Unsupervised Time Series Segmentation for High-Dimensional Body Sensor Network Data Streams

David Haber¹, Andreas A.C. Thomik² & A. Aldo Faisal¹,2,3
¹Dept. of Computing; ²Dept. of Bioengineering; ³MRC Clinical Sciences Centre, Hammersmith Hospital; Imperial College London, London SW7 2AZ
andreas.thomik10@imperial.ac.uk

1. Introduction
Locating behavioural sub-sequences in body sensor network data streams of high temporal and spatial resolution efficiently and accurately is an open challenge.

We present a novel unsupervised machine learning method for segmenting, extracting and recognising motion segments from long and complex time series in O(n).

We apply our method to a kinematic data set consisting of millions of samples in 18 dimensional space and demonstrate its ability to accurately segment and extract motion building blocks. Our method is additionally validated on an artificially-generated dataset.

2. Unsupervised Time Series Segmentation and Recognition in O(n)
The method presented here allows to process long and high-dimensional body sensor network data streams:

1. Reducing Dimensionality
Optimal projection onto low-dimensional space

2. Change Point Analysis
Computes the most likely length for the current segment at each time step. Adapted from [1]

3. Segmentation
Significant drops in run-length are marked as segmentation points and segments are extracted

4. Comparison of Segments
Movement segments are re-sampled and compared using Maximum Cross-Correlation [2]

3. Artificial Data Set
• Randomly generated sequences from the family of functions
  \[ f = \sum_{i=1}^{N} \phi(x_i) + \xi \]
  • Varied dimensionality, number of segment types, noise levels and method parameters
  • 55 parameter combinations, each run 20 times, 1100 runs in total
  • Achieved 86.5% segmentation accuracy on artificial data set, linear complexity in the number of data points

4. Neurotechnology for Kinematics
7 healthy subjects wearing CyberGlove I, 17 tasks from everyday life (eating, typing, opening doors, etc.), 18 dimensional data, 80 Hz, > 1.5M data points [3]

5. Detecting Recurrent Motion Building Blocks
Clusters of similar movement segments are generated using an iterative algorithm. Each cluster represents a basic action (building block) that can be composed into more complex sequences of movement.

• Recurrent building blocks (black in right figure) found across subjects
• Most frequently extracted actions from data set: various forms of hand openings/closings

6. Application to Neuroprosthetics
Our method enables a Big Data driven approach to improved control of neuroprosthetic devices [2].

The behaviour of the undamaged limb is observed to predict the most likely actions of lost limbs in real-time using a two-step process and the clusters previously generated.

• Statistically significantly higher prediction accuracy \((p < 0.05)\)
• Real-time prediction of missing limb movement

7. Discussion
Segmentation/Recognition
• Very high accuracy (86.5%) and extremely robust to most variations in the data and noise
• Recurrent patterns (clusters) found across subjects and the execution of everyday tasks support a modular approach by the brain to generate movement [4, 5]
• Enables Big Data driven approaches to stream processing Body Sensor Network data, Brain-Machine Interfacing and Neuroprosthetics [2]

Acknowledgments
Thanks to Jovana Belic for collecting the data [3]. AT is supported by the National Research Fund, Luxembourg (1229297).

References
[3] Belic, J., MA Thewes, Imperial College, 2010