Background

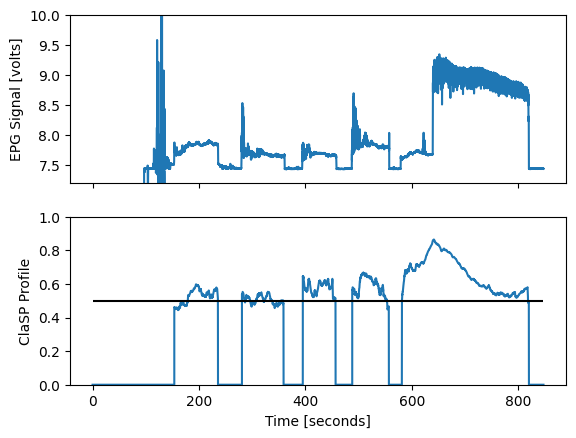
Classification Score Profile (ClaSP) from Ermshaus et al. (2023) is a method for unsupervised segmentation of time-series data that displays periodicity within each segment. ClaSP does this by dividing the time series into windows, normalizing them, and then for each window evaluating the performance of a k-nearest neighbors classifier trained to distinguish the windows before and after each given window. The profile score at each time point is the AUC score of the model distinguishing before and after that time point. As the profile is the AUC score of the k-nearest neighbors models, it is (mostly) bounded between 0.5 (random guessing) and 1 (perfect discrimination) when there is a probe. Profile scores below 0.5 are however possible if the models are guessing *worse* than randomly. ClaSP then goes on to find peaks in the profile and use these to identify segmentation points based on a heuristic. In our case, we are simply interested in the profile score as a channel to be used in our other models.

Implementation Details

We make use of the ClaSPy library that implements the ClaSP algorithm. With the goal of improving segmentation performance and running time, we first split up each EPG recording into separate probes by detecting the periods of NP values between each probe. ClaSP is then run on each probe using a window size determined by SUSS (an algorithm for finding the period of self-similarity in time series provided by ClaSPy) and no ensembling. This produces a score profile which we then append to the original dataframe containing the waveform data. Periods between probes are given a value of 0 in this column.

Example Output

A plot of an EPG recording paired with its CLaSP profile is shown below. Notice that the profile is 0 in regions where there is not a probe taking place and mostly above 0.5 otherwise. Interpreting this plot, we see that the fifth probe has a higher profile than the others. While there could be many reasons for this, it could perhaps be because it has a much stronger signal than the others, especially in the latter three quarters of its duration.



Random Forests (RF) Model

Using our previous RF model with the profile score as a channel did not lead to a change in its accuracy of 0.85. However, it did increase its precision from 0.51 to 0.59 and its recall from 0.39 to 0.40, leading to an increase in F1 from 0.40 to 0.41. While this is an improvement, we would not necessarily expect this new channel to greatly improve the performance of RF as it is a fairly simple model and the profile score is likely more useful in convolutional networks that can make use of peaks in the profile that are themselves associated with transitions.

TCN, etc…

Citations

Ermshaus, Arik, Patrick Schäfer, and Ulf Leser. 2023. Clasp: parameter-free time

series segmentation. Data Mining and Knowledge Discovery 37(3):1262–1300. doi:

10.1007/s10618-023-00923-x. URL http://dx.doi.org/10.1007/s10618-023-00923-x.