Abstract

Partitioning clustering is generally performed using K-modes cluster algorithms, which work well for large datasets. A K-modes technique involve random chosen initial cluster centre (modes) as seed, which lead toward that problem clustering results be regularly reliant on the choice initial cluster centre and non-repeatable cluster structure may be obtain. K-Modes technique has been widely applied to categorical data a clustering in replace means through modes. The pervious algorithms select the attributes on frequency basis but not provided better result. Proposed algorithm select attributes on information gain basis which provide better result. Experimental results showing the proposed technique provided better accuracy.

References

2. Joel Luis Carbonera, Mara Abel, An entropy-based subspace clustering algorithm for
categorical data, 2014 IEEE 26th International Conference on Tools with Artificial Intelligence.
4. Zhexue Huang, A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining.
9. Guo Tao, Ding Xingu, Li Yefeng, Parallel k-modes Algorithm based on MapReduce.

Index Terms

Computer Science  Artificial Intelligence

Keywords

Clustering, Categorical data, K-mean algorithm, K-modes algorithm, Text mining