ISSN:0975-9646

S. Adaekalavan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5970-5977

Huddle Based Harmonic K Means Clustering Using Iterative Relocation on Water Treatment Plant Data Attributes

S. Adaekalavan[1] and Dr. C. Chandrasekar[2]

[1]Asst. Professor, Department of Information Technology, J.J. College of Arts and Science(Autonomous), Pudukkottai, Tamilnadu, India
[2]Associate Professor, Department of Computer Science, Periyar University,Salem, Tamilnadu, India

Abstract-Clustering is one of the important methods in data mining to weight the distance between the cluster objects effectively. Existing K-means algorithm with Affinity Propagation (AP) algorithm captures the structural information of texts, and improve the semi supervised clustering process. Seeds Affinity Propagation (SAP) provides the detailed distance measurement but it takes the longer execution time to perform the clustering process. Weighted Clustering Ensemble (WCE) algorithm on the existing work provides an effectual technique on the temporal data clustering. WCE are not effective in developing the speed clustering on the data mining system. To reduce the execution time on clustering of the large capacity of dataset, Huddle based Harmonic K-means Clustering (HH K-means Clustering) mechanism is proposed in this paper. HH K-means Clustering first performs the initial screening process to divide them into a cluster and non-cluster data form. Secondly, HH K-means Clustering develops the Iterative Relocation (IR) technique to improve the speed of the clustering process. Iterative Relocation (IR) technique calculate and estimates the cluster center objects (i.e.,) Centroids in HH K-means Clustering. The IR technique performs the initialization, relocation of objects and cluster updating process to reduce the computation of distance measure on the every data objects. Huddle groups the data objects using the Harmonic K-means Clustering and calculate the clustering time and clustering rate. So, HH K-means clustering add some improved conditions and factors to enhance the effectiveness of data mining clustering system with approximately 7 % improved cluster rate. Experiment is conducted using Water Treatment Plant Data Set on the factors such as execution time, huddle based clustering rate and entropy level.

Keywords: Iterative Relocation, Distance Measure, Entropy Level, Harmonic K-means clustering, Data Objects, Clustering Time

1. INTRODUCTION

With the advancement in Information Technology and Computer Science, the computer is being applied today in all walks of life. The efficiency and effectiveness of computer software has increased over time. People's produces data of different domain and all the information are stored in the database. Business enterprises, public departments and research organization make of data mining use in many application areas such as artificial intelligence, volcanic activity, biology, customer relationship management, data density , data mining, information retrieval, image processing , machine learning, advertising, medicine, pattern recognition, psychology, statistics and so on.

In the past several years, massive amounts of data are accumulated and stored in different forms, because these data are very difficult to store. So, it leads to the difficult task on getting the valuable information and to achieve the decision-making process. As a result, data mining techniques have emerged, and gradually become a hot spot on attracting a lot of researchers. Data mining is the extraction of hidden analytical information from large databases and works on all dominant new technology with enormous potential to help the companies.

The company makes use of most important information in their data warehouses. Data mining tools forecast future trends and performance, allowing businesses to make practical, knowledge-driven decisions. Data mining discovers explanation through clustering visualization, association and sequential analysis. Clustering analysis is a very important data mining technology. Data clustering is a common technique for data scrutiny, which is used in many fields, includes machine learning, data mining, pattern recognition, image analysis, video analysis and bioinformatics.

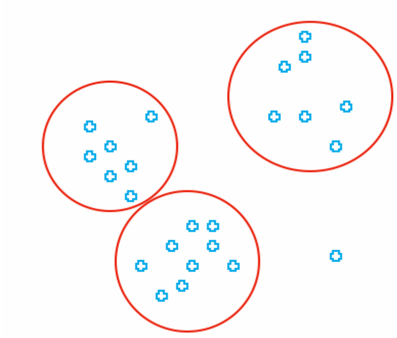

Fig 1 Cluster Analysis in Data Mining

Cluster analysis in data mining is clearly represented through the Fig 1. The similar data objects are grouped together where the objects in space are visually represented. True Query-cluster distance as demonstrated in [8] provides the minimum bounding hyper rectangles (MBR) and hyper spheres (MBS) values. Vector approximation employs scalar quantization, while the hyper plane bounds are movable. Imaginably, the cluster-distance bounds are further tightening thereby optimizing the clustering algorithm distance bounds. On the other hand, most of them come upon difficulties when clusters cover in subspaces with tremendously lowering the dimensionality. In particular, on high-dimensional data the conventional cluster structure identifies the similarity measures in [9] but conventional clustering algorithms take the longer running time.

