Can NLP Models and Methods Be Applied to Eeg Data?

Lino Casanova and **Andreas Marfurt**
lino.casanova@gmail.com

Abstract

Our work introduces an application of Transformer models and transfer learning to electroencephalogram (EEG) data. We start by creating embeddings, then pretrain on large EEG data and finally finetune on the task of sleep stage classification. Related work in the field of sleep phase classification includes the YASA algorithm based on LightGBM. In the processing of EEG data with transformers, Neuro-GPT stands out as a pioneer. The embeddings are created by first segmenting EEG signals into uniform-length parts and utilizing power spectral density (PSD) analysis to derive feature vectors. The embeddings are then constructed from these feature vectors. The model's architecture consists of a BERT encoder followed by a classifier. In a first step, the model was pretrained geusing EEG data from the Temple University Hospital EEG Corpus with a masked language modeling objective. Subsequently, the model was finetuned using only 197 whole-night polysomnographic recordings with annotated sleep stages. The results show that using the PSD analysis to embed EEG data is a valid approach. Furthermore they indicate that pretraining on large EEG datasets enhances performance on downstream tasks. Finetuning on the downstream task of sleep stage classification with 5 classes achieves a mean accuracy of 58%, which is nontrivial but can still be improved. There is a noticable difference between the accuracy of the "Wake", "REM" and "N3" phases at 69% and that of the "N1" and "N2" phases, which is 41%. To summarize, our study shows that applying NLP methodology to EEG data works almost without adaptations. At the same time, there is a lot of room for improvement and we expect further advancements in this field.