Brain Local-Fields Potential Time Series Clustering and Sub Sequence Matching for Behavior Characterization

Advisors

- Andrea Mauri (Lyon 1 Univ., andrea.mauri@univ-lyon1.fr)
- Angela Bonifati (Lyon 1 Univ., angela.bonifati@univ-lyon1.fr)

Context

Local Field Potentials (LFP) refer to electrical signals that are measured in the brain and are thought to reflect the combined synaptic inputs of a particular region. In neuroscience, they are studied to understand the connection between brain activity and behavior (now in primates, with the hope to extend the findings to the characterization of the human brain).
In recent decades, advancements in neural recording techniques have resulted in the generation of increasingly large and intricate datasets in neuroscience. As a consequence, the analysis of these datasets has become a time-consuming task for neuroscientists. To alleviate this burden, there is a growing interest in automating the analysis using machine learning techniques, which can reduce the workload for domain experts. However, when dealing with time series data, specific approaches are needed, especially for variable-length time series commonly encountered in neuroscience experiments. Furthermore, while machine learning algorithms are valuable for processing vast amounts of data that would otherwise require extensive manual effort, they often lack interpretability, which can limit their practical utility.

Assignment

The objective of this thesis is to develop a time series clustering and indexing algorithm to allow neuroscientists to characterize primate behavior from their LFP. For the clustering part, the idea is, starting from a set of *annotated* LFP time series (i.e., each time series is annotated with information about the *action* of the primate), find the clusters that are related to *patterns of actions* of the primate.
For the second part, the assignment is to develop (or adapt) a time series subsequence match algorithm to find those clusters in new time series that aren't annotated (if they are present and where). Subsequence match is a particular problem related to time series data retrieval that consists of, given N time series T1, T2,..., TN of arbitrary lengths, a sequence S and a tolerance e, we want to identify the time series Si ($1 < i < N$) that contain matching subsequences [2].

Tasks:

- Familiarize yourself with the clustering algorithms FeatTS [1] or Time2Feat [3]. Two state-of-the-art human-centered time series clustering algorithms.

- Study the state of the art regarding the time series subsequence match/time series indexing problem.
- Design - or adapt an existing one - and implement a time series subsequence match algorithm
- Evaluate your approach against the state of the art

This thesis is a collaboration between the BD group of the LIRIS lab and the Stem Cell and Brain Research Institute (SBRI) of Lyon1. The neuroscientists will provide real-world datasets and research questions.

Expected abilities:

- Very good programming skills (Python / C++)
- Very good communication skills
- Familiarity with data management techniques
- Familiarity with machine learning methods appreciated

Opportunities: You will have the opportunity to work with top-class researchers in the area of data management and neuroscience, and to be involved in writing a research article documenting the results. The internship also offers the possibility to continue for a PhD scholarship.

References

[1] Bonifati, A., Buono, F. D., Guerra, F., & Tiano, D. (2022). Time2Feat: learning interpretable representations for multivariate time series clustering. Proceedings of the VLDB Endowment, 16(2), 193-201. https://doi.org/10.14778/3565816.3565822

[2] Faloutsos, C., Ranganathan, M., & Manolopoulos, Y. (1994). Fast subsequence matching in time-series databases. ACM Sigmod Record, 23(2), 419-429.

[3] Tiano, D., Bonifati, A., & Ng, R. (2021, May). Feature-driven time series clustering. In 24th International Conference on Extending Database Technology, EDBT 2021. https://hal.science/hal-03548280