rucrdtw: Fast time series subsequence search in R

Philipp H Boersch-Supan[1]

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 alignement 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

Berndt, Donald J, and James Clifford. 1994. "Using Dynamic Time Warping to Find Patterns in Time Series." In *KDD Workshop*, 10:359–70. 16. AAAI. http://www.aaai.org/Library/Workshops/1994/ws94-03-031.php.

Eddelbuettel, Dirk, and Romain Francois. 2011. "Rcpp: Seamless R and C++ Integration." *Journal of Statistical Software* 40 (1): 1–18. doi:10.18637/jss.v040.i08.

Giorgino, Toni. 2009. "Computing and Visualizing Dynamic Time Warping Alignments in R: The Dtw Package." *Journal of Statistical Software* 31 (7): 1–24. doi:10.18637/jss.v031.i07.

Rakthanmanon, Thanawin, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2012. "Searching and Mining Trillions of Time Series Subsequences Under Dynamic Time Warping." In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 262–70. ACM. doi:10.1145/2339530.2339576.
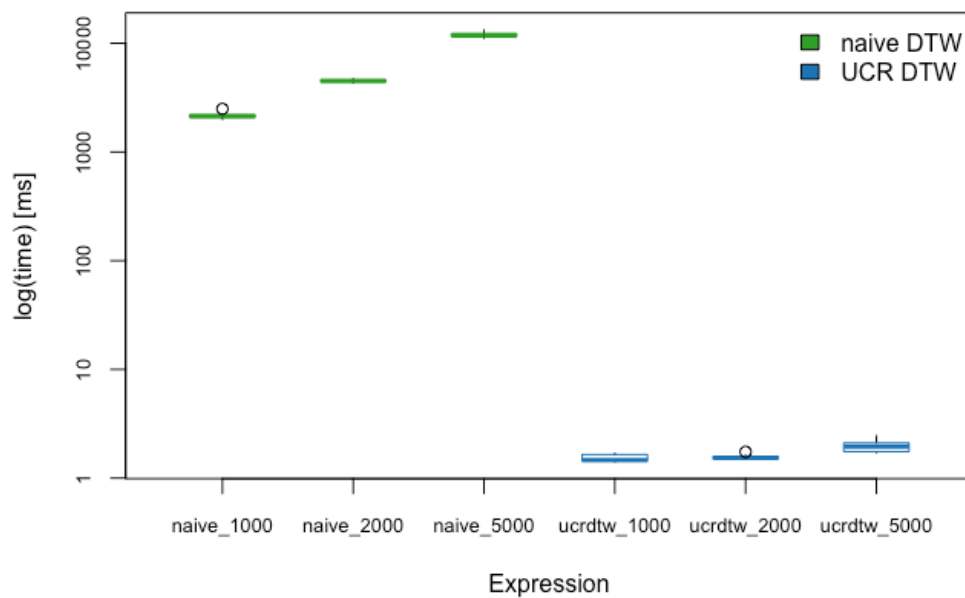
Figure 1: Figure 1: UCR DTW is approximately 3 orders of magnitude faster than a naive sliding-window search using DTW distance.