On cluster stability of
k-NN early classifiable time series

Istvan DAVID[1] and Krisztian BUZA[2]

[1] Department of Computer Science and Information Theory
   Budapest University of Technology and Economics
   `david@cs.bme.hu`
[2] Department of Computer Science and Information Theory
   Budapest University of Technology and Economics
   `buza@cs.bme.hu`

Abstract. Early classifiability is a useful property of time series. The related data mining techniques shorten the time required for classification or pattern-recognition. Some practical examples can be found in the domain of health care or network engineering. (E.g. classification of TCP/IP packets based on the first few partial segments of data.)

On the other hand, this phenomenon raises the question whether the clusters formed over an early classifiable time series database can be early predicted as well, i.e. are these clusters stable or not over time? This particular question has not been examined in the literature yet.

In our paper we define some new concepts to establish the common taxonomy of the problem and we present our latest results. Based on past publications, we use the k-nearest neighbor algorithm for early classification and k-medoids for clustering.

References

XING, Z., PEI, J. and YU, P.S. (2009): Early Prediction on Time Series: A Nearest Neighbor Approach. In: *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*. AAAI Press, Palo Alto, California, 1297–1302.

ACKERMAN, M., BEN-DAVID, S. (2009): Clusterability: A Theoretical Study. *Journal of Machine Learning Research: Workshop and Conference Proceedings, 5, 1-8.*

XING, Z., PEI, J., DONG G. and YU, P.S. (2008): Mining sequence classifiers for early prediction. *SDM08 Proceedings of the 2008 SIAM international conference on data mining, 644-655.*

Keywords