Transfer learning in Brain-Computer Interfaces: Language-Pretrained Transformers for Classifying Electroencephalography

Wolf De Wulf[1,2][0000−0002−0219−3120]

[1] Vrije Universiteit Brussel, Belgium, Artificial Intelligence Lab
[2] University of Edinburgh, United Kingdom, School of Informatics
wolf.de.wulf@ed.ac.uk

Keywords: Deep Learning · Transfer Learning · Brain Computer Interfaces · Natural Language Processing
Supervisors: Prof. Dr. Kevin De Pauw & Prof. Dr. Geraint Wiggins

A Brain-Computer Interface (BCI) is a device that digitises the neurophysiological. By interpreting brain-generated signals, BCIs perform a variety of tasks in both a clinical and an everyday context. Recent advances in BCI research have allowed clinical applications such as early prediction of epileptic seizures (Young et al., 2011; Lin et al., 2018) and spellers for the paralysed (Nijboer et al., 2008; Townsend et al., 2010), as well as everyday applications such as intent recognition for smart living (Zhang et al., 2019; Belkacem et al., 2020) and vehicle and robot control (Aznan et al., 2019; W. Lu et al., 2020). Nevertheless, aside from a number of specific cases, efficient and reliable BCIs are still a distant prospect. Interperson and intraperson variability as well as data scarcity make decoding neuroimaging modalities a complex matter. Similar challenges have been investigated in other research fields, yielding solutions that can inspire new BCI methods. Particularly interesting and novel in the research area of BCIs is the self-supervised learning paradigm, originally cross-pollinated from research areas that have access to vast amounts of unlabelled data, such as Computer Vision (CV; van den Oord et al., 2019; Grill et al., 2020) and Natural Language Processing (NLP; Radford et al., 2019; Devlin et al., 2019; Raffel et al., 2020). The nontransferable nature of the current BCI machine learning state-of-the-art has resulted in variable performance (Ahn & Jun, 2015; Lotte et al., 2018), suggesting that research into self-supervised learning approaches could be fruitful. Various recent works have started laying the foundation for such research (Banville et al., 2019, 2021; Kostas et al., 2021). The work presented in this thesis is part of this effort.

Transfer learning in BCIs has mostly consisted of supervised pretraining followed by supervised finetuning. Pretraining would then consist of initialising the weights of a deep learning model by training it on the data of a number of participants (Dose et al., 2018; Fahimi et al., 2019) or even the runs of a single participant (Schwemmer et al., 2018). In some cases, pretraining happens in a completely different domain, for example, image recognition (Xu et al., 2019). Such research can be interesting as the success or failure of such knowledge transfer can provide insight and inspiration with regards to BCI transfer learning. Recent work by K. Lu et al. (2021) has found that language-pretrained transformers (Vaswani et al., 2017) are surprisingly transferable to a variety of nonlanguage modalities, such as the MNIST handwritten digit benchmark (Deng, 2012) and protein sequence tasks, such as remote homology detection (Rao et al., 2019). Therefore, the work presented in this thesis is to result in two types of insight. On the one hand, inspiration can be gained towards BCI transfer learning. On the other hand, K. Lu et al.'s claim that language-pretrained transformers are universal computation engines is tested towards the complex task of classifying EEG.

Explicitly, the research question studied in this thesis is as follows: "How and to what degree can a language-pretrained transformer transfer to the classification of EEG and does the language-pretraining influence performance?" An attempt is made to finetune the second generation Generative Pretrained Transformer (GPT2; Radford et al., 2019) to a four-class EEG motor imagery classification task, namely, the Graz dataset A (Brunner et al., 2008). To investigate whether or not and to what degree language pretraining is beneficial, the performance of a frozen language-pretrained instance of GPT2 is compared to that of an unfrozen randomly initialised instance. When kept frozen, GPT2 is able to achieve 41.6 % classification accuracy, which lies above the random assignment mark of 25.0 %. Furthermore, it is significantly better than what a randomly initialised instance of GPT2 achieves, i.e., 26.0 % ($p < 0.0001$). Although the achieved performance is incomparable to the state-of-the-art, it can be concluded that positive transfer is, in fact, possible. The unsupervised pretraining phase of GPT2 manages to capture structure in language that is applicable to the classification of EEG. By gradually unfreezing the layers of a language-pretrained instance of GPT2, an attempt is made at gaining more nuanced insight regarding the role of the language-pretraining. While the results are not conclusive, they suggest that some learning can take place and that future work with larger datasets might allow for more insight as well as generally higher classification scores.

Resources The source code accompanying this thesis can be found on GitHub[3]. A summary of all results can be found on Weights & Biases[4].

Acknowledgements Geraint Wiggins and Arnau Dillen merit thanks for their guidance and support in the realisation of this work. The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

[3] https://github.com/wulfdewolf/lpt-for-eeg

[4] https://wandb.ai/wulfdewolf/lpt-for-eeg/reports/Transfer-learning-in-BCI-s-language-pretrained-transformers-for-EEG-classification--VmlldzoxOTIxNDU2?accessToken=r4hzxv3i86ovxcf01fdzcebnnpy79nc57stoew4gasvoboual6f2c93131ra4u1z

References

Ahn, M., & Jun, S. C. (2015, March). Performance variation in motor imagery brain-computer interface: A brief review. *Journal of Neuroscience Methods*, *243*, 103–110. doi: https://doi.org/10.1016/j.jneumeth.2015.01.033

Aznan, N. F. N., Connolly, J. D., Moubayed, N. A., & Breckon, T. (2019). Using Variable Natural Environment Brain-Computer Interface Stimuli for Real-time Humanoid Robot Navigation. *2019 International Conference on Robotics and Automation (ICRA)*. doi: https://doi.org/10.1109/ICRA.2019.8794060

