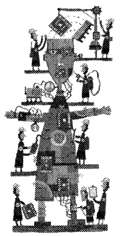


Correlation integral pattern recognition for neural spike trains

Institute of Neuroinformatics

Eidgenössische Technische Hochschule Zürich

Winterthurerstrasse 190, CH-8057 Zurich, Tel: (+41) (0)1 635 30 52, http:// www.ini.ethz.ch

Markus Christen (markus@ini.phys.ethz.ch), Albert Kern, Jan-Jan van der Vyver, Ruedi Stoop

Spike patterns are candidates for neural information encoding. We propose a correlation integral based method for detecting patterns, where we focus on single spike trains. Applications of the method demonstrate the power of correlation integral based pattern recognition. Furthermore, we show the advantages of the method compared to the usual template-based methods for pattern recognition.

1 Introduction

In neural coding theory, the supporters of temporal coding refer to the existence of precise temporal relations in sequences of spike intervals (patterns). Experimentally, patterns have been found in monkey cortical spike trains [1] as well as in the thalamus, and the cortex, of rats [2]. These findings were based on methods for detecting spike patterns in single spike trains [3] as well as in ensembles of spike trains [4]. Most of these methods are template-based, where the statistical significant occurrence of a predefined pattern (template) in a spike train has been analyzed. This is problematic, as the size and the precision of the pattern have to be presupposed. Furthermore, some authors allow extra or missing spikes in patterns [5], which increases the danger of finding statistical artifacts. In patterns in ensembles of spike trains, the problem remains, that one never can be sure if the measured ensemble of spike trains under investigation is complete.

2 Presentation of the Method

Our pattern recognition method is based on the embedding of a time series in a phase space. Let x_t be a sequence of interspike intervals of length T. By choosing an embedding dimension m and a delay rate τ, a series of vectors $\xi_k^{(m)}$ is built: $\xi_k^{(m)} = \{x(t_k), x(t_k+\tau), \dots x(t_k+(m-1)\tau)\}$. If there were patterns present in the time series, the vectors lie on structured manifolds in the embedding space, and otherwise not. To analyze this structure, the correlation integral is used: If the total number of embedded vectors is N, the probability $C_N^{(m)}(\epsilon)$ that the distance between two randomly chosen vectors is smaller than ϵ is estimated by

$$C_N^{(m)}(\epsilon) = \frac{1}{N(N-1)} \sum_{i \neq j} \theta(\epsilon - \|\xi_i^{(m)} - \xi_j^{(m)}\|)$$

where $\theta(x)$ is the heavyside function with $\theta(x) = 0$ for $x \leq 0$ and $\theta(x) = 1$ for $x \geq 0$. Conventionally, the correlation integral is used to determine the correlation dimension [6]. We found the following effect, which can be used for pattern recognition purposes: Patterns in spike trains appear as identifiable steps in the plot of $\log \epsilon$ plotted against the correlation integral $\log C_N^{(m)}(\epsilon)$, because some distances between the embedded vectors are more probable than others, if patterns are present in the time series. Using data with the same probability distribution of the interspike intervals but with no patterns present, this step-like structure does not appear [7].

We also introduce the difference plot to visualize the appearance of patterns. This graph is obtained by plotting $\log \epsilon$ against $\Delta C_N^{(m)}(\epsilon_i) = C_N^{(m)}(\epsilon_{i+1}) - C_N^{(m)}(\epsilon_i)$ for each step $\epsilon_i \rightarrow \epsilon_{i+1}$ and for each embedding dimension. Patterns in this representation show up as peaks.

