Abstract

Classification of categorical data always involves more complexities compared to the numerical data. Because, a firm outline cannot be drawn in case of categorical data. Different types of assumptions are followed by various researchers to treat such kind of data. Again, dissimilarity measures applied in case of numerical data cannot be applied directly in this case. In this paper, a new clustering algorithm for categorical data is proposed. The algorithm is using a newly devised dissimilarity measure. This paper only includes the theoretical description of the proposed algorithm with appropriate example.

References

2. Z. Huang Extensions to the k-means algorithm for clustering large data sets with
CATCLUS – A Proposed Algorithm for Clustering Categorical Data

categorical values Data Mining and Knowledge Discovery, 2 (3) (1998), pp. 283–304


Index Terms

Computer Science

Artificial Intelligence

Keywords

Categorical Data, Clustering, Dissimilarity Measure, Algorithm.