Hierarchically Built Trees with Probability of Placing Clusters

Nebahat Bozkus[1], Stuart Barber[2]

[1]) University of Giresun, Giresun, Turkey
nebahat.bozkus@giresun.edu.tr
[2]) University of Leeds, Leeds, UK

One of the popular questions in hierarchical clustering is how many clusters we have. The available cluster validity indices (CVIs) capture the high percentage of well-separated true components, but their performances deteriorate for indistinct groups or unique data structures. In addition, the available CVIs, or at least the ones we are aware of, find the number of clusters, but they do not show where each cluster needs to be placed on trees. We propose a new algorithm based on non-decimated lifting (NLT) using the 'denoising' abilities of wavelet methods.

In our algorithm, we denoise departures from the centroid of each possible cluster (each node) on a tree by the help of lifting 'one coefficient at a time' (LOCAAT) algorithm. To allocate a cluster at a node, we use denoised departures from the centroids for each possible cluster on the tree. If the denoised result of a possible cluster is small enough, we allocate one of clusters at the node of interest on the tree. However, the nature of the NLT algorithm includes some number of repeated LOCAAT algorithm. Thus, denoising the tree by the NLT algorithm finds a clustering pattern for each repetition. We suggest that we can summarize the multi-denoising results by allocating a probability of being a cluster at each node on the tree, and we can place a cluster at a node where the probability of placing a cluster at this node and all its child nodes are greater than a user defined threshold. We compare the performance of our algorithm with some available methods in the literature using some simulated and real data sets.

Keywords: Cluster Validity Indices; Clustering; Lifting; Wavelets