Harmonic K-means an algorithm (Harmonic K-means) is a system of clustering which permit one portion of data to belong two or more clusters, thereby reduces the execution time. The experiment deals with two of the more clustering algorithms such as Harmonic K-means and K-means algorithm to measure the distance measure. They are described and analyzed based on the distance between the two data objects. In both the algorithms a set of 'n' data objects are given in the four dimensional dataset and 'K' clusters resolve a set of 'n' objects in the given space, called centers. The center value decrease the mean squared distance from each data objects to its nearest center.

Weighted Clustering Ensemble (WCE) algorithm as illustrated in [2] provides an effective enabling technique to make use of different representations. The representation cuts the information loss in a single form and makes use of the mixture of information sources under the temporal data objects. WCE are not effective in developing the clustering on the limited amount of time in data mining system. Feature group K-means clustering algorithm in [10] optimize the result by calculating the two types of subspace weights by limiting the processing time. Feature group K-means clustering is not effective in dividing the features automatically and to perform the grouping operation K-means clustering on the real world applications.

Novel Spatial Clustering Algorithm Based on Delaunay Triangulation (NSCABDT) as demonstrated in [13] determines the neighborhoods based on dividing the features notion. The neighboring region formed in NSCABDT algorithm reflects the neighbor's distribution but the entropy level is of higher ratio on dividing the features. UK-means calculate Expected Distances (ED) among objects and cluster representatives in [12] to systematize the uncertain objects and to diminish pruning overheads. Max-BB (Bounding Box) pruning algorithm potentially reduces the entropy level but cluster-shift operation is not carried out effectively.

Multi viewpoint-based similarity measure in [4] does not describe alternative forms for the relative resemblance. Similarity measure also does not use standard relative similarities according to the different viewpoints. It fails in applying criterion functions for hierarchical clustering algorithms. MAximal Resemblance Data Labeling mechanism for clustering is described in [11] where the unlabeled data objects are labeled into the corresponding cluster based on the novel definite clustering. The changes between different clustering results do not provide the high quality rate on execution.

Clustering of high-dimensional data is not explored to connect other data directly to research directions, include kernel mappings and shared-neighbor clustering. The major drawback in [5] is that the hyperspherical cluster alone detected as K-Means. Additionally, it does not travel around methods for using hubs to repeatedly determine the number of clusters in the data. Clustering algorithm extends affinity propagation in [1] with a narrative asymmetric similarity quantity. The similar objects capture the structural information of texts but it consumes the longer execution time. A semi supervised learning approach develops the knowledge from a minute quantity of labeled objects but the generic seeds cluster construction strategy is not developed.

In this work, focus is made on developing a clustering model for the data objects using the proposed Huddle based Harmonic K-means clustering. Clustering work in data mining is a significant task to improve the performance function. Screening process is carried out in the HH K-means Clustering in which it divides the objects into cluster and non-cluster form. The objects are taken to compute the centroid value using the huddle based harmonic. The mean and median value are also used to exactly plot the distance on the single measure in HH K-means Clustering, thereby reducing the iteration count. The centroid value used and the updating of new objects in the huddle based harmonic cluster are carried out through the decomposed series.

The structure of this paper is as follows. In Section 1, describes the basic problems on clustering using K-means algorithm in the data mining. In Section 2, present an overall view of the Huddle based Harmonic K-means Clustering (HH K-means Clustering) mechanism. Section 3 and 4 outline experiment results with parametric factors and present the result graph for research on clustering. Finally, Section 5 demonstrates the related work and Section 6 concludes the work with better result outcome.

2. HUDDLE BASED HARMONIC K-MEANS CLUSTERING MECHANISM

The HH K-means Clustering main goal is to cluster the K-different data objects within the minimal execution time. The 'K' data objects clustered using the huddle based harmonic is condensed and autonomous. The HH K-means clustering consists of two separate steps. In the first step, HH K-means Clustering performs the initial screening process to divide them into a cluster and non-cluster data form. The screening process attempts to divide the data objects into a finite collection of elements. The screening process is clearly represented in Fig 2.

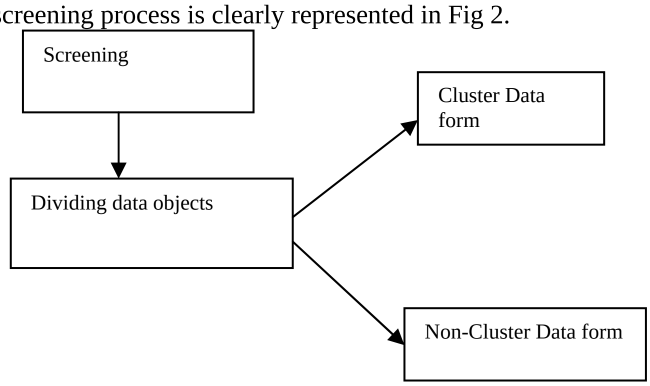

Fig 2 Screening Process HH K-means clustering

The divided data objects are of two forms such as cluster type data objects and non-cluster type data objects. The cluster type data objects are very easy to perform the clustering operation using the huddle based harmonic system. In the second step, Iterative Relocation (IR) technique is used to improve the clustering rate with the minimal execution time. IR technique in HH K-means clustering determines the Centroids of data object and the cluster centers. The mean and median computation is also carried out for centroid computation and IR technique uses the three processing step.