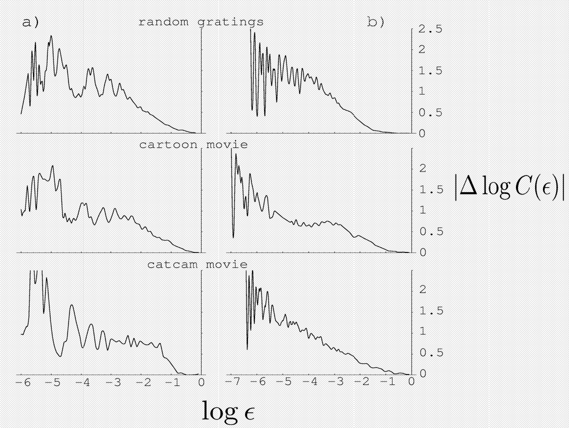
Our method is based on the following assumptions: We consider only single spike trains and do not allow extra or missing spikes. Furthermore, the delay in the embedding can be set to one. For the embedding dimension, different values can be used. The following conclusions can be drawn out of log-log plots and difference plots: First, the appearance of steps or peaks in the plots are a clear indication of the presence of patterns. Second, those ϵ-values or intervals, where steps or peaks are present, indicate the significant time intervals, where no spikes are present.

However, the analysis has to take into account that the number of embedded vectors is finite. Therefore, the method does not give useful results for too small values of ϵ, as the number of points in a small ϵ-environment depends on the number of data points and is not a generic property of the system. In the difference plots, these regions can be detected easily, as suddenly large peaks appear for small ϵ. Usually, files with 5000 data points are sufficient for an analysis.

3 Results

We only present here the application of the method to data from an experiment, which has been performed in the lateral geniculate nucleus (LGN) of a cat. We used data from two cells. Both of them have been exposed to three stimuli: First, to sinusoidal gratings with randomly changing spacial frequency and direction. Second, to a cartoon movie. Third, to a video with natural stimuli. The video has been recorded by putting a camera on the head of a free moving cat (catcam movies). For each class of stimulus between 16 and 40 trials of about 10 to 30 seconds duration have been performed.

For each cell and each stimulus, we generated a single time series of interspike intervals out of the experimental data with length between 6000 to 10000 data points (except catcam movie in neuron a), where we only had 2000 data points). We checked the influence of different ordering the experimental time series and found no relevant effect. We use the difference plot and we sum over the embedding dimensions 1 to 7. The result (figure) show that neuron a) fires in patterns in all three cases, whereas neuron b) does not fire in patterns in neither case.



4 Discussion

The generic statistical characteristic of correlation integral pattern recognition has several advantages to the conventional methods of pattern recognition: It is not necessary to presuppose the size of a pattern in the sense of a template. Concerning precision, we don't have to define any coarse graining once the measurement resolution is given. Correlation based pattern recognition method offers a tool to detect those patterns we consider as relevant from a physiological point of view: statistical significant occurring time intervals where no spikes occur. Furthermore, the LGN-results indicate, that firing in patterns or not firing in patterns is stimulus-independent, which is compatible with our earlier finding of different neuronal firing classes [8].

5 Literature

[1] Lestienne R., Strehler B.L., Brain Res. 437 (1987): 214-238. [2] Tetko I.V., Villa A.E.P., J. Neurosci. Methods 105 (2001): 15-24. [3] Dayhoff, J.E., Gerstein G.L., J. Neurophysiol. 49(6) (1983): 1334-1348. [4] Tetko I.V., Villa A.E.P., J. Neurosci. Methods 105 (2001): 1-14. [5] Strehler B.L., Lestienne R., Proc. Natl. Acad. Sci. U.S.A. 83 (1986): 9812-9816. [6] Grassberger P., Procaccia I., Physica D 13 (1984): 34-54. [7] Stoop R., van der Vyver J.-J., Kern A., NDES 2001 IEEE Conference on Nonlinear Dynamics of Electronic Systems. [8] Stoop R., Blanck D.A., Kern A., van der Vyver J.-J., Christen, M., Lecchini S., Wagner C., Cog. Brain Res. 13 (2002): 293-304.

Acknowledgement and financial support: The autors want to thank Valerio Mante for providing the LGN cat data for this analysis. This work was partially supported by the Swiss National Science Foundation and by a KTI contract with Phonak AG.