Automated Anomaly Detection in Large Sequences

Paul Boniol, Michele Linardi, Federico Roncallo, Themis Palpanas

Alexis Marouani
Grégoire Béchade

19 Décembre 2024

Introduction and context

Classical anomaly detection methods rely on comparing a subsequence to each other subsequence in the time-series. It raises problems when facing :

- Large time-series
- Repeated anomalies

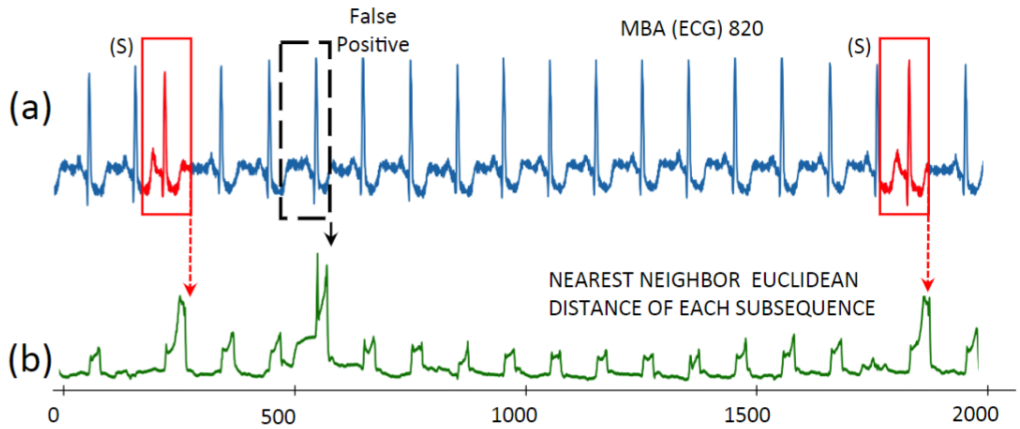

Figure – Heartbeats with several anomalies

Method

The article introduces a new method to perform anomaly detection in large time series, which relies on the introduction of a "normal behaviour" of the time series.

Construction of N_M

- Randomly select subsequences of length $3 \times l$
- Hierarchical clustering of the subsequences
- Select the cluster c that maximises
  $N(c) = \frac{frequency(c)^2 \times coverage(c)}{\Sigma_{x \in \mathbb{C}} dist(center(c), center(x))}$

Idea : The normal cluster is the one that is the most frequent and the most central.

Method

Outliers detection :

For each subsequence of size l in the time series :

- Compute the distances to all subsequences of size l in N_M.
- Label as anomalies the k sequences with the largest distance to M_N, or the one that are above a certain threshold.