The processing step is starts with the initialization, and second process is to relocate the data points into clusters and final point is to update the cluster on insertion of newer data objects. HH K-means clustering performs the initialization using the data objects after the screening process. In the second step, the relocation of objects carried out with the centroid value. The relocation follows the adjoining centroid arrangement process. The data objects arranged to the nearby distance of centroid point in order to reduce the execution time. Finally, the cluster updating carried out on newer data objects insertion in IR technique uses the decomposed series method. Finally, IR technique implementation in HH K-means clustering improves the speed and clustering rate. Architecture Diagram of HH K-means Clustering is described in Fig 3.

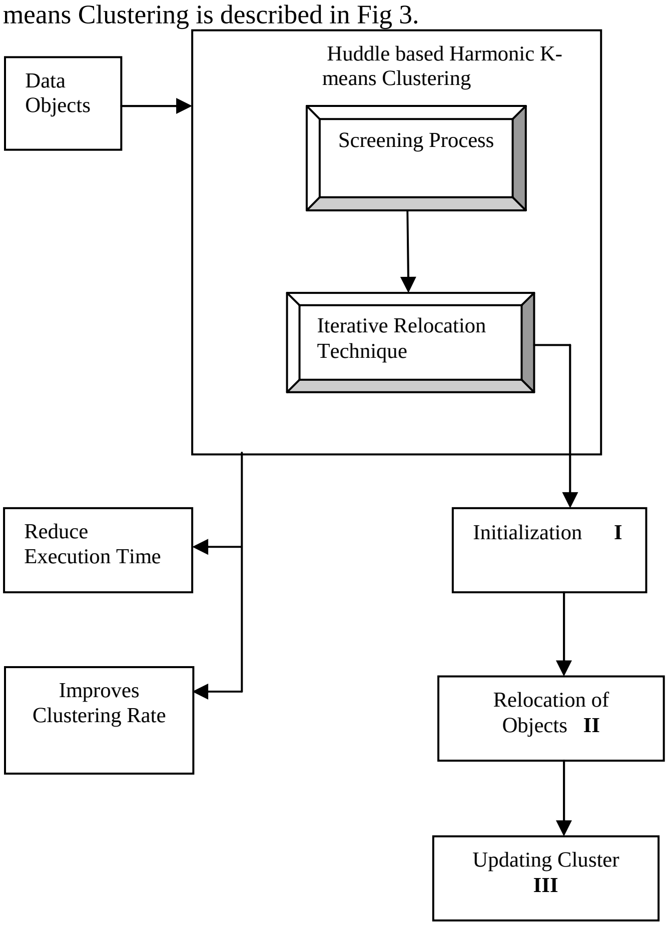

Fig 3 Architecture Diagram of HH K-means

As illustrated in Fig 3, HH K-means clustering takes the Water Treatment Plant Data Set to perform the clustering operation with the minimal execution time. HH K-means clustering initially uses the screening process to separate (i.e.,) divide the data objects effectively. The divided water treatment data object uses the iterative relation technique. The huddle based harmonic which uses the Harmonic Averages of the distances from each data objects center the components to improve the performance function.

HH K-means clustering significantly improves the quality of clustering results by using the IR technique. The initialization of data objects for clustering is performed in the first step, and then the relocation of object is carried out. The relocation carried out using the adjoining centroid arrangement process. The centroid value is used to adjoin (i.e.,) arrange in a nearby distance to reduce the execution time in HH K-means clustering. Finally, updating of cluster work is performed with the decomposed series method. The decomposed work improves the clustering work on the updating of data object series, thereby improves the clustering rate.

2.1 Screening Process

In Huddle based Harmonic work, the screening process is carried out on the raw data objects of the water treatment plant. The processing work in the screening phase of HH K-means clustering is illustrated through Fig 4.

