DA231X Project Proposal
Degree Project at EECS, 2022

Wilhelm Ågren, wagren@kth.se

December 1, 2021

1 Thesis title

Feature extraction on MEG data using Self-Supervised Learning.

2 Background

Functional neuroimaging techniques such as electroencephalography (EEG) and magnetoencephalography (MEG) can be used to monitor subject health or be used in Brain Computer Interfaces (BCI). Labeling medical imagery data can be difficult, and often requires expert knowledge which can be hard to come by [1]. Self-Supervised Learning (SSL) is a deep learning technique which aims to train on unlabeled data using a pretext task. This pretext task should drive the SSL model to be able to predict missing or hidden data, in both temporal and spatial domain. Researchers have in the recent years explored the use of deep learning techniques and SSL on EEG data to for example automatically detect seizures or to classify sleep stages [2][3][4]. Deep learning research on MEG data is more scarce and not as explored as EEG, although an example is [5]. A reason for this scarcity could be that EEG data is more publicly available, with plenty of datasets to choose from, whereas for MEG there are almost no occurrences of publicly available datasets [6].

This proposed project would explore how state-of-the-art SSL algorithms learn latent space representations of MEG data. Furthermore, applying the extracted features, from the latent space, to perform classification and regression tasks on subject labels to investigate if they are represented in the extracted features. The MEG data has been provided by researchers at Karolinska Institutet and Stockholm University as part of the SLEMEG project. This project aims to study the effect of partial sleep deprivation on neurophysiological processes by utilizing MEG. We believe that this proposed project could lay a foundation for future research on SSL and MEG, whilst building upon the established research on EEG data.

3 Research question

Do the extracted features form distinct clusters when applying non-linear dimensionality reductions such as t-SNE?

If the extracted features form clusters, will a hierarchical cluster analysis show that these are robust?

Are the features static along each channel, or do they portray over time in sequences?

How well can a model perform classification and regression tasks on the extracted features based on subject labels such as gender, age, reaction time?

4 Hypothesis

We think the results will show that the extracted features encodes some information regarding subject gender and state, but not regarding age or reaction time. Furthermore, we also believe that the extracted features will form distinct clusters, but we do not know on what basis they will cluster or if they will be robust. Finally, we believe that the extracted features along a channel will portray sequence-characteristics and thus not be static across an entire recording. In what way these sequences of features will show depending on subject labels, we have no current hypothesis about.

5 Research method

The project methodology will be separated into two distinct parts that each aim to help answer the research questions. Firstly, a literature study will be conducted as basis for the written thesis. Literature will be searched for on databases that grant access to KTH students, such as *arXiv* and *IEEE Explore*. More recent machine learning literature in the field of SSL will be favoured over old, but papers detailing the neuroimaging techniques EEG/MEG are all equally favoured to an extent. Secondly, a quantitative study will be conducted where experiments are designed and explored as to answer the overarching research questions. The results statistical significance will be determined by performing some statistical test, such as Two-Sample Kolmogorov-Smirnov for comparing extracted features.

6 Background of the student

I have taken multiple courses which I believe to be useful, relevant, and of appropriate knowledge for this project. First and foremost, I have completed the required course DA2210 *Introduction to the Philosophy of Science and Research Methodology*. Secondly follows a number of completed courses which I think to be relevant and appropriate:

- DD2437 *Artificial Neural Networks and Deep Architectures*,

- DD2421 *Machine Learning*,
- DD2424 *Deep Learning in Data Science*,
- DD2420 *Probabilistic Graphical Models*,
- DD2418 *Language Engineering*.

Furthermore, I have some extra experience with applied deep learning from individual projects which I've worked on during my masters. For example an agent trained using deep reinforcement learning to play the game Snake, and an agent for playing Chess that utilizes a deep CNN to statically evaluate positions.

7 Supervisor and examiner

I have been in contact with Erik Fransén who expresses great interest to serve as supervisor for this project. He was the assigned supervisor for my group in the project course DD2430 and this degree project would follow up and develop what was started in that course. Erik briefly suggested that we ask Arvind Kumar to be examiner for the project.

8 Eligibility

I have been granted exemption from the requirements to start the degree project, more specifically being done with all credits form the bachelors part of my degree. I assume a reason for this was because I only lacked 5 credits, which I am currently reading. This means that at the start of next year I will have completed all necessary credits for my bachelors. I did not miss a single credit during my first year of the masters, so I have the 60 credit requirement completed already. Although, I am still waiting for results in the course DD2430 which lies as a big foundation for this degree project.

9 Study planning

Since I am in this period reading the final credits of my bachelors I am 7.5 credits behind in the masters. These credits I plan to read during my degree project. I have applied to the course LS1419 *English for Employment* which starts in period 3. Reading this course at the same time as my degree project, and finishing the last 5 credits of my bachelors this period, would mean I am done with my degree by the summer.

References

[1] Yuji Roh, Geon Heo, and Steven Euijong Whang. "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective". In: *IEEE Transactions on Knowledge and Data Engineering* 33.4 (2021), pp. 1328–1347. DOI: 10.1109/TKDE.2019.2946162.

[2] Hubert Banville et al. *Uncovering the structure of clinical EEG signals with self-supervised learning*. 2020. arXiv: 2007.16104 [stat.ML].

[3] Siyi Tang et al. *Automated Seizure Detection and Seizure Type Classification From Electroencephalography With a Graph Neural Network and Self-Supervised Pre-Training*. 2021. arXiv: 2104.08336 [eess.SP].

[4] Stanislas Chambon et al. "A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.4 (2018), pp. 758–769. DOI: 10.1109/TNSRE.2018.2813138.

[5] Ivan Zubarev et al. "Adaptive neural network classifier for decoding MEG signals". In: *NeuroImage* 197 (2019), pp. 425–434. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2019.04.068. URL: https://www.sciencedirect.com/science/article/pii/S1053811919303544.

[6] *PhysioNet Databases*. Accessed 01-12-2021. MIT Laboratory for Computational Physiology. URL: https://physionet.org/about/database/.