Structural similarity micro clustering algorithm for local outliers and hubs using dynamic minimum spanning tree

S. John Peter*

Department of Computer Science and Research Center
St. Xavier's College, Palayamkottai
Tamil Nadu
India

Abstract

Clustering is a process of discovering groups of objects such that the objects of the same group are similar, and objects belonging to different groups are dissimilar. Minimum spanning tree based clustering algorithm is capable of detecting clusters with irregular boundaries. Many algorithms find clusters by maximizing the number of intra-cluster edges. While such algorithms find useful and interesting structures, they tend to fail to identify and isolate two kinds of vertices that play special roles – vertices that bridge clusters (hubs) and vertices that are marginally connected to clusters (outliers). Identifying hubs is useful for applications such as viral marketing and epidemiology since hubs are responsible for spreading ideas and disease. In this paper we propose a novel algorithm called Structural Similarity Micro Clustering Algorithm for Local outliers and hubs using Dynamic Minimum Spanning Tree (**SSMCALOHDMST**), which detect clusters, outliers and hubs in data set. The algorithm partition the dataset into optimal number of clusters. 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, where as the second phase of the algorithm detect outliers and hubs.

Index Terms: *Euclidean minimum spanning tree, Subtree, Clustering, Eccentricity, Center, Cluster validity, Cluster Separation, Outliers, Hubs*