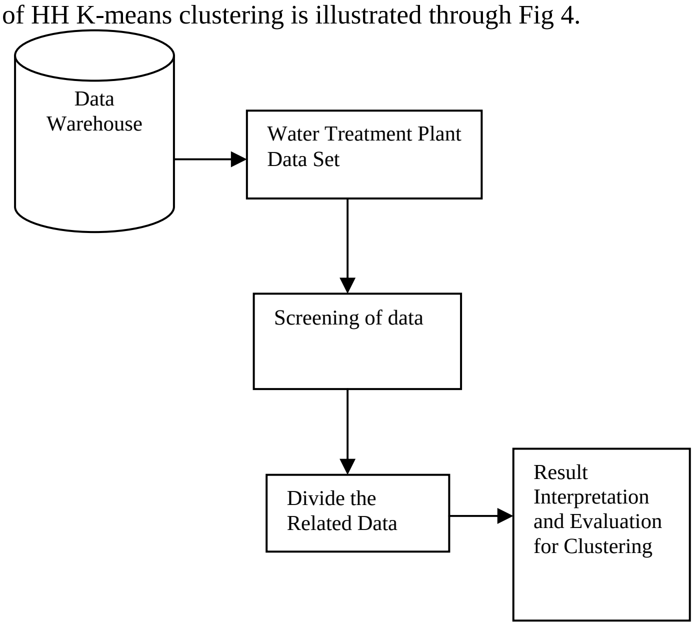

Fig 4 Screening Phase Procedure

The data warehouse contains the large collection of information to perform the K-means clustering. Water treatment data is the process of removing the pollutants (i.e.,) non-cluster able data such as wood chips, leaves, aquatic plants papers, plastics, tins, containers and floating impurities. The removal work is carried out using the screening process and described in Fig 4. The screened water treatment data objects are now divided into related data objects (i.e.,) plant treatment minerals for clustering in HH K-means clustering. The treated water plants are evaluated and result is used to perform the huddle based harmonic clustering to improve the clustering rate.

2.2 Iterative Relocation Process

The Iterative Relocation (IR) technique in HH K-means clustering is developed to provide the quality clustering result on the water treatment data objects. The IR optimizes the cluster by identifying the centroid value with mean and median factor.

2.2.1 Initialization

The iterative process takes the related data objects (i.e.,) minerals for the water treatment processing. The initialization work gives all the information in HH K-means clustering to perform the clustering process. The centroid value of the data objects 'x' and 'y' objects are computed as,

$$C_x = \frac{\sum_{i=0}^{n}(x_i + x_{i+1})(x_i y_{i+1} - }{4} \quad \ldots \text{Eqn (1)}$$

$$C_y = \frac{\sum_{i=0}^{n}(y_i + y_{i+1})(x_i y_{i+1} - }{4} \quad .. \text{Eqn (2)}$$

The centroid value ' ' and ' ' of the data objects 'x' and 'y' are computed. The '4' represents the coordinate points 'i' on the huddle based harmonic clustering. The centroid is computed using the mean value, where mean is the average number of related data objects within the distance. The mean is computed as,

$$\textit{Mean Data objects } (D \quad \ldots \text{Eqn (3)}$$

'D' denotes the overall data objects count used to compute the mean value. The median in the huddle based harmonic clustering is taken as the centroid data objects to perform the clustering, so the computation of distance on the each iteration is reduced.

2.2.2 Relocation of objects

The relocation of data objects in the HH K-means clustering uses the adjoining centroid arrangement process. The adjoining centroid arrangement process relocates the data objects using the median factor computation. The data objects arranged to the nearby distance of centroid point in HH K-means clustering, which reduces the execution time.

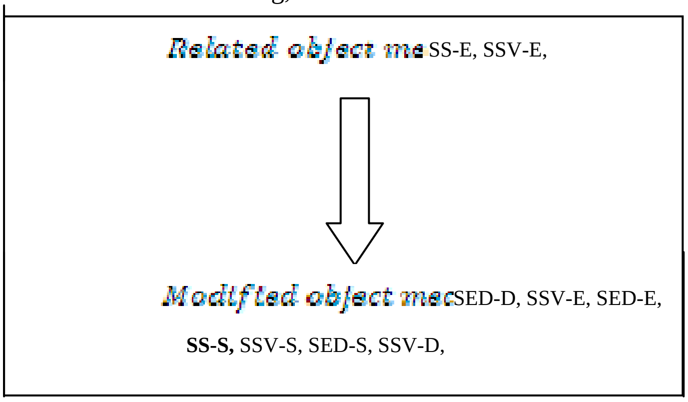

Fig 5 Relocation of objects through median Factor

The modified median value improves the relocation of the factor in HH K-means clustering. [The information for median computation is taken from the Water treatment plant dataset from UCI repository]. The related object median on the SS-E (input suspended solids to plant), SSV-E (input volatile supended solids to plant), SED-E (input sediments to plant), SSV-S (output volatile supended solids), and SED-S (output sediments) is taken as the SED-E to perform the clustering. The relocated data objects as shown in Fig 5, takes SS-S as the median point to perform the Huddle based Harmonic K-means Clustering.

