ISSN: 2321-7782 (Online)
Volume 2, Issue 9, September 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

# An Experimental analysis of Parent Teacher Scale Involvement with help of K Mean Clustering Technique using Matlab

**Dr. Manju Kaushik**[1]
Associate Professor
JECRC University
Jaipur – India

**Bhawana Mathur**[2]
Research Scholar
JECRC University
Jaipur – India

***Abstract:*** ***Researcher collect data through Parent Teacher Scale Involvement psychology Education Test in the form of Questionnaire. Researcher collect random sample of 28 students' data from School, Jaipur as a random selection of Students of Class 1stto 10 th Class. To make cluster, Find Distance point from the cluster, Find Minimum distance from each point of data sets. Assign Points Pi to Centroids Ck. Update centroids .Do Iteration until centroid converges or uses loop. Researcher makes Questionnaire of 34 questions. Three parameters are like these School involvement, home involvement, parent teacher school involvement. With the help of K Means Technique, using Matlab Object Oriented Software System. Through K Means Methodology Computation Large data sets various parameters, Questionnaire's Question, and their values after computational K Means ,Optimized the results in object oriented system like Matlab.***

***Keywords: Questionnaire cluster, Distance point, Minimum distance Centroids, data sets, School Involvement.***

## I. INTRODUCTION

The primary objective of the proposed clustering methodology is to provide a general but illuminating view of a software system that may lead engineers to useful conclusions concerning its maintainability. This data mining technique is useful for Similarity/Dissimilarity analysis; in other words it analyzes what data points are close to each other in a given dataset. This way, mutually exclusive groups of classes are created, according to their similarities and hence the system comprehension and evaluation is facilitated. Thus, maintenance engineers are provided a panoramic view of a system's evolution, which helps them in revising the system's maintainability, studying the classes' behavior from version to version and discovering programming patterns and "unusual" or outlier cases which may require further attention. An approach is to combine all the data sets (the data points corresponding to classes) into a large data set.

## II. LITERATURE REVIEW

"Improving the Accuracy and Efficiency of the k-means Clustering Algorithm", an improvement in K-means clustering is shown. The first phase of K-means clustering algorithm, the initial centroids are determined systematically so as to produce clusters with better accuracy [1.]

"An Efficient K-Means Clustering Algorithm for Reducing Time Complexity using Uniform Distribution Data Points", the uniform distribution of the data points is discussed that how this approach reduce the time complexity of the K-means clustering algorithm [2.] By using this approach the elapsed time is reduced and the cluster is of better quality. A very good method is used for finding the initial centroid. In this initially, the distance between each data points is computed.

"An Iterative Improved k-means Clustering" discuss an iterative approach which is beneficial in reducing the number of iterations from k-mean algorithm, so as to improve the execution time or by reducing the total number of distance calculations [3.] So Iterative improved K-means clustering produces good starting point for the K-means algorithm instead of selecting them randomly. And it will lead to a better cluster at the last result.

 

"Refining Initial Points for K-Means Clustering" discuss Practical approaches to clustering use an iterative procedure (e.g. K-Means, EM) which converges to one of numerous local minima. It is known that these iterative techniques are especially sensitive to initial starting conditions [4].

"Comparison of various clustering algorithms", Comparative study of clustering algorithms across three different datasets is performed. The algorithms under investigation are partitioning based i.e K-means, Farthest First, Expectation maximization and Non Partitioning based i.e Density based, Hierarchical based and Cobweb[5].

"Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," Clustering analysis method is main analytical methods in data mining, the method of clustering algorithm will influence the clustering results directly. The standard k-means clustering algorithm and analyzes the shortcomings of standard k-means algorithm, such as the k-means clustering algorithm has to calculate the distance between each data object and all cluster centers in each iteration, which makes the efficiency of clustering is not high[6].

"An Efficient enhanced k-means clustering algorithm," In k-means clustering, Researcher are given a set of n data points in d-dimensional space $\mathbb{R}^d$ and an integer k and the problem is to determine a set of k points in $\mathbb{R}^d$, called centers, so as to minimize the mean squared distance from each data point to its nearest center[7] .

The issue of identifying iterative records issue is one of the challenging issues in the field of databases. Emergence of modern techniques for scientific data collection has resulted in large scale accumulation of data per-training to diverse fields[8].

"Extensions to the k-means algorithm for clustering large data sets with categorical values," The k-means algorithm is well known for its efficiency in clustering large data sets. Whenever, working only on numeric values prohibits it from being used to cluster real world data containing categorical values[9].

