



of clustering like partitioning based methods, hierarchal based methods, Density based methods, Grid based methods, Model based methods etc. Each method itself consists of variety of clustering algorithms. I am using Enhanced K-means Clustering algorithm which comes under Partitioning methods having distance based environment and OC (Orthogonal Partitioning) Clustering algorithm comes under Grid Based methods having grid based environment. Both algorithms belongs two different environments. The Purpose of Enhanced K-Means Clustering algorithm is to find out better initial centroids with reduced time complexity and whole working of this algorithm is based on K-Means Clustering algorithm where K-Means Clustering Algorithm is Distance based Clustering algorithm which defines distance measures from data instances and also find partitions of the distances as like distance between objects within same clusters is minimized and between different clusters is maximized. The purpose of Orthogonal Partitioning (OC) Clustering algorithm creates a hierarchical grid-based clustering model, which means, it creates axis-parallel (orthogonal) partitions in the input attribute space. O-Cluster separates areas of high density by placing cutting planes through areas of low density. The advantages of using Grid based Environment is objects are represented in multi-resolution grid form with higher processing time and independent number of objects. In generally, we divide clustering algorithms into four categories whose names are Hierarchal clustering, Probabilistic clustering, Exclusive clustering and overlapping clustering. We are using Exclusive type of clustering on two separate environments distance based and grid based. It indicates that cluster must belong to one specific

cluster; that specific cluster must not to be considered into any other cluster. For Example: - such kind of clustering is used in the separation of line shows the difference between the existing clusters lies upper and lower boundary of the lines.

## OUR CONTRIBUTION

In our contribution we are proposing an algorithm named as DRID (Distance+Grid) .DRID is used to merge the clusters. This algorithm is used in two different environments named as Distance based environment and Grid based environment. Here, we are using a common strategy (i.e. DRID) for merging clusters in two different environments. Performance of the DRID must be affected by the input given by the user. Performance may be increase or decrease it further depend on the type of the input applied by the user under some specific environment.

### Basic steps of DRID are follows as:-

**Step1.** Find the Euclidian distance.  $ED = D2 - D1$ .  $D1$  is the distance of the first cluster and  $D2$  is the distance from the second cluster and  $ED$  is the Euclidian distance.

**Step2.** On the basis of calculated Euclidian distance, moves centroid towards it.

**Step3.** Find out again distance after moving centroid. If nearest neighbour distance is not found then mark the reassignment and again move to step2.

**Step4.** If nearest neighbour distance is found then mark assignment and set the centroid.

**Step5:** Then Repeat the loop for  $N$  clusters

Step6. End of the loop.