2.2.3 Cluster Updating Process

The new data object in the HH K-means clustering is updated using the decomposed series method. HH K-means clustering starts with the set of initialization procedure, as described in Section 2.2.1. HH K-means clustering computes the centroid value of the data objects such as

= | , , m| …….. Eqn (4)

Where 'm' is the newly updated data object in clustering. The decomposed series in HH K-means clustering combines the newly updated water treatment plant attributes to the related cluster.

2.3 Huddle based Harmonic Clustering

HH K-means cluster provides the clustering number after the screening process. The screened data objects are then used as an iterative relocation technique to reduce the execution time in HH K-means clustering. HH K-means cluster apply a new distance measurement to compute the centroid value and test the effectiveness of huddle and K-means harmonic combination. The combination of clustering effect thoroughly performs the numerical experiments. The new measurement in HH K-means cluster is described as,

$$HH(x_i y_i) = \sum_{i=1}^{n} \left(1 - \exp\left(-\beta \| x_i - y_i \|^2\right)\right)$$

Eqn (5)

In which is a positive constant, from the distance function of Huddle based Harmonic (HH) Clustering. The distance function HH(x, y) is a monotonically increasing function || x - y ||, namely HH (x, y) increases with the increase of || x - y ||. The steps of huddle based harmonic K-means clustering algorithm is described as follows

Begin

Input: Number of desired clusters, k, and a data objects D= {d1, d2 ….dn} containing 'n' data objects

Output: Set of K-Clusters with lesser execution time

Step 1: Selects 'k' data objects from water treatment plant dataset as initial cluster centers.

Step 2: For data objects 'D'

Step 3: Calculate the distance between each data object (1 <= i<=n) and all k cluster centers and as

harmonic based distance measure function

Step 4: , , m| assign data object to the nearest

cluster.

Step 5: For each data object , find the closest center and

assign to cluster center to perform clustering

Step 6: End For

Step 7: Repeat Step 2 to 6 for clustering operation

End

The Huddle based Harmonic K-means algorithms as described above through algorithmic step does not require to calculate the distance from the each data object to all the centers of k-clusters when executed on each time. HH K-means clustering reduces up a lot of execution time especially for large capacity of datasets such as Water Treatment Plant Data Set. Huddle based Harmonic K-means algorithms is robust by providing the precise experimental results.

3. Experimental Evaluation Of HH K-Means Clustering

Huddle based Harmonic K-means Clustering (HH K-means Clustering) mechanism performs the experiment in the JAVA platform using Water Treatment Plant Data Set from UCI repository. Water Treatment Plant Data Set appears from the every day events of sensors in an urban waste plant. The objective of using Water Treatment Plant Data Set in HH K-means Clustering is to categorize the prepared state of the plant in order to forecast faults through the state variables of the plant at each of the stages treatment process.

Water Treatment Plant Data Set consists of 527 instances 'i' with 38 attributes. Water Treatment Plant Data Set performs the clustering task and all the attributes are of numeric and continuous. Huddle based Harmonic K-means Clustering (HH K-means Clustering) mechanism are compared against the K-means algorithm with Affinity Propagation (AP) algorithm and Weighted Clustering Ensemble (WCE) algorithm. The experiment is conducted on the factors such as execution time, huddle based clustering rate, true positive rate, entropy level and net similarity level.

Execution time is defined as the amount of time consumed to perform the clustering operation using the Huddle based Harmonic. Execution time is measured in terms of seconds (sec).

$$\text{Execution time} = \text{Start Time to cluster } 'D' - \text{End Time on Clustering } 'D'$$

'D' represents the data objects in the data warehouse. The rate at which the high quality clustering performed on the 'D' data objects is called the clustering rate. The cluster rate is always measured in terms of percentage (%). Entropy level is the measure of disorder while updating the clustering process. The disorder is reduced in HH K-means clustering when compared to existing system. True positive rate is the amount of relevant information clustered based on the 'D' related data objects. The Data objects 'D' is used effectively to reduce the running time by improving the true positive rate percentage.

$$\text{True Positive Rate} = \left[\frac{\text{Clustered Objects count} \cap \text{Data objects}}{\text{Data Object Count}} * 100\right]$$

The Data object Count 'D' varied on each experimental evaluation to identify the effectiveness of the system through HH K-means clustering. Net similarity level is defined as the amount of cluster formed with the same data an object (i.e.,) attributes for different water treatment processing, and it is measured in terms of score value.

4. Result Analysis

In section 4, Water Treatment Plant Data Set analyzes the result work with existing K-means algorithm with Seed Affinity Propagation (SAP) algorithm and Weighted Clustering Ensemble (WCE) algorithm. The experimental values are shown through the table structure and graph form.