"An Efficient k-means Clustering Algorithm: Analysis and Implementation", In k-means clustering, we are given a set of n data points in d –dimensional space R d and an integer k and the problem is to determine a set of k points in R d , called centers, so as to minimize the mean squared distance from each data point to its nearest center[10].

"Research issues on K-means Algorithm: An Experimental Trial Using Matlab", It is considered that the k-means algorithm is the best-known squared error-based clustering algorithm, is very simple and can be easily implemented in solving many practical problems[11].

"Effective Communication between Parents and Teachers", Parental involvement enhances academic performance. The more intensely the parent is involved, the greater chance of academic success[12].

## III. METHODOLOGY OF K-MEANS ALGORITHM FOR SIMULINK

**K-means clustering** is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in mining. The method is simulated on matlab (matrix laboratory). The algorithm Used in simulation is given below:

Step 1: Initialization

1.1 Input the number of classes with their attributes and Initialize k=n (K=2)

1.2 Take any random objects as the initial centroids.

1.3 Input number of k cluster and Randomly Select k Points, Assign in cluster Ci

Step 2: Classification

2.1 Compute the distance using most popular distance measure is city block distance or Manhattan between classes and randomly choose objects.

 

2.2 objects are including to the group related to this centroid and For Each points p From data sets and Find Distance point and cluster, Find Minimum distance, Assign points $P_i$ to Centroids $C_{k.}$ and Update centroids.

Step 3: Centroids calculation

3.1 For each group generated in the previous step, its centroid is recalculated.

Step 4: Come together or towards the same point's condition

4.1 Stopping when reaching a given number of iterations.

4.2 Stopping when there is no exchange of objects among groups.

Step 5: If the step 4 is not satisfied then steps 2 to step 4 must be repeated.

Step 6: Produce the population of classes in a given k cluster.

Step 7: Finally perform chi-square test to test the goodness of fit, if it require.

## IV. K MEANS CLUSTERING USING MATLAB

Idx = KMeans(X, K) partitions the points in the n by p data matrix X into K Clusters.

This iterative portioning minimize the sum; overall cluster of the with in cluster sum of point –to – cluster –centroid distances.

Rows of X correspond to points, columns .corresponds to variables. K means returns on n by 1 vector idx containing the cluster indices of each point.

By default, K -Means uses Squard Euclidian distances.

When X is a vector, K Means treats it as an n by 1 data Matrix regardless of its orientation.

[idx,c]= KMeans(X,K) returns the k cluster centroid locations in the k by p matrix c.

[idx,c,sumd]=kMeans

**Table I**
**Dimension of School Involvement**

| Percentile | Range of Score | Category/Parameter |
|---|---|---|
| Above $P_{66}$ | 25 and above | High Parent Involvement |
| $P_{66}$ | 16-24 | Average Parent Involvement |
| $P_{66}$ | 15 and Above | Low Parent Involvement |

**Table II**
**Statistics of the Data Used**

| Parent Teacher involvement in School involvement | No of Student | Dimension (Total Number of Question ) |
|---|---|---|
| Data From Questionnaire | 28 | 10 |

Average individual School involvement= 1/ N ($\sum_{i=1}^{n}(i/n \sum_{i=1}^{n} xi$ ))

Here, it is We take K=3 Clusters, parameters and Values store in a Variables. Parameters are High Parent Involvement, Average Parent Involvement and Low Parent Involvement. Find out maximum similarity between intra Cluster Distance. In Between Inter Cluster Distance have low similarities.

 

## V. EXPERIMENTAL ANALYSIS AFTER SIMULATION

>> a=[21 29 25;22 32 22 ;25 40 26; 22 31 25;26 34 30;24 34 30;24 34 24;23 36 25;27 34 25;25 34 23;25 36 26;22 31 22;25 35 25;27 35 26;25 36 27;24 33 26;;22 31 22;24 34 24;21 34 15;25 33 24;23 40 30;30 38 27;26 34 28;25 36 23; 24 33 26;24 24 21;22 38 30;26 32 23;30 38 30]

>>idx=kmeans(a,3)

>> [idx,c]=kmeans(a,3)

c =

24.9231 34.2308 24.6154; 25.6667 36.8889 28.6667; 22.0000 30.2857 21.7143

>> [idx,c,sumd]=kmeans(a,3)

