Using a Pre-trained Neural Language Model to Make Character Predictions for Brain-Computer Interfaces

Dylan Gaines[1*], Keith Vertanen[1], and CAMBI[2]

[1]Michigan Technological University, Houghton, Michigan, USA; [2]Consortium for Accessible Multimodal Brain-Body Interfaces

*1400 Townsend Drive, Houghton, MI, USA. Email: dcgaines@mtu.edu

Introduction: Because the signals acquired from a BCI user can be quite noisy, often a language model is used to assist the software in selecting the user's intended character. While traditional *n*-gram language models can provide adequate predictions and can be queried quite efficiently once trained, state-of-the-art in natural language processing typically uses a large pre-trained neural network language model such as GPT-2 [1]. In this work we explore some of the speed and accuracy tradeoffs between these two models.

Material, Methods, and Results: We used a training set of 100 phrases to test a range of parameters for our neural language model to determine the most efficient option that still provided high quality predictions. To measure the accuracy of predictions, we used *perplexity*, which describes the branching factor of the search. A lower average perplexity indicates more accurate predictions. We ran a test set of AAC-like phrases [2] through our model with the chosen parameters and through a large 12-gram model that was trained on AAC-like data [3]. We found that a GPT-2 model with 124M parameters trained on text scraped from webpages had a perplexity of 4.82 with an average prediction time of 1.72 s, while a larger GPT-2 model with 355M parameters yielded a better perplexity of 4.18, while increasing the average prediction time to 2.86 s. The 12-gram model had a perplexity of 2.47 with an average prediction time of 0.04 ms.

Discussion and Significance: Character predictions can help to improve users' typing speed and accuracy while using BCI [4]. We observed substantial differences in prediction speed and accuracy between *n*-gram and pre-trained neural language models. While this work showed that the 12-gram model outperformed the GPT-2 model in both perplexity and prediction time, we conjecture that this difference may be due in large part to the text they were trained on. In future work, we plan to adapt our GPT-2 model to AAC-like text to investigate if we are able to achieve further accuracy gains to the point where it might justify the higher prediction time. It is also possible that even larger versions of the GPT-2 model could lead to further improvements.

Acknowledgements: This work was supported by NSF Graduate Research Fellowship 2034833.

References

[1] Radford, A., et al. 2019. Language Models are Unsupervised Multitask Learners. *OpenAI Blog* 1(8): 9.

[2] Costello, J. 2017. *Message Banking, Voice Banking and Legacy Messages*. Boston Children's Hospital.

[3] https://imagineville.org/software/lm/dec19_char/

[4] Speier, W., et al. 2012. Natural Language Processing with Dynamic Classification Improves P300 Speller Accuracy and Bit Rate. *J. Neural Eng.* **9** 016004