International Journal of Academic Research and Development
ISSN: 2455-4197
Impact Factor: RJIF 5.22
www.academicsjournal.com
Volume 3; Issue 1; January 2018; Page No. 760-764

Distributing Secure Clusters over a Network using Enhanced K Means Clustering & Analysis with Different Clustering Mechanisms

[1] Nandini Babbar, [2] Priyanka
[1] M. Tech (CSE), Chaudhary Devi Lal University, Sirsa, Haryana, India
[2] Assistant Professor, Department of CSA, Chaudhary Devi Lal University, Sirsa, Haryana, India

Abstract
Clustering is the process of grouping objects that belong to a same class. Similar objects are grouped in one cluster & dissimilar objects are grouped in another cluster. Clustering methods can be classified into Partitioning Method, Hierarchical Method, Density-based Method, Grid- Based Method, Model-Based Method, and constraint-based Method. In this research, we propose a latest technique to implement learning capacity & decrease computational intensity of a competitive learning, multi-layered neural network using K-means clustering algorithm's-mean clustering algorithm is a well known partitioning method.

Keywords: clustering, hierarchical method, density-based method, grid-based method

Introduction
Unstructured data refers to information that does not have been a pre-defined data model or is not organized with a pre-defined manner. Unstructured knowledge is normally text seriously, but might control data numbers &facts as well. This result in irregularity & uncertainty that make it complex to comprehend using established programs as compared to data stored in fields form in database in documents.

Clustering
By investigative one or more attributes or classes, you more group person pieces of data value as one to form a structure opinion. At a simple stage, clustering is using one or more attributes as your basis for identifying a cluster of correlating the results. Clustering is valuable to identify dissimilar info since this correlates within other examples so you might see where similarities & ranges agree.

Clustering methods
Partitioning Method
Guess we are certain a database of 'n' objects & partitioning method constructs 'k' Partition of data. Each partition would represent a cluster & $k \leq n$.

Hierarchical Methods
This approach develops a hierarchical routing of a given group of data objects. They could classify hierarchical methods on the basis of how hierarchical decomposition is formed.

Agglomerative Approach
This approach is also known as bottom-up approach. In this, we start within each object forming a separate group.

Divisive Approach
This system is also information top-down approach. They have to start within total of objects in the same cluster. In continuous iteration, a cluster is split up into smaller clusters. It is down until each object in one cluster or termination condition holds. This method is rigid, i.e., once a merging or splitting is done, it could never be undone.

Proposed work
Performance Improvement of Web Usage Mining by Using Learning Based K-Mean Clustering through Neural Network
Due to the growing amount of input available online, www had becoming of valuable resources for information retrievals & knowledge discoveries. Web removal technologies are correct solutions for information discovery on the Web. Knowledge wring from mesh could be used to increase presentation for Web information retrievals, question answering, & Web based data warehousing.

K-mean clustering algorithm
K-mean clustering algorithm is a well-known partitioning method. In it system is classified as fit into one of K-groups. Output of separation system is a group of K clusters, some object of the dataset belonging to one cluster. In each cluster there might be a centroid or a cluster representative. In case where we consider real -valued data, the arithmetic mean of attribute vectors of objects within a cluster provides representative; alternative types of centroid might be required in other cases.

Implementation of k means clustering in remote data transmission
Here we create three clusters. These three generated clusters would be transferred from sender to multiple receivers using Socket mechanism. This is done in the Distributed System so that load is not only on the single server. When we run these files we get the corresponding result in the respective host.
In Figure1 there is one server and three receivers. The data

can send according to their IP addresses. When we want to change the receiver, we can send the corresponding address. Actual file contains all three clusters; we can say that the actual file is contained only by the sender who has to send all these clusters to the separate receivers.