[idx,c,sumd,D]=kmeans(a,3)

c =

25.6667 36.8889 28.6667; 24.9231 34.2308 24.6154; 22.0000 30.2857 21.7143

sumd =

132.8889; 54.3077; 132.8571

c =

24.9231 34.2308 24.6154; 25.6667 36.8889 28.6667; 22.0000 30.2857 21.7143;

sumd =

54.3077; 132.8889; 132.8571

D =

42.8994 97.4568 13.4490; 20.3609 81.7901 3.0204; 35.2071 17.2346 121.7347; 19.1302 61.5679 11.3061; 30.2071 10.2346 98.4490; 29.8994 12.9012 86.4490; 1.2840 32.9012 23.0204; 6.9763 21.3457 44.4490; 4.5148 23.5679 49.5918; 2.6686 40.9012 24.4490; 5.0533 8.3457 60.0204; 25.8225 92.5679 0.5918; 0.7456 17.4568 42.0204; 6.8225 12.4568 65.5918; 8.8225 4.0123 69.5918; 4.2840 25.0123 29.7347; 25.8225 92.5679 0.5918; 1.2840 32.9012 23.0204; 107.8994 216.9012 59.8776; 1.8994 37.3457 21.5918; 65.9763 18.5679 164.0204; 45.6686 22.7901 151.4490; 12.6686 8.9012 69.3061; 5.7456 33.3457 43.3061; 4.2840 25.0123 29.7347; 118.5917 227.6790 44.0204; 51.7456 16.4568 128.1633; 8.7456 56.1235 20.5918; 68.9763 21.7901 192.1633.

## VI. OVERVIEW OF K MEANS CLUSTERING ALGORITHMS

The K Means Algorithms can work with compact and hyper Spherical clusters .The Time Complexity of K Means is O (NKa).K Means can be used for large data sets[10].

**K Means uses two phases** Iterative Algorithms to minimize the sum of point to centroid distances, summed overall K Clusters.

1. The first phase uses batch phase updates, Where each iteration consists of reassigning points of their nearest cluster centroid all at once, followed by recalculation of Cluster Centroids .This phase occasionally does not converge to solution that is a local minimum, that is portion of the data where moving any single point to a different cluster increases the total sum of distances.

 

It is used for small data sets .The batch phase is fast but potentially only approximates a solution as a starting point for the second phase.

2. The Second Phases uses online updates, where points are individually reassigned if doing so will reduce the sum of distances and cluster centroids are recomputed after each reassigned .Each iteration during the second phase consists of one pass through all points.

   The Second phase will converge to a local minimum although there may be other local minimum with lower total sum of distances.

## VII. Implementation Issue

The problem of finding the global minimum can only be solved in general by an exhaustive choice of starting points, but using several replicates with random staring points typically results in a solution that is a global minimum.

## VIII. Importance of Parental Involvement

- Significant involvement most likely develops when schools actively seek out ways for parents to get involved.
- Parental involvement lifts teacher morale.
- Parental involvement benefits both children and parents. Parents will gain a better understanding of school curriculum and activities and communicate better with their children.
- Time constraints are the greatest barrier to parental involvement. Collaborate with the teacher to find ways to work around schedules[12].

## IX. Implementation of K Means in Matlab

About Matlab:-Mathematical work is leading developer of Mathematical Computing Software for computing software for engineers and scientist. The product of maths work matlab is a programming environment for algorithm development, data analysis, Visualization and numerical computation .We can use matlab in a wide range of applications, including signal processing, communications, Control design, test and measurement, financial modeling and analysis and computational biology.

For a million of Engineers and academia, Scientist in industry, mat lab is the language of technical Computing.

The two key features of k-means which make it efficient are often regarded as its biggest drawbacks:

- Euclidean distance is used as a metric and variance is used as a measure of cluster scatter.
- The number of clusters k is an input parameter: an inappropriate choice of k may yield poor results. That is why, when performing k-means, it is important to run diagnostic checks for determining the number of clusters in the data set.
- Convergence to a local minimum may produce counterintuitive ("wrong") results.

## X. Scope of K Means Methodology

The scope of the k means methodology is to facilitate maintenance engineers to identify classes which are fault prone and more difficult to understand and maintain as well as to study the evolution of a system from version to version, and its classes' dynamics.

## XI. Convergence of K Means and K Means Start with Local Minima

K-means clustering does not guarantee you global optimum (although I'd not call K-means a "heuristic" technique). However you can do this: run K-means a number of times, each time with different random initial centres seed, and obtain a set of final cluster centres each time. If these sets appear similar enough - in the sense that you can easily identify the "same" final

 

centres across the runs - then you are surely close to the global optimum. Then just average those corresponding final centres across the runs and input the obtained averaged centers as initial ones for one final run. That run is almost sure to give you the global optimum solution.

