International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307

Analysis of Advanced K-Mean Clustering For Various Data Sets In MATLAB

Varsha Bansal[1], Mr. Mahesh Kumar[2]

[1]Ganga Institute of Technology and Management, MDU Rohtak, Haryana, India
[2]Assistant. Professor, Department of Computer Science and Engineering, GITAM, MDU Rohtak, Haryana, India

ABSTRACT

Data mining is a process that uses different technologies, using "patterns" or "knowledge" in the data. These are some data mining techniques related to classification, clustering, and association principles. Among them there are some "similar" things between them, called "balanced" Belongs to other clusters. This means using cluster analysis to find things projects in one group will be different from projects and other group objects. Is a cluster Neutral learning skills. Neutral learning means learning first without knowledge Sample rating K-minerals are a specific way of cluster observations. "K" indicates the number of specific clusters. There are different distance steps to determine Observe the cluster. The purpose of the algorithm is to minimize Measurements between cluster centers and their given observations Track observations in any cluster and eliminate shortest distance measurements Get, finish. This proposed work having global data sets work on MATLAB which found very effectively when it come to underlying structure of modified K-mean. This will have the capability of finding the cluster area efficiently.

Keywords : K-Mean, Global Variables , MATLAB.

I. INTRODUCTION

Clustering populations or data points are divided into groups, such as data points in a group, rather than other data points in other groups. In short, the goal is to separate groups and assign them to clusters with these symptoms. The high dynamic range image area uses devices with limited dynamic range to handle the capture and processing of a wide range of brightness levels worldwide. In this paper, we propose a new algorithm based on the modified clustering method. By using dynamic programming, we can easily resolve significant assemblies and make it a global custom MATLAB.

II. TYPES OF CLUSTERING

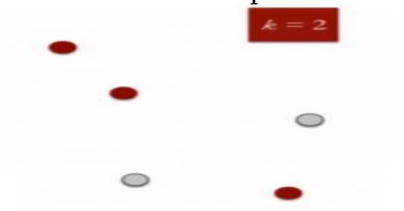
Obviously, clustering can be divided into two subgroups:

- **Hard clustering:** In a difficult cluster, each graph belongs to a poly point. For example, in the above example, each customer is placed in a group of 10 groups.
- **Soft clusters:** In a soft cluster, instead of placing each data point in a separate cluster, the possibility or possibility of the data is included in these clusters. For example, each of the above scenarios may be located in 10 clusters of any retail store.

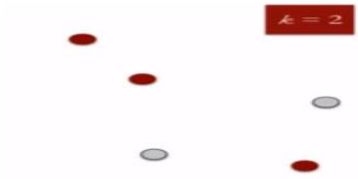
III. K-MEANS CLUSTERING

K means is an iterative clustering algorithm that aims to find local maxima in each iteration. This algorithm works in these 5 steps:

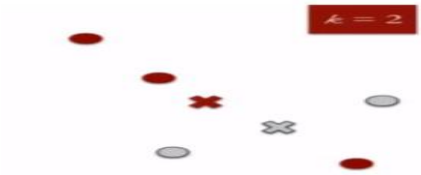
1. Specify the **desired number of clusters K**: Let us choose k=2 for these 5 data points in 2-D space.



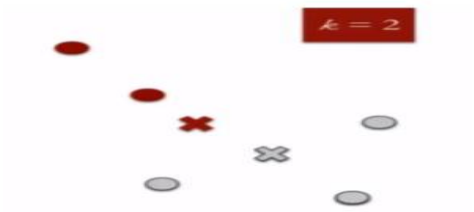
2. **Randomly assign** each data point to a cluster: Let's assign three points in cluster 1 shown using red color and two points in cluster 2 shown using grey colors.



3. **Compute cluster centroids:** The centroid of data points in the red cluster is shown using red cross and those in grey cluster using grey cross.



4. **Re-assign each point** to the closest cluster centroid : Note that only the data point at the bottom is assigned to the red cluster even though its closer to the centroid of grey cluster. Thus, Here assign that data point into grey cluster.



5. **Re-compute cluster centroids** : Now, re-computing the centroids for both the clusters.

Repeat steps 4 and 5 until no improvements are possible: Similarly, we'll repeat the 4th and 5th steps until we'll reach global optima. When there will be no further switching of data points between two clusters for two successive repeats. It will mark the termination of the algorithm if not explicitly mentioned.