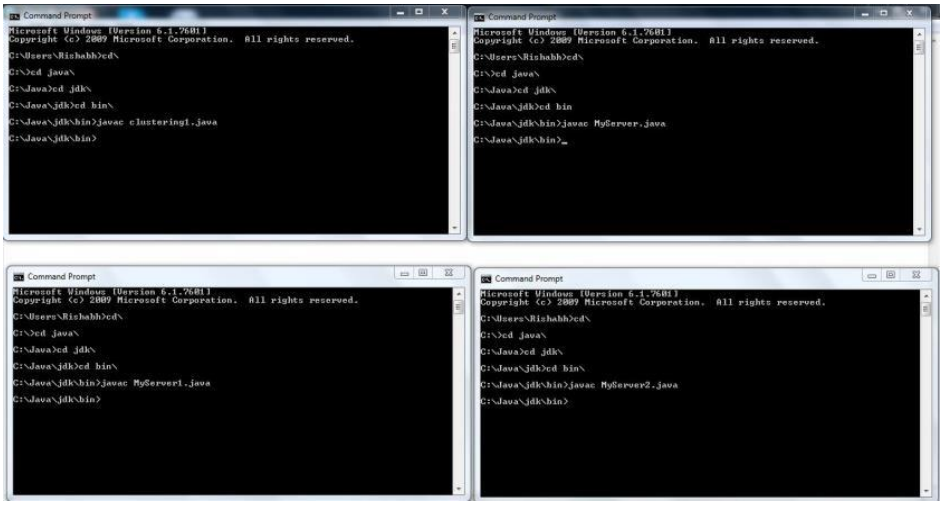

Fig 1: Connection between Client and Server

Fig 2: Cluster at Receiver 1

The receiver 1 is the first using the first cluster which is send by the sender. My Server is the first receiver name and c1.txt is the text file which contains all the items in the cluster one.

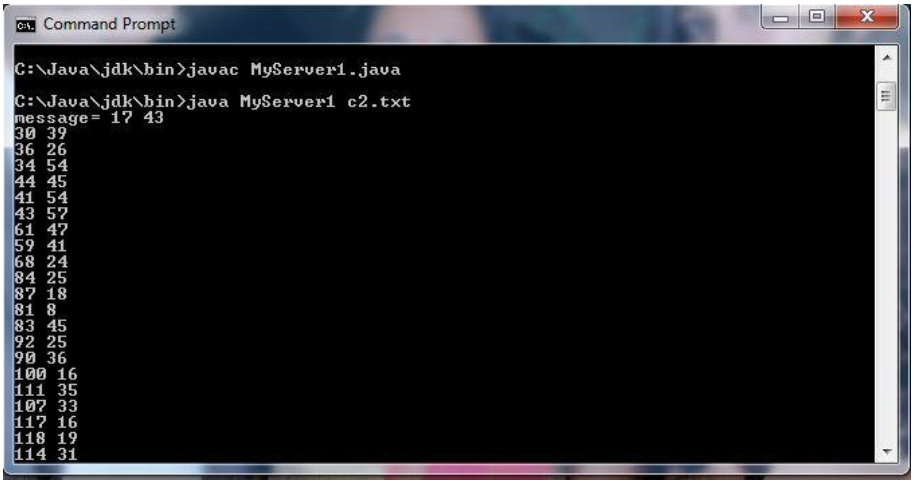

Fig 3: Cluster at Receiver 2

Next is the receiver 2, which is using the second cluster, which is sent by the sender. MyServer1 is the second receiver name and c2.txt is the text file which contains all the items in the cluster two.

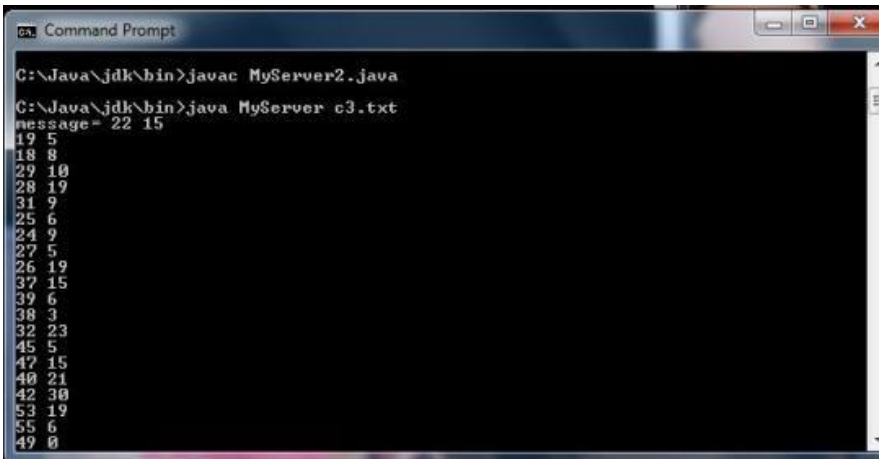

Fig 4: Clusters at Receiver 3

Implementation of dataset in weka

We create the database with the extension of csv (comma separated version). Then open that database in the WEKA tool. Then select the attributes of that particular database. And then check for the different the cluster algorithms.