K-means assignes, at each iteration, each object to the closest cluster centre. After all objects were thus assigned, the K centres are updated. It thus appears that a centre moves further towards the set of objects that were already "its" objects. That's why each iteration is an improvement, and the optimum - local or global, dependent on the initial centres choice - is reached. The optimized function is the pooled within-cluster sum-of-squares (because mean is the locus of minimal SS deviations from it), which is equivalent to minimizing the pooled within-cluster sum of pairwise squared Euclidean distances normalized by the respective number-of-objects in a cluster.

## XII. Conclusion and Future Work

In this research work, the development of a methodology based on the clustering data mining technique was presented. It consists of two steps: I. A separate clustering step for every version of a system to assist software system's evaluation in means of maintainability. II. A macro-clustering analysis in order to study the system's dynamics from version to version.

It is Sensitivity to initial conditions. A clustering with smaller K can have a lower SSE than a poor Clustering with Higher K.

## XIII. Experiments Result

c =

24.9231 34.2308 24.6154; 25.6667 36.8889 28.6667; 22.0000 30.2857 21.7143

sumd =

54.3077; 132.8889; 132.8571.

Test are done on the numbers of iterations in K means algorithms for reaching the minimum .The total sum of distances decreases at each iteration as K Means reassigns points between clusters and recomputed cluster Centroids. Following results shows that K Means on various parameters for analyzing the maintainability should be analyzed at the design phase of software development life cycle itself.

Various such data mining clustering techniques can be analyzed on large data sets. Our work with matlab is simple and efficient for statistical analysis Through K-Means clusters datasets using heuristics, it is based on other clustering and location problems. Hence, there is a need to experiment, improve and other algorithms. This can be done faster and effectively using Matlab.

Through K Means Large Data Set of School with the help of various parameters like High Parent Involvement, Average Parent Involvement and Low Parent Involvement, they minimize the sum of point to centroid distances, computationally large data sets Change into optimized Data Sets in Object Oriented Software MATLAB.

Due to Parental Involvement having an awareness of and involvement in school-work, understanding of the interaction between parenting skills and student success in schooling, and a commitment to consistent communication with educators about student progress.

## References

1. 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 2009 Vol I WCE 2009, London, U.K, July 1 - 3, 2009.
2. D. Napoleon & P. Ganga lakshmi, "An Efficient K-Means Clustering Algorithm for Reducing Time Complexity using Uniform Distribution Data Points", IEEE, 2010.

 

3. Madhuri A. Dalal & Nareshkumar D. Harale "An Iterative Improved k-means Clustering" Proc. of Int. Conf. on Advances in Computer Engineering, 2011.
4. Paul S. Bradley & Usama M. Fayyad, "Refining Initial Points for K-Means Clustering", 15th International Conference on Machine Learning, ICML98.
5. Osama Abu Abbas "Comparison of various clustering algorithms" The International Arab Journal of Information Technology, Vol. 5, No. 3, July 2008.
6. Shi Na; Liu Xumin; Guan Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on , vol., no., pp.63,67, 2-4 April 2010 doi: 10.1109/IITSI.2010.74
7. Fahim A.M, Salem A. M, Torkey A and Ramadan M. A, "An Efficient enhanced k-means clustering algorithm," Journal of Zhejiang University, 10(7):1626–1633, 2006.
8. Huang Z, "terin Data Mining and Knowledge Discovery, (2):283–304, 1998.
9. Jiawei Han M. K, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, An Imprint of Elsevier, 2006.
10. Kanungo T., M.D.Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman and Angela Y. Wu ,"An Efficient k-means Clustering Algorithm:Analysis and Implementation".
11. Ortega J. P., Rojas M. D. R. B., Garcia M. J. S.J., "Research issues on K-means Algorithm: An Experimental Trial Using Matlab".
12. King S. H.,Goodwin, A. L, "Effective Communication between Parents and Teachers", Center for Assessment and Intervention Fischler School of Education and Human services Nova South-eastern University,2002.
13. Kaushik M., Mathur B. , "Data Analysis of Students Marks with Descriptive Statistics", in International Journal on Recent and Innovation Trends in Computing and Communication(IJRITCC), ISSN: 2321-8169 ,May 2014, Volume: 2 Issue: 5 1188– 1190, ISRA Journal Impact Factor: 1.376 ,Scientific Journal Impact Factor (SJIF) Value for 2013 is 5.098,Page No.1188-1191.
14. Kaushik M., Mathur B. "Comparative Study of K-Means and Hierarchical Clustering Techniques"(Paper ID- IJSHRE-2653), Volume 2 Issue 6, International Journal of Software & Hardware Research in Engineering (IJSHRE) ISSN NO: 2347-4890, SJIF - 2.669,Page No -93-98.

 