Discovering Local Outliers using Dynamic Minimum Spanning Tree with Self-Detection of Best Number of Clusters

Abstract

Detecting outliers in database (as unusual objects) using Clustering and Distance-based approach is a big desire. Minimum spanning tree based clustering algorithm is capable of detecting clusters with irregular boundaries. In this paper we propose a new algorithm to detect outliers based on minimum spanning tree clustering and distance-based approach. Outlier detection is an extremely important task in a wide variety of application. The algorithm partition the dataset into optimal number of clusters. Small clusters are then determined and considered as outliers. The rest of the outliers (if any) are then detected in the clusters using Distance-based method. The algorithm uses a new cluster validation criterion based on the geometric property of data partition of the dataset 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, where as the second phase of the algorithm detect outliers in the clusters. The key
Discovering Local Outliers using Dynamic Minimum Spanning Tree with Self-Detection of Best Number of Clusters

A notable feature of our approach is its capacity to combine the best features of Distance-based and Clustering-based outlier detection to find noise-free/error-free clusters for a given dataset without the necessity of any input parameters.

Reference

- F. Angiulli, and C. Pizzuti, “Outlier Mining in Large High-Dimensional Data sets”, IEEE Transactions on Knowledge and Data Engineering, 17(2): 203-215, 2005
Discovering Local Outliers using Dynamic Minimum Spanning Tree with Self-Detection of Best Number of Clusters


Discovering Local Outliers using Dynamic Minimum Spanning Tree with Self-Detection of Best Number of Clusters

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

**Index Terms**

Computer Science

Information Retrieval

**Key words**

Euclidean minimum spanning tree

Subtree

Clustering

Eccentricity
Cluster validity

Cluster Separation

Small Clusters

Outliers