The purpose of the data cluster is also called clustering, cluster analysis, which is the natural mode of discovering patterns, points or objects. Visitors (Marriott-Westter Online Dictionary, 2008) Cluster Analysis explains that as a data scoring technique, it shows whether people in a population have different clusters in different types of clusters.The operational definition of clustering can be described below. Look at the representation of things and find K groups based on similar equations. Display The culture can vary depending on the shape, size and density. The noise in the data exceeds seven in cluster detection. An ideal cluster can be described as a set of compact and discrete points. In fact, a cluster is a spiritual institution in the context of things. Its meaning and interpretation require domain knowledge. However, when human beings are three-dimensional potential optimal clustering students, we need automatic algorithms for high-dimensional data. The number of clusters for this number and this number is unknown, resulting in thousands of clustering algorithms that have been released and appear.Cluster analysis is prevalent in any discipline that involves analysis of multivariate data.

IV. METHODOLOGY

K-Means algorithm is the most popular distributed distributed cluster technology. This is an unwanted algorithm for clustering. It selects the center well, selects the center and its intensity and attribute data points, and finds the distance. The data points are similar to the clusters that are assigned to the center. Go away Calculate the new "center" center, so form a cluster by finding data points near the cluster.

The K-Means algorithm stage can be mentioned as follows:

1. Select from random numbers and make them initially centered.
2. Select a data point from the collection, compare it with each center, and if you encounter the data point center, assign it to the center's cluster.
3. When assigned to each data point cluster, each total recalculates the cluster's central value.
4. Erase 2 to 3 feet unless the data point reaches its last cluster to another cluster (shut off quality met).

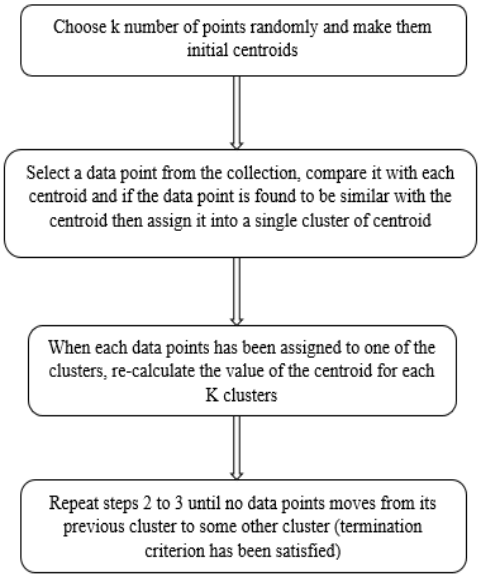


Fig1: K-Means Clustering Algorithm

Flow of the K-Mean (Modified) Implementation

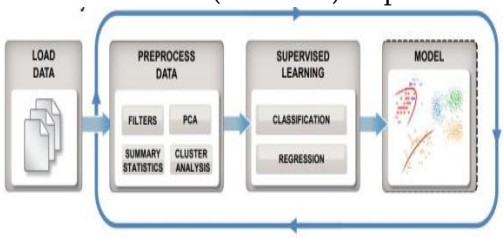


Fig2: Process flow of the K-Mean

V. SIMULATION & RESULT

We basically chose three items as the three main cluster centers. We point to the cluster centers "1, 2, 3, and 4". Use attachments to make a cluster and make clusters near the cluster center. After creating the cluster, we will find the value of the cluster. According to the new meaning, we will create or create a new cluster. As long as we write these values as red, green, blue or dark blue, this process still exists.

