R Package - hclustHDCP

Description

This package deals with one of the classical problem in statistics, called change-point detection method. It attempts to detect abrupt significant changes in probability distributions for a sequence of observations arranged in chronological order. We implement change-point detection for high dimensional observations. This method detects change-points based on hierarchical clustering. The algorithm initially considers each observation as a cluster and at each subsequent step merges two closest consecutive clusters based on linkages (single, average or complete). Hence the number of clusters decreases by one in each step of the algorithm. In case the number of change-points to be detected is known, the algorithm stops when the desired number of clusters is reached. It returns the change-point locations where the cluster membership changes. However, if the number of change-points to detect is unknown, then we make use of some clustering evaluation algorithm like Penalized Dunn index to estimate optimal number of change-points. Here we also implement the algorithm using a different distance function, called the mean absolute deviation of distances (MADD distance) instead of usual Euclidean distance while calculating the linkage methods. This enables the method to perform well in low sample size situations as well for most of the occasions.

Installation Instruction

devtools::install_github("DawnTrisha/hclustHDCP")

Remaining work

The following works need to be performed :

1. Complete coding DetectChangePoint_NumUnknown.R for detecting change-point when the number of change-points to detect is *unknown*
2. Code calculation of MADD distance and generalized MADD distance to be used in the algorithm
3. Make vignette
4. Complete documentation
5. Time permitting add other cluster evaluation method for comparison of performance