ATOL: Measure Vectorization for Automatic Topologically-Oriented Learning, Supplementary Materials

1 Topological score for time series, an industrial application

We present an industrial application for time series, in a case where the learning problem is hard and no obvious solutions are to be found. This dataset consists in the following experiments: using commercially available simulator of a Japanese city road circuit course, about a hundred subjects are monitored and the intervals between successive heartbeats are recorded (RRI data sampled at 4Hz) for a 80 minutes drive that includes two periods of high-speed driving at the beginning and at the end of the experiment, and a low-speed driving period in the middle designed to induce sleepiness. For each experiment, an expert annotation (labeled NEDO score) produced from visual observation of the driver is made available, indicating sleepiness on a 1 to 5 class scale. We show four such experiments in Figure 1 (the RR-intervals have then been normalised).

This problem of retrieving the sleepiness level based on RRI levels is hard and ill-posed: there are strong individual differences in perceived reaction to a given situation, a single experiment per subject to learn behaviour from, and apparent noise or absence of signal in annotations, see e.g. subject 3 in Figure 1. Nevertheless we propose to use the Atol framework to produce features meant reflect the sleepiness level in subjects based on RRI variations. The intent is that even though this will poorly reflect the latent sleepiness level, this could be enough to allow to catch jumps in the perceived attention level. The framework can readily be applied to time series in any given dimension and used to produce topological features. For this application we will follow a classical path: *(i)* use a sliding window decomposition on the RRI time-series, *(ii)* use a time-delay embedding to transform said window into a point cloud, *(iii)* apply persistent homology analysis (we use DTM-filtration [**?**]) to produce persistence diagrams and *(iv)* vectorise persistence diagrams using Algorithm 1.

We concatenate those features with the mean and standard deviation statistics on the sliding-window. As for learning, we compute a learner based on other individuals' features regressed to their NEDO scores, and use it to generate a score based on Atol features (see middle and bottom row in Figure 2). Although this score imperfectly reflects the underlying NEDO score for a given patient, is can still have some uses. We set to detect two jumps on this topologically-augmented score using a Gaussian Kernel. We also compute a regressor based on the standard features without additional topological features, for comparison purposes, and also detect two jumps on this standard score.

Figure 2 shows two example results of our analysis. Each panel (top and bottom) consists in three time-series: the (hidden) NEDO score (top row), the Atol-score computed from a regressor based on topological features (middle row), and a standard score computed from a regressor based solely on standard features (bottom row). The changes of colour from blue to red and to blue indicates the changes in the experimental design for the driving simulation, i.e. the red portion indicates low-speed driving whereas the blue portions indicate high-speed driving periods. The black dotted lines indicate jumps detected from the Atol representation, whereas the red dotted lines indicate jumps detected from the standard representation. In the top panel, the two series of jumps are concomitant, and almost an exact match to the underlying changes in the experimental design. In the bottom panel, an improvement over the standard score is caught with the Atol score that better reflects the changes in latent NEDO score for this subject, two the point that the detected jumps are an exact match for the changes in experimental conditions. Overall, the Atol score has less spikes and more regularity than the standard score, which is expected as the topological features are extracted posterior to a time-delay embedding procedure.