Table 1 Tabulation of Execution Time

Data Objects 'D'	Execution Time (sec)		
	SAP Algorithm	WCE algorithm	HH K-means Clustering
5	102	105	89
10	110	104	92
15	142	121	114
20	151	138	125
25	156	149	134
30	179	165	152
35	212	192	176

Table 1 describes the execution time based on the data objects 'D'. The execution time is decreased in the HH K-means Clustering, since the centroid value is computed qith mean and median factor.

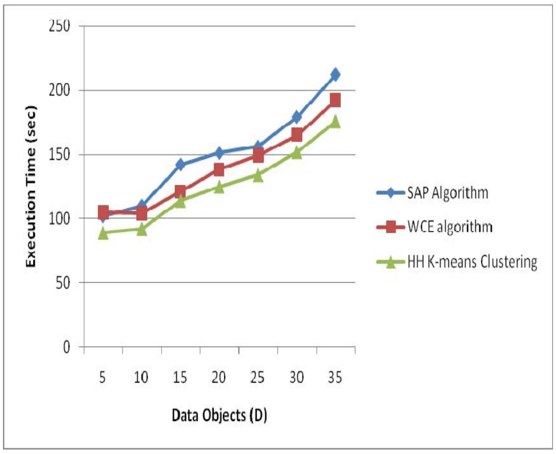

Fig 5 Execution Time Measure

Fig 5 describes the execution time based on the data objects. The Iterative Relocation (IR) technique computes the centroid value which is used to adjoin (i.e.,) arrange in a nearby distance to reduce the execution time in

HH K-means clustering. The mean and median computation is also carried out for centroid computation, where the execution time is reduced to 12 – 19 % when compared with SAP Algorithm [1]. The data objects arranged to the nearby distance of centroid point in HH K-means clustering, which reduces the execution time by 5 – 15 % when compared with WCE algorithm [2].

Table 2 Tabulation of Clustering Rate

No. of Objects	Clustering Rate (Success %)		
	SAP Algorithm	WCE algorithm	HH K-means Clustering
7	76	81	86
14	79	83	91
21	80	89	93
28	82	90	95
35	83	91	96
42	84	91	97
49	85	90	98

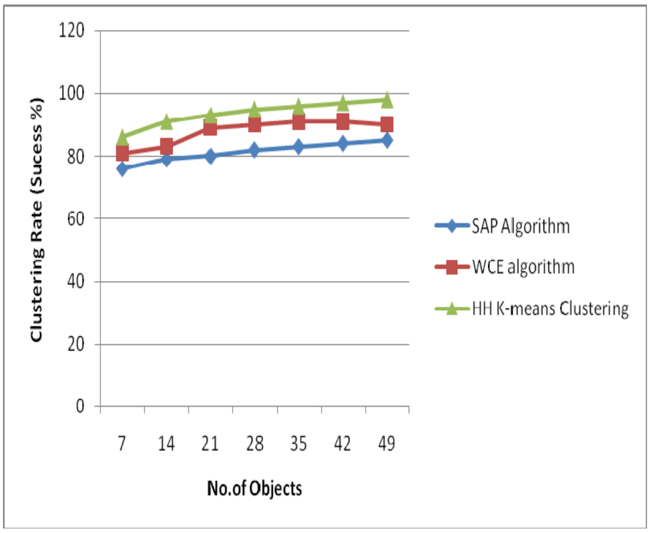

Fig 6 Performance of Clustering Rate

Table 2 and Fig 6 illustrate the cluster rate based on the objects 'D' used for the experimental work. HH K-means cluster apply a new distance measurement to compute the centroid value. The combination of clustering effect thoroughly performs the numerical experiments and improves the cluster rate by 13-16 % when compared with the SAP Algorithm [1]. Huddle and K-means harmonic combination improves the cluster rate by 4- 9 %when compared with the WCE algorithm [2] using the screening process at the initial stage of processing.

Table 3 Tabulation of Entropy Level

Iterations	Entropy Level (Fractional Percent)		
	SAP Algorithm	WCE algorithm	HH K-means Clustering
2	0.32	0.29	0.28
4	0.33	0.31	0.30
6	0.36	0.35	0.32
8	0.39	0.36	0.34
10	0.41	0.40	0.37
12	0.45	0.41	0.39
14	0.49	0.46	0.41