Table 1: Dataset used in WEKA tool

	A	B	C	D
1	id	name	age	salary
2	1	Naresh	32	20000
3	2	Ram	35	20000
4	3	Pooja	32	40000
5	4	Dimple	32	40000
6	5	Priyanka	32	40000
7	6	Renu	32	40000
8	7	Nandini	32	40000
9	8	Ramveer	32	20000
10	9	Param	32	20000
11	10	Purva	34	20000
12	11	Neha	34	20000
13	12	Rishabh	34	30000
14	13	Gouri	34	30000
15	14	Khushi	34	30000
16	15	Suman	34	30000
17	16	Gautam	34	30000
18	17	Santosh	34	32000
19	18	Aastha	34	32000
20	19	Radhe	34	32000
21	20	Sonu	34	32000
22	21	Sanjay	35	30000
23	22	Deepu	35	30000
24	23	Kiran	35	20000
25	24	Laddu	35	20000

dataset

We get the following conclusion from the above dataset.

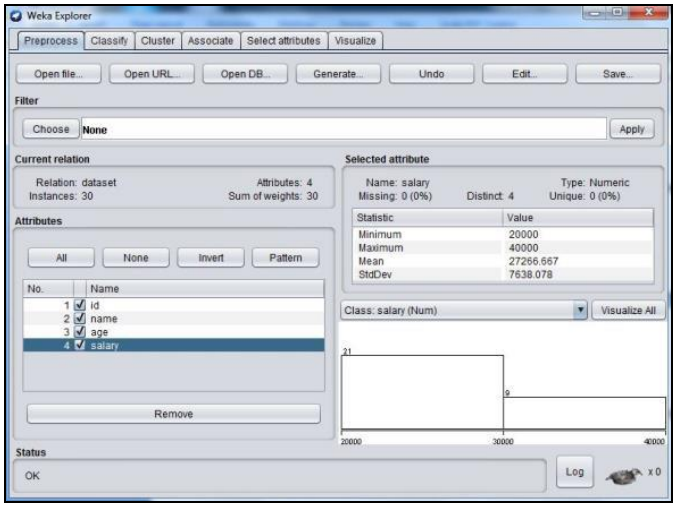

Fig 5: Conclusions from Dataset

There are four attributes in the dataset. There are total 30 instances in this dataset. The mean is 27266.667
The standard deviation is 7638.078

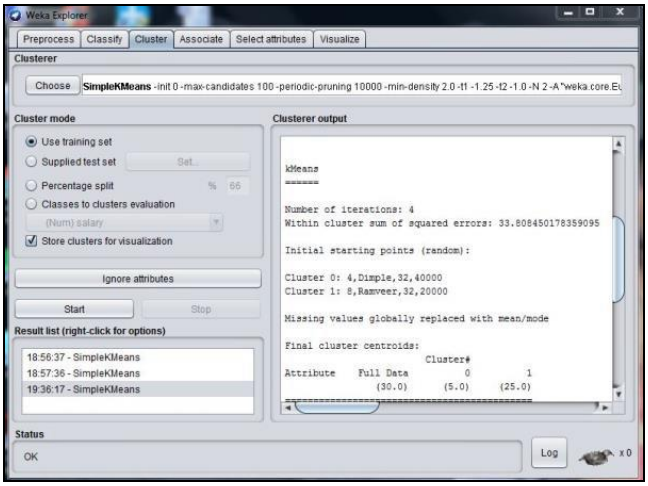

Fig 6: K Means Clustering

Now check for Hierarchical Clustering Algorithm

In algorithm of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types:

Agglomerative: Every observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

Divisive: all comments start in one cluster, & splits are performed recursively as one move down the hierarchy.

