly tries to make a better choice of medoids. Therefore, the algorithm is often called as representative object based algorithm.

#### 4. Experimental Setup

Data mining concepts are used in different applications as per the need, demand, nature of the problem and domain. In this research, the clustering process is achieved using a distance method. The clustering process is aimed to minimize the expenditure of the business application and increase the benefits of the business. The algorithms are implemented in the real time connection oriented telecommunication data and the results are discussed. In this process, the communication connection structure is evaluated and reconstructed using clustering techniques for the effective data distribution. The data distribution process is affected by the connected server, distance and number of connections available in the specific server. The distance factor create an impact on the creation of the infrastructure using cable, cost of the cable, manpower, maintenance and the data distribution based on the bandwidth. Therefore, the data access points are considered as data points and planned to optimize the network using clustering concepts. After the clustering process, the number of connections for each server is changed.

The data set collected from a broadband service provider at Chennai city. The connection oriented data set contains 285,520 data connection points with 27 servers with locations. The 27 servers are treated as 27 clusters in this work and they are called as data centers. The user points are called as data access points. There are 12 data sets available. One data set for each month. The data set contains information about distance, type of connection (single user, multi user), data transfer capacity (256, 512, 1024, 2048), area code and the server number in which the data points are connected. This representation is based on the connections established from the month of January to December. The collected data consists of the connection establishment month, area and the connected data center, type of the data service and the volume of data used in the year. The total connection of data access points are connected according to the geographical location. This connection is made, based on the demand of the customer which

is provided by the service provider. The total number of connected data points for each and every month is given in **Table 1**. The distance between data centers and its data access points are available in meters. The servers (Data centers) are treated as center point in the clustering process. Normally, the distance between one server and the same server is zero.

The total number of data access points (number of user points before clustering) in all the 27 servers are given in **Table 2**. The data access points are currently distributed unevenly based on the request of the user at every month and the availability of the nearest data centers. But, this caused the issues of the traffic and the distribution of data. It affects the quality of the server to the user as well as the service provider. In the exiting connection, the data center 27 has 279 connections alone. This is the minimum number of connection in a particular data center. But, the first data center has 14,101 data access points, which is the maximum of all the data centers. After the clustering process, this may vary from server to server.

**Table 1.** Total connections in data set.

Month	Connection Numbers		Total Connections
	From	From	
Jan	1	25,094	25,094
Feb	25,095	45,416	20,322
Mar	45,417	66,977	21,561
Apr	66,978	94,928	27,951
May	94,929	122,271	27,343
Jun	122,272	143,345	21,074
Jul	143,346	168,791	25,446
Aug	168,792	189,130	20,339
Sep	189,131	209,432	20,302
Oct	209,433	231,583	22,151
Nov	231,584	261,107	29,524
Dec	261,108	285,520	24,413
	Grand Total		285,520

**Table 2.** Number of connections in servers (before clustering).

Server	Total	Server	Total	Server	Total
1	14,101	10	13,019	19	12,845
2	9607	11	12,922	20	11,534
3	9493	12	12,923	21	9778
4	11,024	13	13,203	22	8144
5	12,662	14	13,040	23	6697
6	12,964	15	12,904	24	5130
7	13,085	16	13,000	25	3363
8	13,098	17	13,120	26	1819
9	12,933	18	12,833	27	279

## 5. Clustering Process and Experimental Results

The k-Mean and k-Medoids algorithms are implemented using MATLAB software and the results are discussed in this section. In the implementation process, the data set is processed based on the distance. Initially, first data center and the first month (January) are selected. For the selected data center and for the month, the distance is reconstructed and stored in the process matrix. The reconstruction of distance is made by using the Pythagoras theorem. After the reconstruction of the distance, the data access points are clustered using any one of the taken algorithm. In this process, the number of user points in each and every data center is reassigned. Therefore, each data center has some new number of data access points after the process. The computational time and the number of connections in each server are stored in tables. This means that the starting and ending time of clustering process is stored in tables. Next, by choosing the same first data center, the second month (February) data access points are chosen and clustered using the chosen algorithm. This process is repeated upto the last month (December) data. After processing the 12th month data by choosing the first server, the second server is chosen and the process is repeated up to 27 servers. Hence, the number of connections in each data center is considered as application impact and the process time is considered as computation impact. The algorithmic steps involved in the clustering process are summarized below.