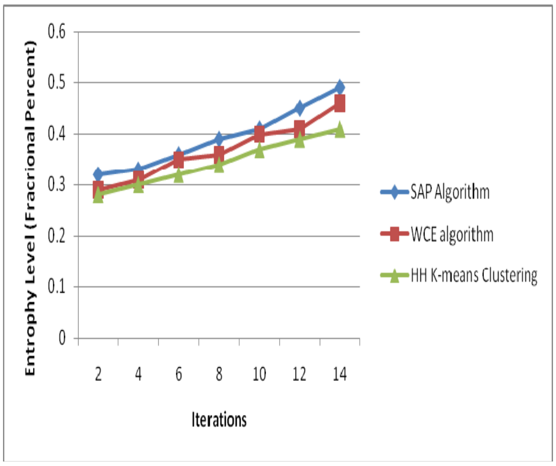

Fig 7 Measure of Entropy Level

Fig 7 illustrates the entropy level (i.e.,) disorder level on Huddle and K-means harmonic clustering process. The entropy level is reduced to 9 – 16 % when compared with the SAP Algorithm [1] by using the decomposed series method. As the iteration gets increased, the entropy level is reduced and measured in terms of fractional percent. The decomposed series in HH K-means clustering combines the newly updated water treatment plant attributes to the related cluster by reducing the entropy level by 3 – 10 % when compared with the WCE algorithm [2].

Table 4 True Positive Rate Tabulation

Object Count 'D'	True Positive Rate (%)		
	SAP Algorithm	WCE algorithm	HH K-means Clustering
5	73.61	78.45	80.12
10	74.25	81.15	85.45
15	76.82	83.78	86.66
20	78.74	86.42	89.45
25	79.89	88.73	92.78
30	82.45	90.45	93.15
35	83.63	91.72	95.86
40	83.81	93.19	97.49

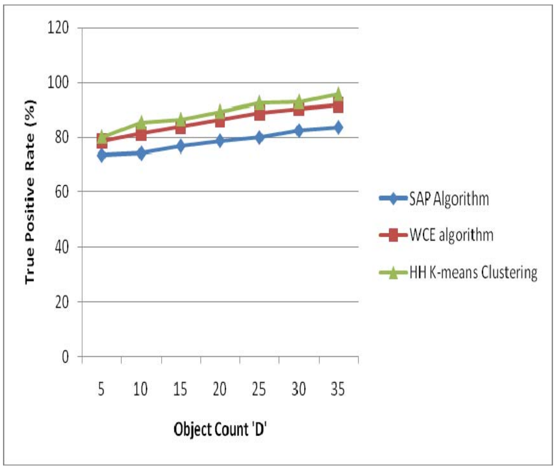

Fig 8 True Positive Rate Measure

Fig 8 demonstrates the true positive rate based on the object count 'D'. The HH K-means clustering computes the centroid value of the data objects 'x' and 'y' to easily attain the higher true positive rate. The positive rate is improved by 8 – 16 % when compared with the SAP Algorithm [1] by using the mean computation [as shown in Eqn (3)] of HH K-means clustering. The true positive rate is also enhanced by 2 – 5 % when compared with the WCE algorithm [2] by using constant while processing.

Table 5 Tabulation for Net Similarity Level

No. of instances	Net Similarity Level (Similarity Score Value)		
	SAP Algorithm	WCE algorithm	HH K-means Clustering
50	33	37	40
100	65	69	77
150	90	90	102
200	145	166	175
250	178	190	205
300	263	297	310
350	282	302	334

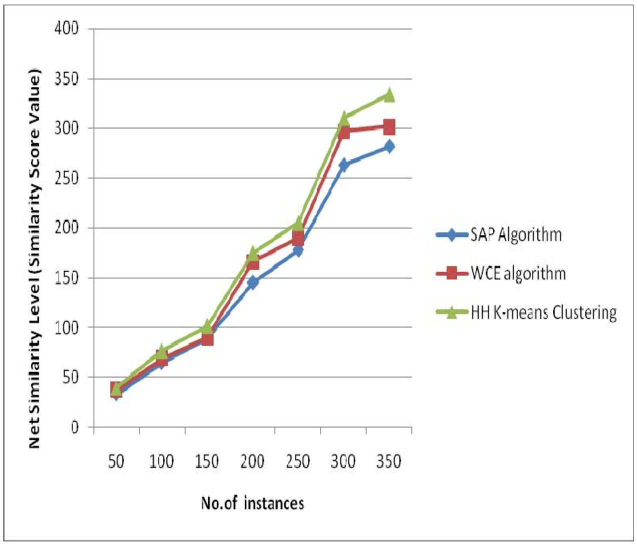

Fig. 9 Net Similarity Level Measure

Fig 9 describes the net similarity level based on the instances. As the instances count increases, the HH K-means Clustering improves the net similarity level by 13 – 21 % when compared with the SAP Algorithm [1]. The net similarity level improved while using adjoining centroid arrangement process on the IR technique. HH K-means clustering arrange objects to the nearby distance of centroid point thereby improves the similarity level by 4 – 13 % when compared with the WCE algorithm [2].

Finally, HH K-means Clustering improves the performance function by computing the centroid value. The mean and median value are also used to exactly plot the distance on the single measure in HH K-means Clustering, thereby reducing the running time. HH K-means clustering provide improved performance result on all the parametric factors used in the data mining clustering system.