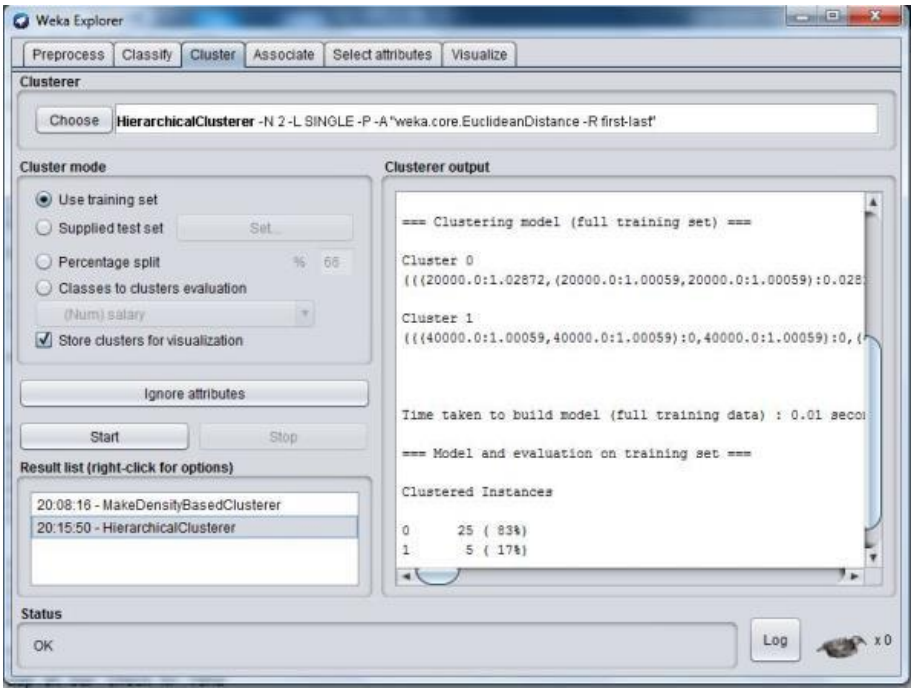

Fig 7: Hierarchical Clustering Algorithm

Now check for farthest first clustering algorithm

The farthest first clustering of a bounded metric space is a sequence of points in the space, where the first point is selected arbitrarily and each successive point is as far as possible from the set of previously-selected points.

