AI for Adaptive Brain-Computer Interfaces: Challenges and Opportunities

Ruixue Liu
Worcester Polytechnic Institute
Worcester, MA 01609, USA
rliu2@wpi.edu

Erin T. Solovey
Worcester Polytechnic Institute
Worcester, MA 01609, USA
esolovey@wpi.edu

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
Research in brain-computer interfaces (BCIs) has explored a set of methods to capture information about users' state from brain signals, with the goal of developing adaptive interfaces. The rise of modern artificial intelligence and machine learning methods has been central to the advancement of the BCI field. While these methods show promise for addressing some problems for brain data modeling, there also exist some challenges. We discuss the challenges faced when using deep learning in this field, including the tension caused by the small sizes of brain datasets, issues related to the generalizability of the models and the need for deriving interpretable results from these models. Furthermore, we discuss the approaches we are taking to address these challenges.

Author Keywords
Deep learning; brain-computer interfaces.

Introduction
Advances in interaction techniques have made computer systems increasingly user-friendly. However, the use of computer systems still depends on a machine-mandated sequence of explicit commands from users. At the same time, humans have limited perceptual and communication capabilities. This presents a communication bottleneck and source of potential error in human-computer interac-

tion [12]. To optimize user experiences and improve task performances, we need computer systems to incorporate more information beyond the explicitly given commands and build a model of the user. Based on the model of the user, the computer would then be able to "understand" the user and adapt its behavior to the individual user according to users' changing states. For example, if an online learning system can gain insight into the learner's cognitive states, it can adapt its behavior to the learner's changing states and provide a better learning experience.

We are exploring the use of measurements of brain activity from fNIRS brain sensors alongside student log data to understand important mental activities during learning [3, 6]. It will allow us to explore novel human-computer interaction paradigms for utilizing sensors that provide passive, continuous, implicit input to interactive systems.

Background

Relevant information to build the user model includes the user's intentions, subjective interpretations, and emotions. Research in BCIs has explored a set of methods to capture such information from brain signals. Some notable methods are functional Magnetic Resonance Imaging (fMRI), Electroencephalography (EEG), and function Near-Infrared Spectroscopy (fNIRS). fMRI is an effective technique for brain function imaging in clinical or laboratory settings, but it is not practical for realistic human-computer interaction settings because it is expensive and vulnerable to the existence of metal objects and head movement. EEG has been the main technology used in brain-computer interface research due to its low cost, portability, and high temporal resolution. fNIRS is a relatively new neuroimaging device which measures the changes in oxygenated and deoxygenated blood in the cortex, and also has potential for HCI.

To enable brain data to be used as input to adaptive interactive systems, there is a need to build models that can automatically predict user's states based on the brain signals. With the advancements in machine learning, researchers have attempted to move from offline statistical analysis of the brain data to real-time automated classification of users' state. Traditional machine learning methods, such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), k-nearest neighbor (kNN), and Hidden Markov Models (HMM) have been widely applied to detect patterns from brain data [7, 8]. These classifiers gained popularity in the BCIs research community because of their simplicity and low computational requirements. However, these machine learning classifiers can not adequately consider the spatial-temporal dynamics of brain data. The results of these classifiers depend heavily on the features extracted from data. Because brain data signals are time-series, many 'standard' machine learning methods are not designed to identify and extract informative features from brain signals.

Challenges

More recent work has investigated using deep learning methods for brain data classification, including Convolutional Neural Networks (CNN), Deep Belief Networks (DBN), and Long Short-Term Memory (LSTM) network [10, 15, 1, 2]. These approaches have shown their ability to detect spatial patterns and short-long-term dependencies from time-series data by automatically extract higher-level features. However, there also exist challenges.

Size of Brain Datasets

Collecting and labeling brain data is costly and time-consuming. As such, the sizes of brain datasets are usually small. At the same time, deep learning methods require a large number of training examples, and insufficient training data could lead to the poor performance of deep learning models. For real-world applications, the accuracy of the models need to be improved [9].