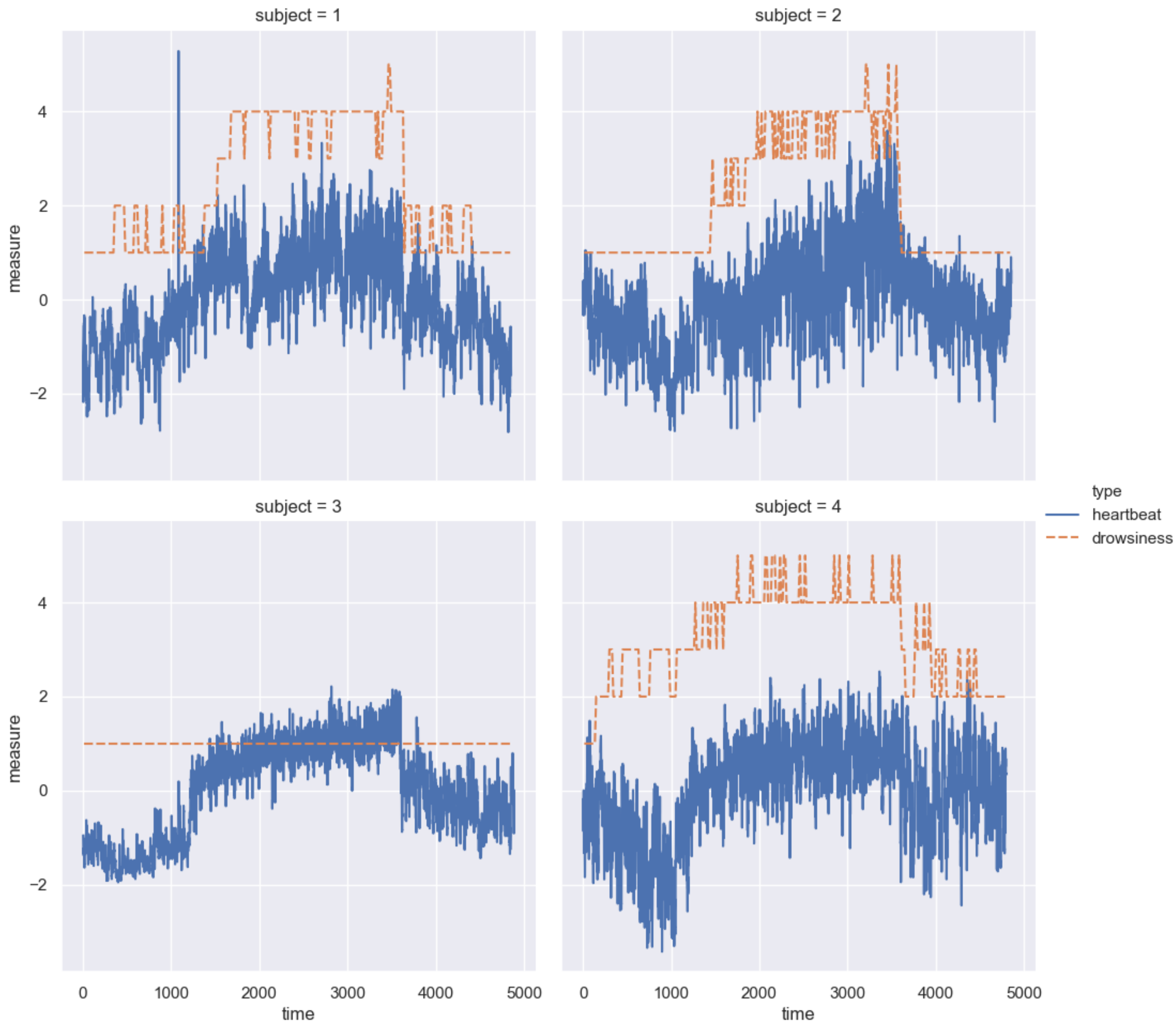

Figure 1: Normalized RRI time-series (blue) and annotated NEDO score (orange) for four subjects with the simulation time (x-axis, in seconds).

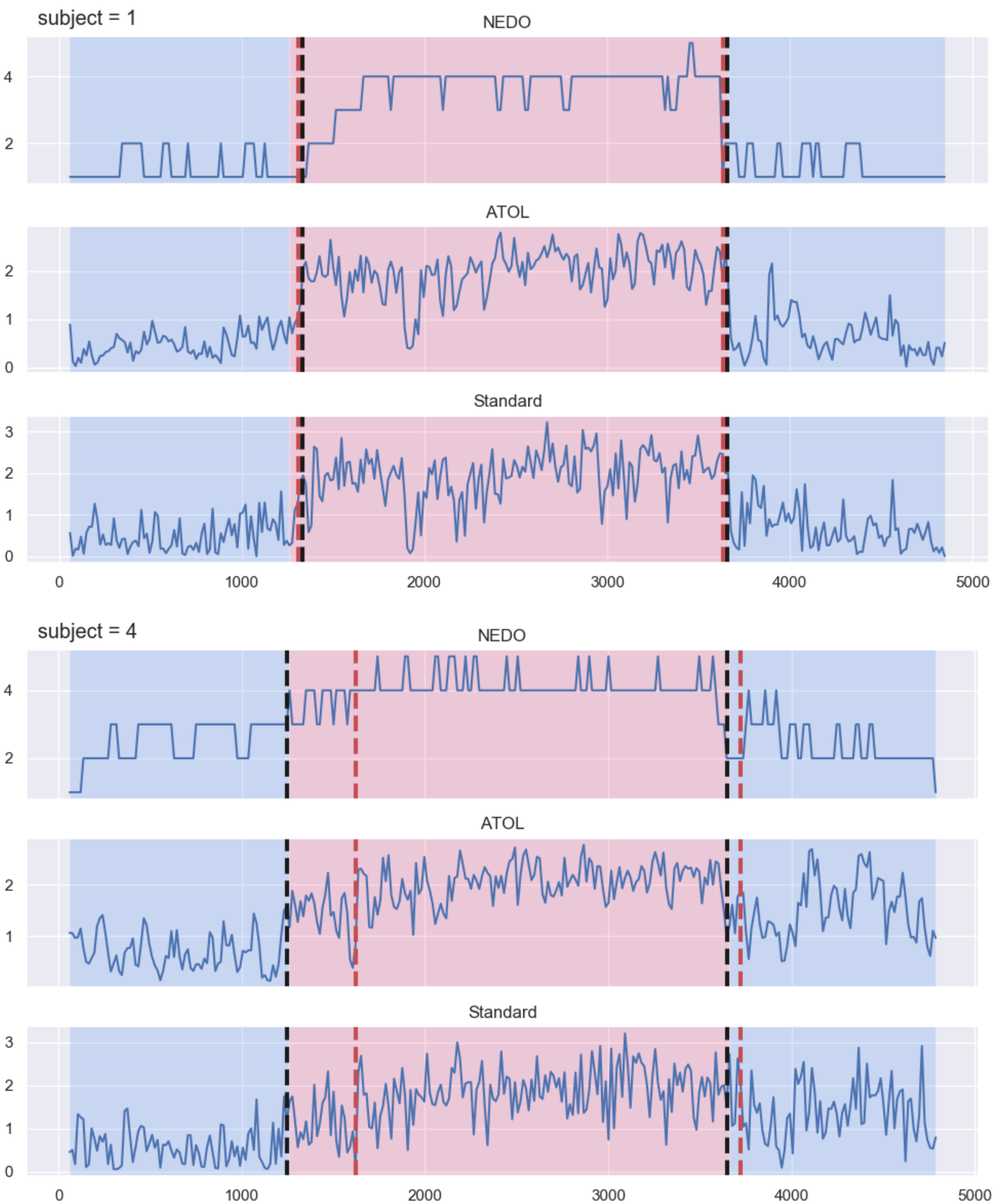

Figure 2: For two subjects (1 and 4), results of NEDO score (top row) regression and of a 2-jumps detection procedure, from Atol procedure (Atol-regression in middle row, yields jumps in black dashes) and standard features alone (standard-regression in bottom row, yields jumps in red dashes). Red zones indicates low-speed and blue zones indicates high-speed section in the experiment.