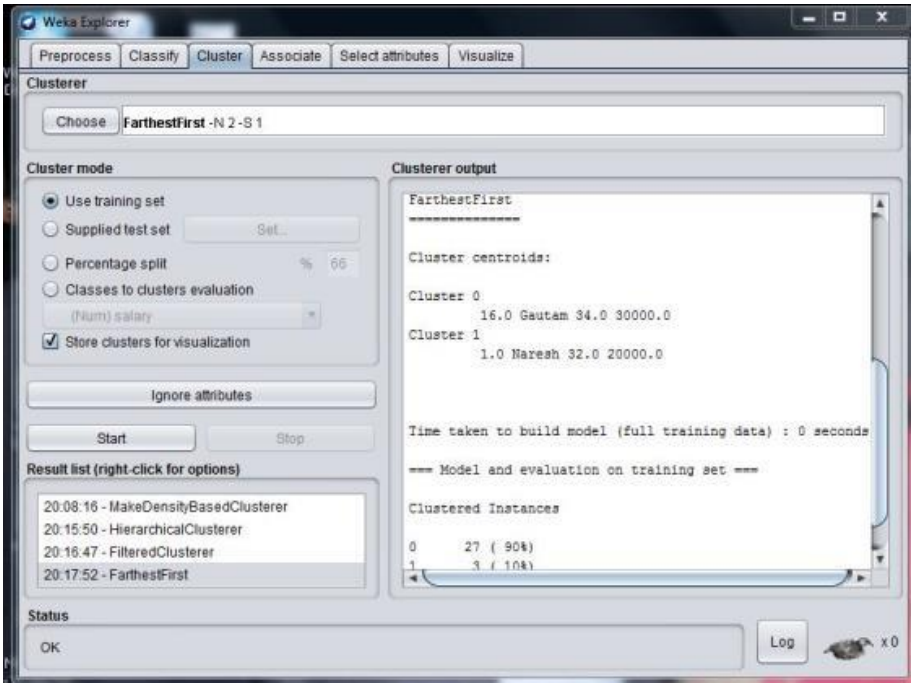

Fig 8: Farthest First Clustering Algorithm

Summary of different clustering mechanisms

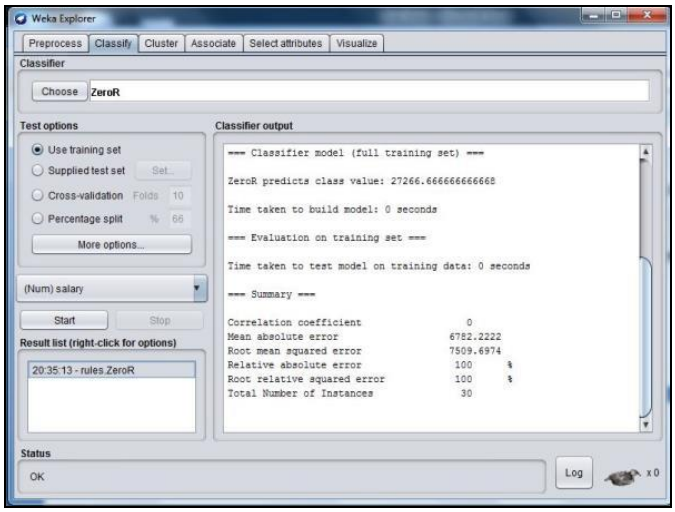

Fig 9: Summary

This figure provides with the summary of the classification done yet.

It will determine the total percentage of the mean absolute error, root mean squared error, relative absolute error and total number of instances occurred.

Conclusion

Clustering is the process of grouping objects that belong to a same class. Similar objects are grouped in one cluster & dissimilar objects are grouped in another cluster. Clustering methods can be classified into Partitioning Method, Hierarchical Method, Density-based Method, Grid- Based Method, Model-Based Method, and constraint-based Method. Clustering analysis is used in several applications like market research, pattern recognition, and data analysis. The major advantage of the Grid Based method is fast processing time.

References

1. Fahim AM. Wrote on An Efficient enhanced k-means clustering algorithm International Journal of Database Theory & Application. 2006.
2. Hong Liu 1, Xiaohong Yu. Wrote on Application Research of k-means Clustering Algorithm in Image Retrieval System. International Journal of Database Theory & Application, 2009.
3. Bhagyashree Umale. Overview of K-means & Expectation Maximization Algorithm for Document Clustering International Conference on Quality Up-gradation in Engineering, Science & Technology, 2011.
4. David Jensen, Jennifer Neville. Data Mining in Social Networks. Computer Science Department Faculty Publication, 2012.
5. Vishal Shrivastava. A Study of Various Clustering Algorithms on Retail Sales Data of Computing, Communications & Networking. International Journal of Advanced Research in Computer Science and Software Engineering, 2012.
6. Kalyani Raval M. Data Mining Techniques International Journal of Advanced Research in Computer Science and Software Engineering, 2012.
7. Navjot Kaur. Wrote on Efficient K-Means Clustering Algorithm Using Ranking Method In Data Mining. International Journal of Research, 2012.
8. Neelamadhab. Survey of Data Mining Applications & Feature Scope. International Journal of Computer Science, Engineering and Information Technology, 2012.
9. Atul Kumar Pandey. Data Mining Clustering Techniques in Prediction of Heart Disease using Attribute Selection Method. International Journal of Science, Engineering & Technology Research (IJSETR). 2013.
10. Bhoj Raj Sharma. Clustering Algorithms: Study & Performance Evaluation Using Weka Tool. International Journal of Current Engineering & Technology, 2013.
11. Nikita Jain, Vishal Srivastava. Data mining techniques: A Survey paper IOSR Journal of Computer Engineering (IOSR-JCE). 2013.
12. Aarti Sharma. Application of Data Mining – A Survey Paper. International Journal of Trend in Research and Development, 2014.
13. Shweta Srivastava. Clustering Techniques Analysis for Microarray Data. International Journal of Computer Science & Mobile Computing A Monthly Journal of Computer Science & Information Technology IJCSMC, 2014.
14. Muhammad Husain Zafar. A Clustering Based Study of Classification Algorithms. International Journal of Database Theory & Application, 2015.
15. Asir Antony D. Performance Analysis on Clustering Approaches for Gene Expression Data. International Journal of Advanced Research in Computer & Communication Engineering, 2016.
16. Fayyad Usama, Piatetsky-Shapiro, Gregory, Smyth Padhraic. From Data Mining to Knowledge Discovery in Databases (PDF), 1996. Retrieved 17 December 2008.
17. Clifton Christopher. Encyclopedia Britannica: Definition of Data Mining, 2010. Retrieved.
18. Hastie Trevor, Tibshirani Robert, Friedman Jerome. The Elements of Statistical Learning: Data Mining, Inference & Prediction, 2009. Retrieved 7 August 2012.