

# Online Unsupervised State Recognition in Sensor Data (Supplementary Materials)

Julien Eberle, Tri Kurniawan Wijaya, and Karl Aberer

School of Computer and Communication Sciences

École Polytechnique Fédérale de Lausanne (EPFL)

CH-1015 Lausanne, Switzerland

Email: {julien.eberle, tri-kurniawan.wijaya, karl.aberer}@epfl.ch

One of the goals of Spclust and StateFinder is to produce a symbolic time series to support higher level applications, without converting it back to sensor's original measurement values. Since symbolic time series is much shorter than its original version, this property is desirable, especially due to the limited sensor's storage and computational power. Thus, we do not discuss the process of reverting symbols back to its original value in the main paper. However, for the sake of completeness, below we illustrate how one can revert symbolic time series to its original values.

## *A. Converting symbols level 0 generated by Spclust to its original values*

In this case, we could simply use the cluster centroids to approximate the original values of the symbols. Note that, using Spclust, we have one-to-one mapping between clusters and symbols.

## *B. Converting RLE compressed triples to non-compressed triples.*

Each symbol is repeated  $n$  times with  $n = (t_e - t_s)/r$ , where  $r$  is the sampling rate of the original data. As sensors may have a variable sampling rate or some gaps, this transformation could produce more/less triples than the original ones.

## *C. Converting symbols from level 1 or higher to one level lower.*

The main idea, is to use the Segment Transition Matrix To find the starting symbol, we use an heuristic that aims to find a symbol that has the lowest incoming transition probability and a positive outgoing transition probability. Formally, we take the symbol  $i$  such as the sum of the elements in the  $i$ th row, except the diagonal is greater than 0 and the sum of the elements on the  $i$ th column, except the diagonal, is minimal. Then according to the transition probability to the next symbol, we build a sequence. Even though we might produce a slightly different sequence from the original one, it eventually has the same symbol distribution.