Abstract

Clustering is one of the most important research areas in the field of data mining. In simple words, clustering is a division of data into different groups. Data are grouped into clusters in such a way that data of the same group are similar and those in other groups are dissimilar. It aims to minimize intra-class similarity while to maximize interclass dissimilarity. Clustering is an unsupervised learning technique. Clustering is useful to obtain interesting patterns and structures from a large set of data. Clustering can be applied in many areas, such as marketing studies, DNA analysis, city planning, text mining, and web documents classification. Large datasets with many attributes make the task of clustering complex. Many methods have been developed to deal with these problems. In this paper, two well known partitioning based methods – k-means and k-medoids are compared over health data. This paper also proposes an improved k-means medoids clustering algorithm. The proposed algorithm is evaluated using the health dataset i.e Liver dataset and compare the results with other previous algorithms. The proposed algorithm is more effective in terms of computation time as compared to K means and K-medoids clustering algorithm. The algorithms under consideration, is evaluated with Rand Index, Jaccard Coefficient, Folkes and Mallows and Run Time as four metrics.
Evolving Efficient Clustering Patterns in Liver Patient Data through Data Mining Techniques

Experimental results are obtained on WEKA, a data mining tool.

References

- Raymond, T. Ng and Jiawei Han 2002, "CLARANS: A Method for Clustering Obejcts for Spatial Data Mining", IEE Transactions on Knowledge and Data Engineering, Vol. 14, No. 5, pp. 1003–1016.
Comparitive study &quot; National Conference on recent trends in Engineering & Technology.

Index Terms

Computer Science Algorithms

Keywords

Rand index (RI) Jaccard Coefficient Folkes and Mallows (FM) index Silhouette Index