Model Generalizability

For real-world applications, users could engage in various tasks while interacting with a system. Therefore, to enable brain data to be used as input for real-word adaptive interactive systems, there is a need to build robust models that can automatically predict user's states across different tasks. However, previous work in deep learning for BCIs have been focused on modeling a single specific task, and the models built are specially designed using task-specific knowledge. At the same time, different classification tasks based on brain data could have different characteristics. For example, the sizes of datasets could be different depending on the number of trials in an experiment and the number of participants. As such, most models built in previous research fail to achieve satisfactory results when being applied to other tasks, and it remains unclear how to build a model that can generalize across different tasks [5].

Interpretable Results

In addition to the outputs from the models, the ability to explain these results is also important for researchers in the BCI community. A main goal in the field of neuroscience is to understand how certain brain activation is linked to specific cognitive processes. Therefore, explaining how the models make specific predictions would be helpful not only in verifying the validity of the model by comparing to the findings in the neuroscience literature, but also in providing insights for further investigation [13]. Furthermore, for designing and building adaptive BCIs, interpretable results can help users understand the behavior of the models and thus enable meaningful interactions with the users. However, little research has been focusing on improving the interpretability of deep learning models for brain data.

Current Work

To solve the challenges posed by the small sizes of brain datasets and models' generalizability, we will investigate self-supervised learning and transfer learning techniques. Self-supervised learning trains a model using labels that are naturally part of the input data, without requiring separate external labels. Transfer learning can store knowledge gained from solving one task and applying it to another related task. By utilizing these techniques, we will have access to pre-trained models, which will then be used as the starting point of the specific task with fewer training data. This would help us improve the performance of models trained on small datasets [4, 11]. Moreover, the transfer learning methods would allow us to investigate the generalizability of models by applying the same model on different tasks. To solve the challenge of model interpretability for brain data, we are planning to research methods for transforming pre-trained deep learning models into explanatory graphs [14].

REFERENCES

[1] Salma Alhagry, Aly Aly Fahmy, and Reda A El-Khoribi. 2017. Emotion recognition based on EEG using LSTM recurrent neural network. *Emotion* 8, 10 (2017), 355–358.

[2] Johannes Hennrich, Christian Herff, Dominic Heger, and Tanja Schultz. 2015. Investigating deep learning for fNIRS based BCI. In *2015 37th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2844–2847.

[3] Shelby Keating, Erin Walker, Anil Motupali, and Erin Solovey. 2016. Toward real-time brain sensing for learning assessment: Building a rich dataset. In *Proceedings of the 2016 CHI Conference Extended*

Abstracts on Human Factors in Computing Systems. 1698–1705.

[4] Alexander Kolesnikov, Xiaohua Zhai, and Lucas Beyer. 2019. Revisiting self-supervised visual representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 1920–1929.

[5] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering* 15, 5 (2018), 056013.

[6] Ruixue Liu, Erin Walker, and Erin Solovey. 2017. Toward neuroadaptive personal learning environments. In *The First Biannual Neuroadaptive Technology Conference*. 59.

[7] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. 2018. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering* 15, 3 (2018), 031005.

[8] Noman Naseer and Keum-Shik Hong. 2015. fNIRS-based brain-computer interfaces: a review. *Frontiers in human neuroscience* 9 (2015), 3.

[9] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. 2019. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering* 16, 5 (2019), 051001.

[10] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping* 38, 11 (2017), 5391–5420.

[11] Chuanqi Tan, Fuchun Sun, Bin Fang, Tao Kong, and Wenchang Zhang. 2019. Autoencoder-based transfer learning in brain–computer interface for rehabilitation robot. *International Journal of Advanced Robotic Systems* 16, 2 (2019), 1729881419840860.

[12] Erin Treacy Solovey, Daniel Afergan, Evan M Peck, Samuel W Hincks, and Robert JK Jacob. 2015. Designing implicit interfaces for physiological computing: Guidelines and lessons learned using fNIRS. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 6 (2015), 1–27.

[13] Yujun Yan, Jiong Zhu, Marlena Duda, Eric Solarz, Chandra Sripada, and Danai Koutra. 2019. Groupinn: Grouping-based interpretable neural network for classification of limited, noisy brain data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 772–782.

[14] Quan-shi Zhang and Song-Chun Zhu. 2018. Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering* 19, 1 (2018), 27–39.

[15] Wei-Long Zheng, Jia-Yi Zhu, Yong Peng, and Bao-Liang Lu. 2014. EEG-based emotion classification using deep belief networks. In *2014 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.