rucrdtw: Fast time series subsequence search in R

Philipp H Boersch-Supan

1 Department of Geography and Emerging Pathogens Institute, University of Florida

Summary

Dynamic Time Warping (DTW) methods provide algorithms to optimally map a given time series onto all or part of another time series (Berndt and Clifford 1994). The remaining cumulative distance between the series after the alignment is a useful distance metric in time series data mining applications for tasks such as classification, clustering, and anomaly detection.

Calculating a DTW alignment is computationally relatively expensive, and as a consequence DTW is often a bottleneck in time series data mining applications. The UCR Suite (Rakthanmanon et al. 2012) provides a highly optimized algorithm for best-match subsequence searches that avoids unnecessary distance computations and thereby enables fast DTW and Euclidean Distance queries even in data sets containing trillions of observations.

A broad suite of DTW algorithms is implemented in R in the dtw package (Giorgino 2009). The rucrdtw R package provides complementary functionality for fast similarity searches by providing R bindings for the UCR Suite via Rcpp (Eddelbuettel and Francois 2011). In addition to queries and data stored in text files, rucrdtw also implements methods for queries and/or data that are held in memory as R objects, as well as a method to do fast similarity searches against reference libraries of time series.

References


Figure 1: UCR DTW is approximately 3 orders of magnitude faster than a naive sliding-window search using DTW distance.