Abstract

Minimum spanning tree-based clustering algorithm is capable of detecting clusters with irregular boundaries. Detecting outliers using clustering algorithm is a big desire. Outlier detection is an extremely important task in a wide variety of application. In this paper we propose a minimum spanning tree-based clustering algorithm for detecting outliers. The
algorithm partition the dataset into optimal number of clusters. Outliers are detected in the clusters based on outlyingness factor of each point (objects) in the cluster. The algorithm uses a new cluster validation criterion based on the geometric property of data partition of the data set in order to find the proper number of clusters. The algorithm works in two phases. The first phase of the algorithm creates optimal number of clusters, whereas the second phase of the algorithm detects outliers.

Reference

- D. Agrawal, and C. Pizzuti, Outlier Mining in Large High-Dimensional Data sets, IEEE Transactions on Knowledge and Data Engineering, 17(2): 203-215, 2005.
Outlier Removal Clustering through Minimum Spanning Tree

Outlier Removal Clustering through Minimum Spanning Tree

- Stefan Wuchty and Peter F. Stadler. “Centers of Complex Networks”. 2006

Index Terms

Computer Science
Pattern Recognition

Key words
Euclidean minimum spanning tree
Eccentricity
Cluster validity
Cluster Separation
Outlyingness factor
Outliers