- 1) Selection of Algorithm from k-Means or k-Medoids
- 2) Selection of data center
- 3) Calculate the distance between data access points and servers based on selected data center
- 4) Selection of monthly data
- 5) Implementation of the selected algorithm and cluster the distance
- 6) According to the processed cluster, the data points are reassigned to the data center
- 7) Observe the cluster process start time and completion time
- 8) Summaries the number of connection in each data center
- 9) Implement the step (2) to (8) to all the different data sets
- 10) Represent the data connection according to the newly assigned data center

Usually, clustering approaches yields different kind of results, this depends on the nature and chosen application of the problems. Next section discusses the results and interpretations about the results of the algorithms produced as output in the process.

### 5.1. Experimental Results

The k-Mean clustering algorithm is implemented as per the discussion above. **Table 3** is the processing time of both k-Means and k-Medoids algorithm. The total elapsed time by k-Means algorithm to cluster first month data set for the entire 27 data center is given in third column. Processing time for the first month data is 16.653 seconds and so on. Average processing time of all the 12



**Table 3.** Results of clustering algorithms (processing time).

Month	k-Means Algorithm			k-Medoids Algorithm		
	Start Time	End Time	Elapsed Time	Start Time	End Time	Elapsed Time
1	29.018	45.671	16.653	41.047	15.6	34.505
2	29.363	38.942	9.579	0.460	33.7	33.205
3	27.144	4.152	37.008	20.023	56.5	36.491
4	43.083	1.773	18.690	38.515	18.3	39.768
5	41.498	2.019	20.521	57.559	35.8	38.246
6	42.292	58.14	15.848	14.983	47.2	32.247
7	37.016	53.722	16.706	25.383	56.9	31.531
8	33.336	42.165	8.829	35.442	8.4	32.958
9	21.446	29.496	8.050	46.303	19.2	32.890
10	8.745	24.782	16.037	57.410	34	36.543
11	3.969	35.325	31.356	12.646	56.5	43.837
12	16.26	32.078	15.818	36.650	13.2	36.568
Average Process Time			17.925	Average Process Time		35.732

month data is available in the last row, which is found to be 17.925 sec. The data points in each data center points are clustered (distributed) using the k-Means algorithm based on the neighborhood distance. The minimum time taken by the algorithm is 8.050 seconds and the maximum is 37.008 seconds. To avoid lengthy discussion, the number of data points created by the algorithm is not shown. The first server is chosen as a center point, in the same way the second server is chosen and the clustering process is repeated.

In the similar fashion, the results of k-Medoids algorithms are also given in the **Table 3**. In this table, it is easy to find the total processing time for the first month data is 34.505 seconds and so on. The average processing time for the 12 month data is 35.732 seconds. Both algorithms are executed 12 times (up to choosing 12 servers as center point) and the average results are listed in **Table 4**. The average processing time of 12 runs of the k-Means clustering is observed. The algorithm takes a minimum of 15.919 Seconds in the 7th run and a maximum of 19.796 Seconds in the 5th run. From the same table, a minimum of 31.924 seconds and a maximum of 57.659 seconds are taken by k-Medoids algorithm to complete the clustering process. After the clustering process, the number of data points created by both the algorithms is given in **Table 5**. It is very clear that from the results of the algorithm, the distribution of data points in each and every data center is almost closer to one another. In the distribution process, data access points are equally divided for all the data centers. This makes the requirement of the clustering process.

**Table 4.** Processing time.

Run	k-Means	k-Medoids	Run	k-Means	k-Medoids
1	17.925	35.732	7	15.919	39.614
2	17.142	31.924	8	17.224	37.899
3	16.822	33.417	9	18.700	50.904
4	19.664	37.467	10	18.984	57.659
5	19.796	35.511	11	19.474	43.953
6	16.129	38.747	12	16.243	42.728
Average				17.84	40.46

**Table 5.** Total number of connections (after clustering).

Data Center	k-Means	k-Medoids	Data Center	k-Means	k-Medoids
1	10,543	10,333	15	10,276	10,502
2	11,185	11,361	16	10,840	10,746
3	10,816	10,639	17	10,567	10,605
4	10,702	10,544	18	10,559	10,538
5	10,249	11,243	19	10,663	10,303
6	10,773	11,038	20	10,744	9808
7	10,935	10,361	21	10,089	10,438
8	10,372	10,457	22	10,391	10,355
9	10,586	10,877	23	9853	10,363
10	10,311	10,888	24	10,763	10,569
11	10,506	10,375	25	10,487	10,014
12	10,754	10,632	26	10,472	10,593
13	10,519	10,613	27	10,731	10,902
14	10,833	10,424	Total	285,520	285,520

