Novel transformer-based architectures for time series classification based on symbolic data representation
ABSTRACT FOR THE GENERAL PUBLIC

Subject matter

Classification, in general, is the process of predicting the class label of an object. We describe an object using measurable attributes, also called features. The same set of features is extracted for each object, and a decision-making algorithm, called a classifier, is used to distinguish between objects of different classes. In the case of time series classification, attributes are consecutive observations ordered in time. A prominent group of methods in time series classification are dictionary approaches. They convert raw time series values into symbols from a predefined alphabet. The symbolic data representation is then used as a basis for time series classification.

In this project, we plan to provide novel algorithms and data processing pipelines for time series classification using symbolic data representation. The focus will be placed on methods that are able to handle challenging data sets. The project will result in the development of new neural network-based models. They will utilize an architecture termed *transformer*. This architecture is at the moment in the centre of all modern natural language processing tasks. The plain transformer model is suitable for sequence to sequence learning. We will need to use it in a predictive task, which will require appropriate adaptations.

The motivation for pursuing this research path stems from the observed inadequacies in how the existing time series classifiers handle challenging data sets and the lack of a systematic approach to neural approaches based on symbolic time series representations. We plan to utilize the methodological advances of neural architectures known from the natural language processing domain and transfer these concepts to the field of time series classification.

Motivation and originality

We plan to deliver two types of transformer-based architectures: encoder-decoder and and encoder-only. To the best of our knowledge, our studies will be the first so far that will directly target the development of transformer-based methods for time series data represented in the form of a sequence of symbols. Successful completion of these tasks will substantially enlarge the selection of available methods in multiple real-world domains where time series need to be classified.

This work is on the borderline of natural language processing and time series classification. Let us emphasize the huge methodological development observed in the past few years in the natural language processing domain. In that regard, we need to mention the recent decoder-only neural network model (a variant of the transformer architecture). The decoder-only architecture executes sequence-to-sequence data processing tasks (it is thus unsuitable for direct class label generation). Decoder-only networks became the cornerstone of the generative pre-trained transformers (GPTs). With this, we want to underline that the project will deal with the most advanced neural architectures used at this moment in natural language processing. We will adapt the methodological advancements from this domain to the domain of time series classification.