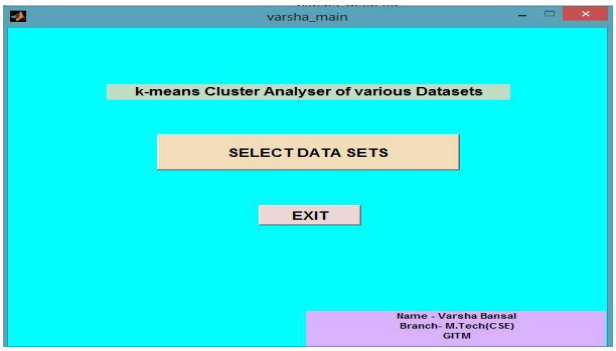


Fig 3: Basic Layout of Proposed GUI in MATLAB

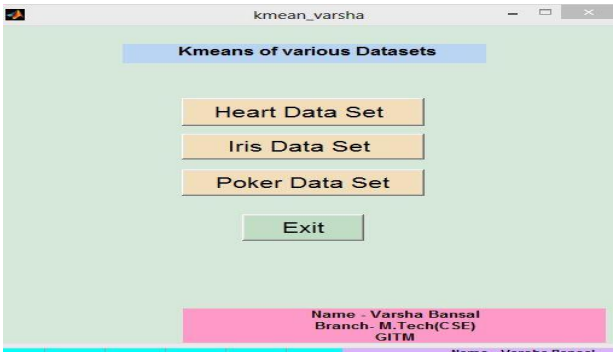


Fig 4: Various Data set for execution

```
MIN value =
    5.2655

MAX value =
  173.5936

MAX Purity overall =
   43.3984

Fitness =
    9.2010

Elapsed time is 0.355094 seconds.
>>
```

Fig 5: Values of Heart Data Sets

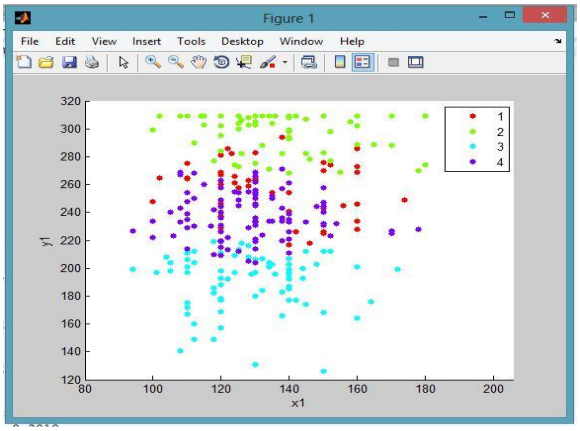


Fig 6: K-Mean Cluster image of Heart Data Sets

```
MIN value =
    0.6618

MAX value =
   64.6328

MAX Purity overall =
   16.1582

Fitness =
   11.4770

Elapsed time is 0.150038 seconds.
>>
```

Fig 7: Values of Iris Data Sets

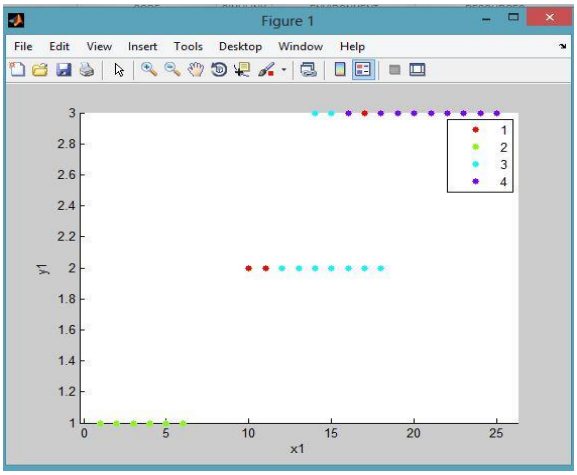


Fig 8: K-Mean Cluster image of Iris Data Sets

```
MIN value =
    2.9544

MAX value =
   17.1158

MAX Purity overall =
    4.2790

Fitness =
    0.9953

Elapsed time is 0.114882 seconds.
>>
```