## 5.2. Summary and Discussions

The total numbers of 285,520 data access points are clustered which are available in 27 data centers by the chosen two algorithms. Based on the distance between the data access points and data centers, the performance and efficiency of the clustering process is analyzed. The k-Means algorithm assigns a minimum of 9853 data points and a maximum of 11,185 data points after the clustering. The minimum and maximum data points assigned by the k-Medoids method is 9808 and 11361 respectively. **Figure 1** shows that the minimum and maximum data access points assigned by the proposed two algorithms. Note that the minimum and the maximum data points are created by the k-Medoids method. But, in particular, the difference between minimum and the maximum is very less in both the methods. In **Table 4**, average execution time for 12 runs is given in last row to both algorithms. **Figure 2** is the graphical representation of average time

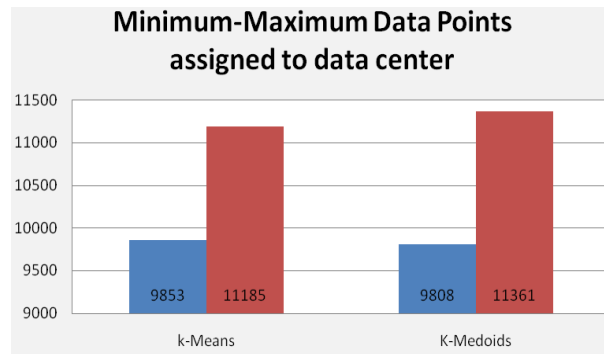


Figure 1. Min & max data access points.

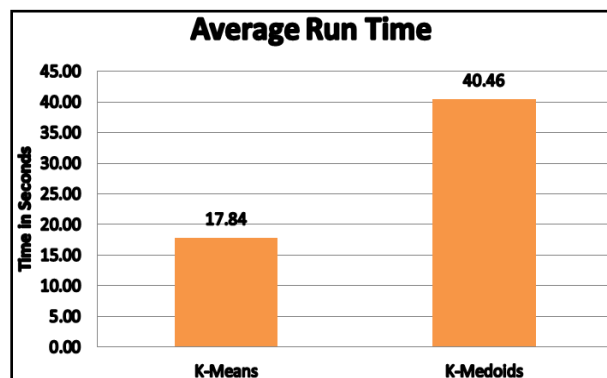


Figure 2. Result comparison.

of the algorithms. From this figure, it is easy to identify that the differences between performance of given two algorithms. Based on the result of several executions of these two algorithms in the MATLAB software, the clustered results are analyzed. According to the efficiency of the algorithms, the performance of k-Means method is better than the k-Medoids methods. It is evident that from the **Figure 2**, the k-Medoids algorithm takes more time than the k-Means algorithm. Thus, for the telecommunication data, the k-Means algorithm is performance wise best because of its structure and simplicity.

## 6. Conclusion

Cluster analysis is still an active field of development. Many cluster analysis techniques do not have a strong formal basis. Cluster analysis is a rather ad-hoc field. There are a wide variety of clustering techniques. Comparisons among different clustering techniques are difficult. All techniques seem to impose a certain structure on the data and yet few authors describe the type of limitations being imposed. In spite of all these problems, clustering analysis is a useful (and interesting) field. In summary, clustering is an interesting, useful, and challenging problem. It has great potential in applications like object recognition, image segmentation, and information filtering and retrieval. However, it is possible to exploit this potential only after making several carefully chosen designs and application. From the experimental approach, by several executions of the program

for proposed algorithms in this research work, following results were obtained. Usually, the time complexity varies from one processor to another processor, which depends on the speed and the type of the system. The advantage of the k-Means algorithm is its favorable execution time. Its drawback is that the user has to know in advance how many clusters are searched for. From the experimental analysis, the distribution of number of connections for each and every server, produced by both the algorithms after clustering process is almost even. The computational time of k-Means algorithm is less than the k-Medoids algorithm. Further, k-Means algorithm stamps its superiority in terms of its lesser execution time. Finally, this work concludes that the k-Means algorithm is better than the k-Medoids algorithm for the chosen connection oriented telecommunication data.