## RESULT ANALYSIS

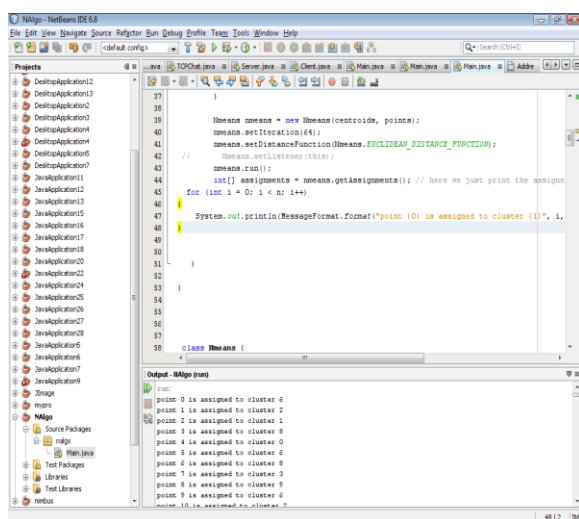


Figure 1: DRID-In Distance Based Environment

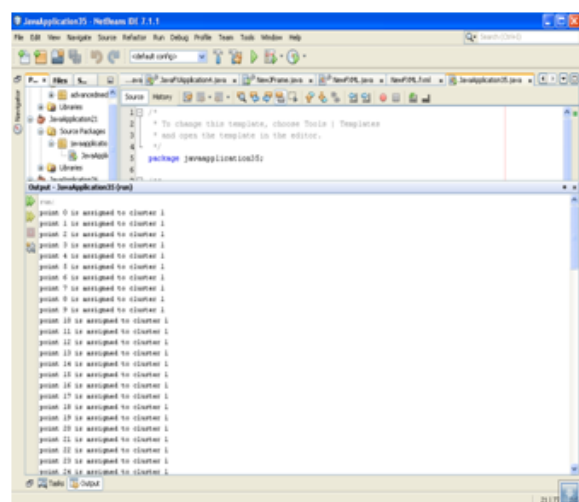


Figure 2: DRID-In Grid Based Environment

When we run DRID in distance based environment then we entered a set of hundred values of array as an input value then the output of the algorithm assigns a **different different ClusterID** for each element of an array at some specific point. And in case of grid based environment when we entered a set of hundred array as an input value then the output of the algorithm assigns a **Same ClusterID** for

each element of an array at some specific point. ClusterID is generated by the algorithm automatically. ClusterID shows the distance between specific clusters with its neighbour cluster. Clusters either are placed nearest or farthest from each other. In this way clusterID helps to find the location between two clusters. But the performance of the algorithm totally depends on the type of dataset used in specific environment. Like Distance based environment is much comfortable to deal with numeric type values but Grid based environment is not much comfortable. So, from Figure1 the performance of the DRID in Distance based environment is greater than that of Grid based environment from Figure2 because of Grid based environment shows that each point of array is located at the same clusterID where as distance based environment shows different location (i.e-ClusterID). In case of Grid based environment clusters are described by intervals along attributes axes and corresponding centroids and histograms as in input form. This same algorithm DRID can also be applied in grid based environment Just difference in input applied by the user as per the requirement of the algorithm.

## CONCLUSIONS

We propose a “DRID – A new algorithmic approach” that merge two different-different environments, Distanced based environment and Grid based environment that results by merging of two clusters, which are placed close to each other increase the performance, reduce noise problem and reduce the execution time. A common strategy is to be followed in two different-different environments. Distance based algorithm(K-means algorithm) is

suitable for optimization type of problem and Grid based algorithm(Orthogonal Partitioning clustering ) algorithm is suitable for the set of cutting of hyper planes problems.

## FUTURE WORK

In future it can be extended by providing some additional changes in this approach so that “DRID” automatically detects the type of data set available and on the basis of data set, it can automatically choose the type of environment for merging.

## REFERENCES

[1] Manoranjan dash, Huan liu, xiaowei Xu. Merging distance and density based clustering.

[2]Ping YU, Data mining in library Reader Management.2011 International Conference on Network Computing and Information Security.

[3] Dr. J. Akilandeswari, A survey of partitioning clustering algorithms. International Journal of Enterprise Computing and Business Systems.

[4] Jiann-Cherng Shieh, Yung-Shun Lin. Bibliomining User Behaviors in the Library. Journal of Educational Media & Library Sciences.2006.

[5] Hsiao-Tieh Pu. Explore improving the utilization of library resources by bibliomining. Journal of Library Association of the Republic of China.2006.

[6] ZhaoHui Tang, Jamie MacLennan.Data Mining with SQL Server 2005.Wiley Publishing Inc, 2005.

[7]P. S. Bradley, U. Fayyad, and C. Reina. Scaling clustering algorithms to large databases. In Proceedings of the 4th

International Conference on Knowledge Discovery & Data Mining (KDD'2002).

[8] Nicholson, S. The Bibliomining Process: Data Warehousing and Data Mining for Library Decision-Making. Information Technology and Libraries. 2003.

[9] Seth Paul, Jamie MacLennan, Zhaohui Tang. Data Mining Tutorial. Microsoft Corporation.2005.

[10]Jiawei Han andMichelin Kamber. Data Mining: Concepts and Techniques.Morgan Kaufmann Publishers, 2000.

[11] M Steinbach, G Karypis, and V. Kumar. A comparison of document clustering techniques. In KDD Workshop on Text Mining, 2000.

[12] H. Abolhassani, M.Madhvi, 2009. Harmony K-means Algorithm for document clustering. Data Mining Knowledge Discovery.18:370-391.

[13] P. Berkhin, 2002 Survey of Clustering Data Mining Techniques.Technical Report , Accure Software , San Jose, Caiff.

## ACKNOWLEDGEMENT

There are a bunch of people to thank for this paper, including Mr. Iqbal Singh. This paper would not exist but for their faith in me, and I offer them my heartfelt thanks.