

Automatic Extraction of Clusters from Hierarchical Clustering Representations

Jörg Sander, Xuejie Qin, Zhiyong Lu, Nan Niu, Alex Kovarsky
Department of Computing Science
University of Alberta
Edmonton, AB, Canada T6G 2E8
{ joerg, xuq, zhiyong, nan , kovarsky }@cs.ualberta.ca

Abstract

Hierarchical clustering algorithms are typically more effective in detecting the true clustering structure of a data set than a partitioning algorithm. However, hierarchical clustering algorithms do not actually create clusters, but compute only a hierarchical representation of the data set. This makes them unsuitable as an automatic preprocessing step for other algorithms that operate on detected clusters. This is true for both dendrograms and reachability plots, which have been proposed as hierarchical clustering representations, and which have different advantages and disadvantages. In this paper we introduce a method that automatically determines the significant clusters found by a hierarchical clustering method from a reachability plot. To be able to extract the clusters also from dendrograms, we also show that dendrograms and reachability plots contain essentially the same information and that one can easily convert them into each other. This makes it possible to use hierarchical clustering as an automatic preprocessing step that requires no user interaction to manually select cluster from a hierarchical clustering representation.