## References

- [1] Han, J. and Kamber, M. (2006) Data Mining: Concepts and Techniques. 2nd Edition, Morgan Kaufmann Publishers, New Delhi.
- [2] Jain, A.K., Murty, M.N. and Flynn, P.J. (1999) Data Clustering: A Review. *ACM Computing Surveys*, **31**. <https://doi.org/10.1145/331499.331504>
- [3] Berkhin, P. (2002) Survey of Clustering Data Mining Techniques. Technical Report, Accrue Software, Inc.
- [4] Hartigan, J.A. (1975) Clustering Algorithms. Wiley Publishers.
- [5] Bradley. P.S., Fayyad, U.M. and Reina, C.A. (1998) Scaling Clustering Algorithms to Large Databases. *Proceedings of the 4th International Conference on Knowledge Discovery & Data Mining*, AAAI Press, Menlo Park, CA, 9-15 .
- [6] Bhukya, D.P., Ramachandram, S. and Reeta Sony, A.L. (2010) Performance Evaluation of Partition Based Clustering Algorithms in Grid Environment Using Design of Experiments. *International Journal of Reviews in Computing*, **4**, 46-53.
- [7] Leiva-Valdebenito, S.A. and Torres-Aviles, F.J. (2010) A Review of the Most Common Partition Algorithms in Cluster Analysis: A Comparative Study. *Colombian Journal of Statistics*, **33**, 321-339.
- [8] Napoleon, D. and Ganga Lakshmi, P. (2010) An Enhanced k-Means Algorithm to Improve the Efficiency Using Normal Distribution Data Points. *Int. Journal on Computer Science and Engineering*, **2**, 2409-2413.
- [9] Benderskaya. E.N. and Zhukova, S.V. (2011) Self-Organized Clustering and Classification: A Unified Approach via Distributed Chaotic Computing. *International Symposium on Distributed Computing and Artificial Intelligence, Advances in Intelligent and Soft Computing*, **91**, 423-431.
- [10] Shanmugam, N., Suryanarayana, A.B., Chandrashekar, S.D. and Manjunath, C.N. (2011) A Novel Approach to Medical Image Segmentation. *Journal of Computer Science*, **7**, 657-663. <https://doi.org/10.3844/jcssp.2011.657.663>
- [11] Kim, S.S. and Yang, S.O. (2011) Wireless Sensor Gathering Data during Long Time. Telecommunication Systems.
- [12] Jain, A.K. and Dubes, R.C. (1988) Algorithms for Clustering Data. Prentice Hall Inc., Englewood Cliffs, New Jersey.
- [13] Velmurugan, T. and Santhanam, T. (2010) Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distribu-

tions of Data Points. *Journal of Computer Science*, **6**, 363-368.

<https://doi.org/10.3844/jcssp.2010.363.368>

- [14] Rakhlin, A. and Caponnetto, A. (2007) Stability of k-Means Clustering. *Advances in Neural Information Processing Systems*, **12**, 216-222.
- [15] Dharmarajan, A. and Velmurugan, T. (2016) Efficiency of k-Means and k-Medoids Clustering Algorithms Using Lung Cancer Dataset. *Int. Journal of Data Mining Techniques and Applications*, **5**, 150-156.  
<https://doi.org/10.20894/IJDMTA.102.005.002.011>
- [16] Velmurugan, T. (2012) Efficiency of k-Means and k-Medoids Algorithms for Clustering Arbitrary Data Points. *Int. J. Computer Technology & Applications*, **3**, 1758-1764.
- [17] Park, H.S., Lee, J.S. and Jun, C.H. (2009) A k-Means-Like Algorithm for k-Medoids Clustering and Its Performance. Department of Industrial and Management Engineering, POSTECH, South Korea.
- [18] Yu, Y.Q., Xin, W., Liu, G.N., Li, H., Li, P. and Lin, H. (2017) A Combinatorial Clustering Method for Sequential Fraud Detection. *IEEE International Conference on Service Systems and Service Management*, 1-6.
- [19] Vijayakumar, M. and Parvathi, R.M.S. (2010) Concept Mining of High Volume Data Streams in Network Traffic Using Hierarchical Clustering. *European Journal of Scientific Research*, **39**, 234-242.