Pattern-based Time Series Segmentation

L. Carpentier[1,3], L. Feremans[2], W. Meert[1], M. Verbeke[1,3]

[1] Departement of Computer Science, KU Leuven

[2] Departement of Computer Science, University of Antwerp

[3] Flanders Make@KU Leuven

e-mail: `louis.carpentier@kuleuven.be`

Keywords: Semantic segmentation, Multivariate time series, Pattern mining

In recent years, the wide availability of low-cost, high resolution sensors has led to a dramatic increase in monitoring capabilities. Across a wide range of application areas, increasingly long sequences of measurements (i.e., time series) are being monitored. The monitored process is typically composed of multiple states. *Time series semantic segmentation* aims at automatically uncovering these hidden states from the time series data in an unsupervised manner [3, 4].

State transitions often do not lead to abrupt changes, but happen gradually. For example, in motion capture analysis, there is a gradual transition from the state *sitting* to the state *standing*, namely *standing up* [2]. State-of-the-art time series semantic segmentation algorithms have only focused on identifying discrete state transitions through change point detection [3, 4]. As noted in [3], detection of gradual changes is "outside the scope of *current* time series segmentation methods".

We propose Pattern-based Time Series Segmentation (PaTSS), a novel, domain-agnostic semantic segmentation algorithm that can learn gradual state transitions in time series data. PaTSS performs a semantic segmentation for gradual state transitions based on an embedding space derived from mined sequential patterns. It achieves this by performing the following steps:

1. Segment the time series using multi-resolution sliding windows, and transform each segment into a symbolic representation. By considering multiple resolutions, PaTSS captures both long and short term behavior.

2. Mine frequent sequential patterns in the symbolic representations, thus learning the frequent shapes and behavior of the time series. For this step, PaTSS leverages the well-established field of frequent pattern mining [1].

3. Embed the time domain using the mined frequent patterns. For each pattern and each time unit, the embedding value is set to the relative support of that pattern if it covers the time unit, and to zero otherwise.

4. Identify the semantic segments which have a similar embedding (and consequently similar behavior) and learn the likelihood of each semantic segment occurring at a certain time.

By learning distributions over the different semantic segments, PaTSS can identify gradual state transitions, namely when the likelihood of some segment decreases while the probability of another one increases.

PaTSS has two major advantages, besides being able to identify gradual state transitions. First, PaTSS can identify reoccurring behavior in the time series, because the goal is to group similar behavior. This is in contrast to state-of-the-art semantic segmentation procedures, which separate dissimilar behavior through change point detection. Second, PaTSS uses frequent patterns to embed the time series. This ensures that the decision making process is highly explainable because every decision can be linked to a small set of easy-to-interpret patterns.

References

[1] C. C. Aggarwal et al. *Data mining: the textbook*. Vol. 1. Springer, 2015. DOI: 10.1007/978-3-319-07821-2.

[2] F. Chamroukhi et al. "Joint segmentation of multivariate time series with hidden process regression for human activity recognition". In: *Neurocomputing* 120 (2013), pp. 633–644. DOI: 10.1016/j.neucom.2013.04.003.

[3] A. Ermshaus, P. Schäfer, and U. Leser. "ClaSP: parameter-free time series segmentation". In: *Data Mining and Knowledge Discovery* (2023), pp. 1–39. DOI: 10.1007/s10618-023-00923-x.

[4] S. Gharghabi et al. "Domain agnostic online semantic segmentation for multi-dimensional time series". In: *Data mining and knowledge discovery* 33 (2019), pp. 96–130. DOI: 10.1007/s10618-018-0589-3.