Shapelet-neural-networks for weakly supervised problems

PhD position at IRISA, Rennes

Simon Malinowski - smalinow@irisa.fr

In the time series analysis domain, very efficient methods have been developed recently for supervised tasks (e.g. classification). Amongst them, shapelet-based models are known to be efficient both in terms of accuracy and complexity [1, 2, 3].

However, in a wide range of applications, very little amount of supervised information is available which prevents from using the above-cited methods directly.

Recently, some efforts have been dedicated to design unsupervised methods for time series analysis [4, 5, 6, 7]. These works mainly focus on the particular task of time series clustering.

The aim of this thesis is to explore several weakly supervised tasks for time series analysis. For that purpose, we will be particularly interested in bridging the gap between shapelets and neural networks, in order to learn efficient representations for time series in a weakly supervised context.

In [6], we designed LDPS, a model combining shapelet and siamese networks in order to embed time series in a space where Euclidean distance mimics a widely used similarity measure for time series analysis (DTW). We aim at extending this framework to the following tasks :

- time series indexing under DTW : this task is known to be very challenging [8]. We expect that an anytime extension of the LDPS framework would be of great help for this task.
- metric learning and semi-supervised learning : we will be interested in extending the LDPS framework for situations where only a few labels are available (semi-supervised task) or where supervised information is available as must-link/cannot-link constraints (as in the metric learning framework)

Other tasks (eg. domain adaptation) will also be considered.

The methods developed in this thesis will be applied to multimedia data in the context of efficient document retrieval.

[1] Jon Hills, Jason Lines, Edgaras Baranauskas, James Mapp, and Anthony Bagnall. Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery*, 28(4) :851–881, 2014.

[2] Josif Grabocka, Nicolas Schilling, Martin Wistuba, and Lars Schmidt-Thieme. Learning time-series shapelets. pages 392–401, 2014.

[3] Lexiang Ye and Eamonn Keogh. Time series shapelets : a new primitive for data mining. pages 947–956, 2009.

[4] Jesin Zakaria, Abdullah Mueen, and Eamonn Keogh. Clustering time series using unsupervised-shapelets. pages 785–794, 2012.

[5] Qin Zhang, Jia Wu, Hong Yang, Yingjie Tian, and Chengqi Zhang. Unsupervised feature learning from time series. 2016.

[6] Arnaud Lods, Simon Malinowski, Romain Tavenard, and Laurent Amsaleg. Learning DTW-Preserving Shapelets. In *International Symposium on Intelligent Data Analysis*, pages 198–209, October 2017.

[7] Marco Cuturi and Mathieu Blondel. Soft-DTW : a differentiable loss function for time-series. pages 894–903. PMLR, 8 2017.

[8] Eamonn Keogh and Chotirat A. Ratanamahatana. Exact Indexing of Dynamic Time Warping. 7, 2005.