Comparison of statistical methods for spatio-temporal pattern detection in multivariate point processes

Alessandra Stella¹, Pietro Quaglio¹, Laura Sacerdotè³, Sonja Grün¹,²

¹ Institute of Neuroscience and Medicine INM-6 and Institute for Advanced Simulation (IAS-6) and JARA-Brain Institute I (INM-6), Jülich Research Centre and JARA, Jülich, Germany
² Theoretical Systems Neurobiology, RWTH Aachen University, Germany
³ Department of Mathematics “G. Peano”, University of Torino, Torino, Italy

Contact: a.stella@fz-juelich.de

Introduction

Hebb’s hypothesis [1] states that information processing in the brain is performed by groups of neurons called assemblies. Correlated activity between neurons of the cell assembly is considered as the signature of its activation [2, 3]. Several methods have been designed to extract such correlation and dependency structures from massively parallel spike trains. Here we concentrate on two methods that specifically detect significant repetitions of Spatio Temporal Patterns (STP), i.e. patterns with temporal delays between the composing spikes.

Motivation

- extracting correlation structures from data is a non-trivial problem
- combinatorial explosion of patterns to test for statistical and computational issues
- many methods available in literature, however with different hypothesis and different performances
- need for evaluation and comparison [9]

Goal

Compare two statistical methods - SPADE [4,5] and CAD [6] - for STP detection in multivariate point processes in terms of:

- constraints and limitations
- statistical performance
- computational performance

in order to obtain a possible workflow for the analysis of suitable electrophysiological data.

Methods

SPADE

Spike Pattern Detection and Evaluation [4,5]

SPADE detects all possible repetitions of spike patterns with a mining technique called Frequent Itemset Mining (FIM, Panels A,B,C). The found and counted patterns are evaluated for their statistical significance by a bootstrap technique (Pattern Spectrum Filtering, Panels D,E,F) employing the generation of surrogate data by spike dithering [7]. (Figure on the right)

CAD

Cell Assembly Detection [6]

CAD consists of a statistical test (assuming Poisson distribution of the spike trains) applied to pairwise correlations. It builds on an agglomerative recursive algorithm, starting by extracting all significantly correlated pairs of neurons. The resulting correlated processes are then further correlated with individual spike trains for higher order patterns. (Figure on the left)

Computational Performances

We study the computational time required by the two methods to analyze independent data (no pattern injection - under the null hypothesis of independence) in function of the recording length.

Since CAD does not employ surrogate generation, the analysis is faster by many orders of magnitude with respect to SPADE.

Conclusions

We compare analytically the two methods to see how well they capture structures of pairwise correlation.

- p-values decrease for both methods as the number of occurrences n_AB of a candidate pattern increases, and decrease with the recording length
- less repetitions of a pattern are needed for CAD to obtain significance with respect to SPADE.

We compare the two methods by generating artificial data simulating parallel spike trains with assembly activity. We model massively parallel spike trains by a multi-dimensional stationary Poisson process, into which we superimpose STPs with varying size (number of neurons involved) and frequency (number of occurrences). For each configuration we generate many data sets. Then, we evaluate SPADE and CAD by counting how well they detect the injected patterns (False Positives (FP: incorrect pattern detected) and False Negatives (FN: pattern injected not detected) counts).

- SPADE is almost always able to retrieve the correct pattern
- CAD performs well for patterns with high occurrence rate
- retrieves only subsets of the pattern for medium occurrence rate
- cannot detect the pattern for low occurrence rate.

Table of Comparison between SPADE and CAD

<table>
<thead>
<tr>
<th></th>
<th>SPADE</th>
<th>CAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise Correlation</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Independent Data</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Higher Order STP - Low Number of Occurrences</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Higher Order STP - High Number of Occurrences</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Computational Performances</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

References


Acknowledgments: Helmholtz Portfolio Human (SHINE), Human Brain Project HBP Grant No. 720701 (SGAD) and No. 788671 (SGAD), DFG RTG 1906, the DFG SPP Priority Program 1865 (TR 1153/11-1) and University of Torino.