Abstract

Most of the earlier work on clustering has mainly been focused on numerical data whose inherent geometric properties can be exploited to naturally define distance functions between data points. Working only on numeric values prohibits it from being used to cluster real-world data containing categorical values. Recently, the problem of clustering categorical data has started drawing interest. The k-means algorithm is well known for its efficiency in this respect. It is also well known for its efficiency in clustering large data sets. However, in this paper we use the k-means algorithm to categorical domains by assigning rank value to the attributes.

References

Clustering of Categorical Data by Assigning Rank through Statistical Approach

- Alan Agresti 2nd edition 2007” An Introduction to Categorical Data Analysis”; John Wiley & Sons, Inc.,

Index Terms

Computer Science Data Mining

Keywords

Categorical Data K-mean Rank Value