Banville, H., Albuquerque, I., Hyvärinen, A., Moffat, G., Engemann, D.-A., & Gramfort, A. (2019, October). Self-Supervised Representation Learning from Electroencephalography Signals. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1–6). doi: https://doi.org/10.1109/MLSP.2019.8918693

Banville, H., Chehab, O., Hyvärinen, A., Engemann, D.-A., & Gramfort, A. (2021, March). Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering*, *18*(4), 046020. doi: https://doi.org/10.1088/1741-2552/abca18

Belkacem, A. N., Jamil, N., Palmer, J. A., Ouhbi, S., & Chen, C. (2020). Brain Computer Interfaces for Improving the Quality of Life of Older Adults and Elderly Patients. *Frontiers in Neuroscience*, *14*. doi: https://doi.org/10.3389/fnins.2020.00692

Brunner, C., Leeb, R., Muller-Putz, G. R., & Schlogl, A. (2008). BCI Competition 2008 – Graz data set A. *Dataset*.

Deng, L. (2012). The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, *29*(6), 141–142.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*.

Dose, H., Møller, J. S., Iversen, H. K., & Puthusserypady, S. (2018, December). An end-to-end deep learning approach to MI-EEG signal classification for BCIs. *Expert Systems with Applications*, *114*, 532–542. doi: https://doi.org/10.1016/j.eswa.2018.08.031

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T.-S., Ang, K. K., & Guan, C. (2019, January). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *Journal of Neural Engineering*, *16*(2), 026007. doi: https://doi.org/10.1088/1741-2552/aaf3f6

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., ... Valko, M. (2020). Bootstrap your own latent - a new approach to self-supervised learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 21271–21284). Curran Associates, Inc.

Kostas, D., Aroca-Ouellette, S., & Rudzicz, F. (2021). BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From

Massive Amounts of EEG Data. *Frontiers in Human Neuroscience*, *15*. doi: https://doi.org/10.3389/fnhum.2021.653659

Lin, S.-K., Istiqomah, Wang, L.-C., Lin, C.-Y., & Chiueh, H. (2018). An Ultra-Low Power Smart Headband for Real-Time Epileptic Seizure Detection. *IEEE Journal of Translational Engineering in Health and Medicine*. doi: https://doi.org/10.1109/JTEHM.2018.2861882

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018, June). A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update. *Journal of Neural Engineering*, *15*(3), 031005. doi: https://doi.org/10.1088/1741-2552/aab2f2

Lu, K., Grover, A., Abbeel, P., & Mordatch, I. (2021, June). Pretrained Transformers as Universal Computation Engines. *arXiv:2103.05247 [cs]*.

Lu, W., Wei, Y., Yuan, J., Deng, Y., & Song, A. (2020). Tractor Assistant Driving Control Method Based on EEG Combined With RNN-TL Deep Learning Algorithm. *undefined*.

Nijboer, F., Sellers, E., Mellinger, J., Jordan, M., Matuz, T., Furdea, A., ... Kübler, A. (2008, August). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, *119*(8), 1909–1916. doi: https://doi.org/10.1016/j.clinph.2008.03.034

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. *OpenAI*.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, *21*(140), 1–67.

Rao, R., Bhattacharya, N., Thomas, N., Duan, Y., Chen, X., Canny, J., ... Song, Y. S. (2019, December). Evaluating Protein Transfer Learning with TAPE. *Advances in neural information processing systems*, *32*, 9689–9701.

Schwemmer, M. A., Skomrock, N. D., Sederberg, P. B., Ting, J. E., Sharma, G., Bockbrader, M. A., & Friedenberg, D. A. (2018, November). Meeting brain–computer interface user performance expectations using a deep neural network decoding framework. *Nature Medicine*, *24*(11), 1669–1676. doi: https://doi.org/10.1038/s41591-018-0171-y

Townsend, G., LaPallo, B. K., Boulay, C., Krusienski, D., Frye, G. E., Hauser, C. K., ... Sellers, E. (2010). A novel P300-based brain–computer interface stimulus presentation paradigm: Moving beyond rows and columns. *Clinical Neurophysiology*. doi: https://doi.org/10.1016/j.clinph.2010.01.030

van den Oord, A., Li, Y., & Vinyals, O. (2019, January). Representation Learning with Contrastive Predictive Coding. *arXiv:1807.03748 [cs, stat]*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is All you Need. In *Advances in Neural*

Information Processing Systems (Vol. 30). Curran Associates, Inc.

Xu, G., Shen, X., Chen, S., Zong, Y., Zhang, C., Yue, H., ... Che, W. (2019). A Deep Transfer Convolutional Neural Network Framework for EEG Signal Classification. *IEEE Access*, *7*, 112767–112776. doi: https://doi.org/10.1109/ACCESS.2019.2930958

Young, C., Liang, S.-F., Chang, D., Liao, Y.-C., Shaw, F., & Hsieh, C.-H. (2011). A Portable Wireless Online Closed-Loop Seizure Controller in Freely Moving Rats. *IEEE Transactions on Instrumentation and Measurement*. doi: https://doi.org/10.1109/TIM.2010.2050358

Zhang, X., Yao, L., Zhang, S., Kanhere, S., Sheng, M., & Liu, Y. (2019). Internet of Things Meets Brain–Computer Interface: A Unified Deep Learning Framework for Enabling Human-Thing Cognitive Interactivity. *IEEE Internet of Things Journal*. doi: https://doi.org/10.1109/JIOT.2018.2877786