Fig 9: Values of Poker Data Sets

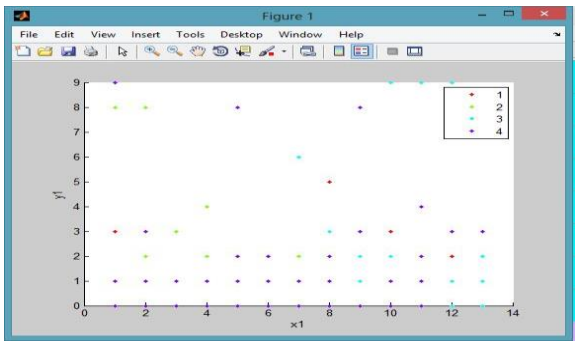


Fig 10: K-Mean Cluster image of Poker Data Sets

VI. CONCLUSION & FUTURE WORK

K-Means is a famous cluster algorithm that has successfully implemented various issues. However, its application is usually limited to small databases. Mahat is a cloud computing that leads to K-Means that runs on the MATLAB system. Due to their affordable and valuable features, this platform may be a passionate technology to solve the huge problem of data that was not in the past. This article deals with a brief introduction around a large document area. It is impossible to describe the detail of each algorithm, this article specifically discusses the methods of assembly, the most used, the pyramid pools and the K-Means sets and the features that apply to the MATLAB Use This idea has been discussed and it involves the necessity of the obligation of any obligation to the party that essentially adds the provider to the provider. Your useful tools are in further study.

VI. ACKNOWLEDGEMENT

I would like to express my appreciation to Mr. Mahesh Kumar (Asstt. Professor) for his guidance and support.Without his valuable assistance, this work would not have been completed I am also thankful to Mrs. Neetu Sharma, the Head of Department of Compuer science in Ganga institute of technology and management for her valuable cooperation and assistance.

VII. REFERENCES

1. Michael Steinbach,George Karypis ,Vipin Kumar, “AComparison of Document Clustering Techniques”.
2. Kris Buchanan, Daniel Gaytan, Lu Xu, Chris Dilay, and David Hilton, “Spatial K-means Clustering of HF Noise Trends in Southern California Waters”.
3. Magnus Oskarsson, “Temporally Consistent Tone Mapping of Images and Video Using Optimal K-means Clustering” , J Math Imaging Vis (2017) 57: pp. 225–238, 2016.
4. Deruo Cheng, Yiqiong Shi, Tong Lin, “Hybrid K-Means Clustering and Support Vector Machine Method for Via and Metal Line Detections in Delayered IC Images” IEEE, 2018

5. Tao Lei, Xiaohong Jia, Yanning Zhang, Lifeng He, "Significantly Fast and Robust Fuzzy C-Means Clustering Algorithm Based on Morphological Reconstruction and Membership Filtering", IEEE, 2017.
6. Ekta Joshi, Dr. D. A. Parikh, "An Improved K-Means Clustering Algorithm" (4)2 : 239-244, 2018.
7. Arpit Bansal, Mayur Sharma, Shalini Goel, "Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining", International Journal of Computer Applications (0975 – 8887) Volume 157 – No 6, January 2017.
8. Marco Capóa, Aritz Pérez a , Jose A. Lozanoa, "An efficient approximation to the K-means clustering for massive data", Knowledge-Based Systems 000 (2016) 1–14, 2016.
9. Yanhui Guo1 , Rong Xia2 , Abdulkadir S¸ engu¨r , Kemal Polat, "A novel image segmentation approach
based on neutrosophic c-means clustering and indeterminacy filtering", 2016.
10. Takayuki Iguchi1 , Dustin G. Mixon1 ,Jesse Peterson1 ,Soledad Villar, "Probably certifiably correct k-means clustering", 2016.
11.]S. Javadi , S.M. Hashemy , K. Mohammadi , K.W.F. Howard , A. Neshat, "Classification of aquifer vulnerability using K-means cluster analysis", Journal of Hydrology 549 (2017) 27–37, 2017.
12. Sina Khanmohammadi, Naiier Adibeig, Samaneh Shanehbandy, "An Improved Overlapping k-Means Clustering Method for Medical Applications", September 16, 2016.
13. Hongfu Liu, Junjie Wu, Tongliang Liu, Dacheng Tao, and Yun Fu, "Spectral Ensemble Clustering via Weighted K-means: Theoretical and Practical Evidence", Ieee Transactions On Knowledge And Data Engineering, Vol. Xxx, No. Xxx, January 2017.
14. Jiahu Qin, Weiming Fu, Huijun Gao, Wei Xing Zheng, "Distributed k-Means Algorithm and Fuzzy c-Means Algorithm for Sensor Networks Based on Multiagent Consensus Theory", IEEE, 2016.
15. Tian-Shi Xu, Hsiao-Dong Chiang, Guang-Yi Liu, and Chin-Woo Tan, "Hierarchical K-means Method for Clustering Large-Scale Advanced Metering Infrastructure Data",IEEE, 2015.
16. Miin-Shen Yanga , Yessica Nataliani, "Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters",2017.