

CS 780 - Homework 1
Due 10/10

Read 3 papers, one from each of the categories below. For each paper, write a one-page report that consists of:

1 paragraph: summarizing the paper, in your own words (do not just copy the abstract or the conclusion).
1 paragraph: **your** own thoughts on the paper, e.g. advantages/disadvantages/limitations of the algorithm(s) proposed, ways that you may improve them, etc.

Required:

Must-read for anyone who wants to do research in time series

Keogh, E. and Kasetty, S. (2002). [On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration](#). In the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. July 23 - 26, 2002. Edmonton, Alberta, Canada. pp 102-111.

Pick one from this list (algorithms we covered in class):

SAX – extended version (original paper published in 2003)

Jessica Lin, Eamonn J. Keogh, Li Wei, Stefano Lonardi: [Experiencing SAX: a novel symbolic representation of time series](#). Data Min. Knowl. Discov. 15(2): 107-144 (2007)

VizTree: Visualization tool for motif discovery and anomaly detection

Jessica Lin, Eamonn J. Keogh, Stefano Lonardi, Jeffrey P. Lankford, Donna M. Nystrom: [Visually mining and monitoring massive time series](#). KDD 2004: 460-469

R-K Band: Learning the best warping band for DTW

Ratanamahatana, C. A. and Keogh, E. (2004). [Making Time-series Classification More Accurate Using Learned Constraints](#). In proceedings of SIAM International Conference on Data Mining (SDM '04), Lake Buena Vista, Florida, April 22-24, 2004. pp. 11-22.

Compression-based structural similarity (based on co-compressibility)

Keogh, E., Lonardi, S. and Ratanamahatana, C. (2004). [Towards Parameter-Free Data Mining](#). In proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, WA, Aug 22-25, 2004.

Subsequence clustering

Eamonn J. Keogh, Jessica Lin: [Clustering of time-series subsequences is meaningless: implications for previous and future research](#). Knowl. Inf. Syst. 8(2): 154-177 (2005)

Time series discord discovery – the original paper

E. Keogh, J. Lin and A. Fu (2005). [Efficiently Finding the Most Unusual Time Series Subsequence](#). In Proc. of the 5th IEEE International Conference on Data Mining (ICDM 2005), pp. 226 - 233. Houston, Texas, Nov 27-30, 2005.

iSAX

Jin Shieh and Eamonn Keogh (2008) [iSAX: Indexing and Mining Terabyte Sized Time Series](#). SIGKDD 2008.

Grammar-based motif discovery (Sequitur)

Yuan Li, Jessica Lin, and Tim Oates. 2012. [Visualizing variable-length time series motifs](#). In Proceedings of the 2012 SIAM International Conference on Data Mining. Anaheim, CA. Apr 26-28. Pages 895-906.

Pick another one from this list (Nice algorithms that we didn't have time to cover in class)

A good motif discovery algorithm

Abdullah Mueen, Eamonn Keogh, Qiang Zhu, Sydney Cash, Brandon Westover (2009). [Exact Discovery of Time Series Motifs](#). SDM 2009.

Shapelets are subsequences that are the most representatives of a class. This is similar to building a decision tree for time series. A nice paper on time series classification.

Thanawin Rakthanmanon and Eamonn Keogh. [Fast-Shapelets: A Scalable Algorithm for Discovering Time Series Shapelets](#). SDM 2013

A "fix" for subsequence clustering

Thanawin Rakthanmanon, Eamonn Keogh, Stefano Lonardi, and Scott Evans (2011). [Time Series Epenthesis: Clustering Time Series Streams Requires Ignoring Some Data](#). ICDM 2011.