Active Learning for Neuroscience Time Series Analysis with Deep Learning

Qitong Wang
LIPADE, UPCité
qitong.wang@etu.u-paris.fr

Stephen Whitmarsh
Paris Brain Institute (ICM)
stephen.whitmarsh@icm-institute.org

Jérôme Cartailler
Lariboisière - Saint Louis Hospitals, UPCité
jerome.cartailler@inserm.fr

Themis Palpanas
LIPADE, UPCité
& Institut Universitaire de France (IUF)
themis@mi.parisdescartes.fr

[***Motivation***] Electroencephalogram (EEG) is an essential method for diagnosing a variety of neurological disorders. However, the practical value of state-of-the-art techniques in clinical practice has remained limited. One field in which this is apparent is epilepsy. Currently, even with the advent of computer-assisted techniques, the detection of epileptic patterns remains an almost human (visual) task. Manual detection of many hours of multi-channel EEG recordings is time-consuming for the clinician and burdens a hospital's human resources. This internship aims to design effective deployment schemes of modern deep learning techniques for epilepsy detection in EEG [1].

[***Objectives***] The goal of this project is to build automated tools (with minimal manual intervention, if any) to facilitate real-world EEG analysis by clinicians and neuroscientists. These tools will use fine-tuned, or trained-from-scratch deep learning models, where only a *small* (i.e., manageable) number of examples will need to be labeled by clinicians. The target outputs of the project are 1) to publish high-level research papers in neuroscience and computer science venues, and 2) to release open-source software to benefit the EEG analysis community.

[***Challenges***] The project's main challenges lie in the complexity of *real-world* EEG datasets. The epileptic spike and other patterns possess different lengths in the time domain, exist across several different channels or frequency ranges, and vary among different patients, deteriorating the performance of many existing heuristic methods. Deep learning solutions may help address some of these problems, though, the inter-patient diversity prevents trained models from being employed on other patients, and the successful deployment of deep network models on individual patients demands thousands of expert-labeled examples, defeating the goal of automated detection.

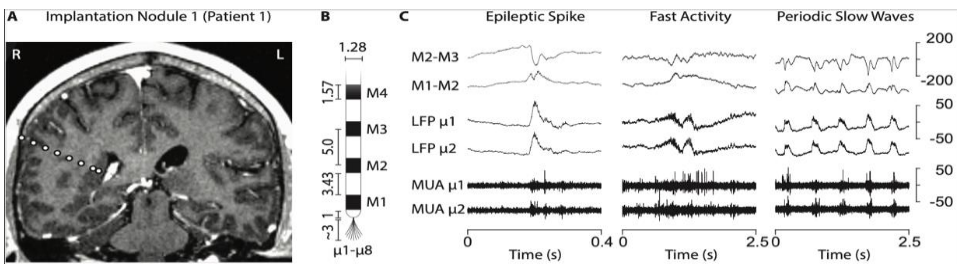

Figure 1. An illustration of our real-world dataset [1].

[***Proposed Work and Implementation***] This project will develop novel model training strategies that only require a small number of labeled data, in order to detect epileptic patterns in real-world stereotaxic EEG (sEEG) datasets, based on our existing code bases, including a ResNet-style deep learning architecture as the backend pattern detection algorithm and a Python web tool as the frontend expert intervention interface. The datasets involved in this project will consist of real-world datasets from our neuroscience collaborators, who are part of this project's team. For example, an ICM dataset we can use is a 3-day recording (32kHz, 20+ channels) from seven patients that carry implanted electrodes, as shown in Figure 1. The total size of the raw datasets is 1.7TB.

The main task of this project is to train a neural network from scratch with minimal expert-labeled examples using active learning strategies [2]: conduct an experimental analysis to determine the proper model capacity [3] (represented by the number of layers, the dimensional of latent features, etc.) and fitting the ICM dataset; design a suitable example selection strategy (eg, entropy-based scoring, consensus-based scoring, diversity-based scoring, and a mixture of these strategies); investigate model training strategy for the active example selection (e.g., soft-labelling, auto-labelling trick); explore active transfer learning strategies that take into account model drift between patients.

[***Prerequisites***] Excellent Python skills, very good knowledge of deep learning frameworks (PyTorch/GPU, etc.) and libraries in data analysis workflow (NumPy, Matplotlib, etc.). Research/projects on deep learning or data analysis is a plus.

[***How to apply***] Email CV and transcripts to Prof. Themis Palpanas: themis@mi.parisdescartes.fr

References

[1] Valerio Frazzini, Stephen Whitmarsh, et al. "Human periventricular nodular heterotopia shows several interictal epileptic patterns, associated with hyperexcitability of neuronal firing". Frontiers of Neurology (in press)

[2] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, Xin Wang: A Survey of Deep Active Learning. ACM Comput. Surv. 54(9): 180:1-180:40 (2022)