5. Related Work

Clustering has been predictable as a significant and valuable capability in the data mining field. Fuzzy similarity based self-constructing algorithm as demonstrated in [3] have one mined out feature for every cluster. The extracted feature, equivalent to a cluster, is a weighted mixture of the words restricted in the cluster. By fuzzy similarity based self-constructing algorithm provides the derived membership functions which are equivalent for the real allocation of the training data. DENsity Conscious Subspace clustering in [7] follows the divide-and-conquer scheme to competently determine clusters. The DENsity Conscious Subspace clustering refers to the observable fact with effective region density on different subspace cardinalities.

Fast clustering-based feature selection algorithm (FAST) as described in [6] partitions the clusters using the minimum-spanning tree (MST) clustering method. FAST algorithm powerfully relates the target classes from each cluster to form a subset of features but does not calculate the different correlations. Profile Support Vector Machine (PSVM) partitions the training examples into clusters in [14] and calculate the correlation point by building a linear SVM model for each cluster. The PSVM consumes more execution time while computing the distance measure on each clustering process.

6. Conclusion

Huddle based Harmonic K-means Clustering extensively used for clustering large sets of data objects. Through the distance metric computation on the huddle based harmonic achieves the better clustering effects with the minimal execution time. HH K-means Clustering effectively performs the screening process by removing the non-cluster data. Iterative Relocation (IR) technique is also developed effectively to improve the speed of the clustering process with initialization, relocation of objects and cluster updating process. Huddle based Harmonic with K-means data objects add some improved conditions and factors to develop the Clustering algorithm. Theoretical analysis and experimental result shows that HH K-means Clustering reduces the execution time by 9.751 % and improves the cluster rate of about 6.69 % when compared with WCE algorithm. The result shows the improved net similarity level and minimal entropy level in Huddle based Harmonic K-means Algorithm.

References

[1] Renchu Guan., Xiaohu Shi., Maurizio Marchese., Chen Yang., and Yanchun Liang., "Text Clustering with Seeds Affinity Propagation," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 4, APRIL 2011

[2] Yun Yang., and Ke Chen., "Temporal Data Clustering via Weighted Clustering Ensemble with Different Representations," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 2, FEBRUARY 2011

[3] Jung-Yi Jiang., Ren-Jia Liou., and Shie-Jue Lee., "A Fuzzy Self-Constructing Feature Clustering Algorithm for Text Classification," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 3,2011

[4] Duc Thang Nguyen, Lihui Chen., and Chee Keong Chan., "Clustering with Multiviewpoint-Based Similarity Measure," IEEE

TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 6, JUNE 2012
[5] Nenad Tomasev, Milos Radovanovic, Dunja Mladenic, and Mirjana Ivanovic., "The Role of Hubness in Clustering High-Dimensional Data," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, REVISED JANUARY 2013
[6] Qinbao Song., Jingjie N.i, and Guangtao Wang., "A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 1, JANUARY 2013
[7] Yi-Hong Chu, Jen-Wei Huang, Kun-Ta Chuang, De-Nian Yang., and Ming-Syan Chen., "Density Conscious Subspace Clustering for High-Dimensional Data," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 1, JANUARY 2010
[8] Sharadh Ramaswamy., and Kenneth Rose., " Adaptive Cluster Distance Bounding for High-Dimensional Indexing," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011
[9] HANS-PETER KRIEGEL., PEER KROGER., and ARTHUR ZIMEK., "Clustering High-Dimensional Data: A Survey on Subspace Clustering, Pattern-Based Clustering, and Correlation Clustering," ACM Transactions on Knowledge Discovery from Data, Vol. 3, No. 1, Article 1, Publication date: March 2009.
[10] Xiaojun Chen., YunmingYe., XiaofeiXu., JoshuaZhexueHuang., "A feature group weighting method for subspace clustering of high-dimensional data," Pattern Recognition., Elsevier journal., 2012
[11] Hung-Leng Chen., Ming-Syan Chen., and Su-Chen Lin., "Catching the Trend: A Framework for Clustering Concept-Drifting Categorical Data," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 5, MAY 2009
[12] Ben Kao., Sau Dan Lee., Foris K.F. Lee., David Wai-lok Cheung., and Wai-Shing Ho., "Clustering Uncertain Data Using Voronoi Diagrams and R-Tree Index," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 9, SEPTEMBER 2010
[13] Xiankun Yang., Weihong Cu., "Novel Spatial Clustering Algorithm Based on Delaunay Triangulation," Journal of Software Engineering & Applications, 2010
[14] Haibin Cheng., Pang-Ning Tan., and Rong Jin., "Efficient Algorithm for Localized Support Vector Machine," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010