Deep Learning in Brain Computer Interfaces

Petia Georgieva

petia@ua.pt

University of Aveiro, Department of Electronics Telecommunications and Informatics
Institute of Electronics and Informatics Engineering of Aveiro -Intelligent Robotics and Systems Lab
Aveiro, Portugal

Figure 1: Brain Computer Interface (BCI) concept

KEYWORDS

Deep Learning, Convolutional Neural Networks, Brain Computer Interface, Affective Neuro Computing

ACM Reference Format:

Petia Georgieva. 2019. Deep Learning in Brain Computer Interfaces. In *Proceedings of 9th Balkan Conference on Informatics (BCI'19)*. September 26–28, 2019, Sofia, Bulgaria. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3351556.3351594>

1 INTRODUCTION

Deep Learning (DL) refers to computational models composed of multiple processing layers capable of learning representations of data with multiple levels of abstraction [1]. Since the seminal work of Lecun *et al.* [2] on training convolutional networks using back-propagation and gradient-based optimization, DL has outperformed other machine learning techniques in many domains of science and industry. Hinton *et al.* [3], achieved breakthrough on the task of

automatic speech recognition, the first major DL industrial application. DL has achieved great success in recognition tasks within a wide range of applications including computer vision (images, videos), natural language processing (text), bioinformatics, to mention a few.

Convolutional Neural Networks (CNN) lie at the core of the most powerful DL architectures working with images and video data, primarily due to their ability to extract representations that are robust to partial translation and deformation of input patterns. However, despite the success story of CNNs they remain relatively unexplored in the Brain-Computer Interfaces (BCI).

BCIs provide an alternative communication channel between the human brain and a machine. Mental activity is reflected into electrophysiological signals such as Electroencephalogram (EEG). The BCI system detects such changes and transforms them into control signals, which can be used for moving objects, writing letters, opening doors, changing TV channels or other everyday household activities (Fig. 1). The main difficulty in building a noninvasive BCI is to effectively separate the brain signal from the noise and find a meaningful signal interpretation. Identifying which spatio-temporal features of the EEG recordings are the most relevant as well as selecting the most appropriate EEG frequency components is the key to infer the subjects intentions. The identification strategies are usually based on the input of human experts to identify and mark particular features of the EEG signal. DL methods have the potential to automatically extract features from wider range data (in-depth feature learning) which may have been previously

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BCI'19, September 26–28, 2019, Sofia, Bulgaria

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-7193-3/19/09...\$15.00

<https://doi.org/10.1145/3351556.3351594>

overlooked from focusing solely on specific electrodes or frequencies.

A challenging application of EEG-based BCIs is to find correlates between human emotions and the registered EEG signals, known as Affective Neuro Computing (ANC) [4], [5]. EEG-based ANC systems are useful for brain studies, psychology research and clinical applications as a complement to the human medical expertise.

2 DEEP LEARNING IN ANC

In the ANC context, deep learning is still at an early explorative stage, with Echo State Networks (ESN) demonstrating robustness with respect to data variability among subjects [6]. ESN belongs to the class of recurrent neural networks (RNN) where the parameters of the hidden recurrent layer (called reservoir) are generated randomly and only the ESN output parameters are trained. This ESN property speeds up the model training however there is a risk to become unstable. An improved ESN adaptation rule was proposed in [7], coined Intrinsic Plasticity (IP) to guarantee the reservoir parameters tend to equilibrium states and observed that these states are concentrated in different regions depending on the inputs. If two input vectors are close in the input space, they will result in close equilibrium points in the reservoir state. These findings defined the feature selection approach: the EEG recordings are mapped into the reservoir equilibrium states. The dominant features are extracted iteratively from low dimensional combinations of these states. The ESN-based software sensor outperformed state of the art methods for detecting the positive and negative valence of human emotions across subjects.

The default DL approach is to process directly raw ERPs, without using domain knowledge of the data to create input features. However, the collaborative alternative to build DL model that computes internal representations of features created by human experts seems to come up with less complex and more interpretable deep architectures, something that is worth to further explore.

3 CHALLENGES AND OPPORTUNITIES

The deep learning models for brain computer interfaces are mainly based on advances in convolutional and recurrent neural networks and efficient parallel solvers. Their major advantage is that they reduce the need of feature engineering, which is one of the most difficult part in machine learning practice. However, there are many challenges in adopting deep learning in real practice.

First, DL models require a large amount of labeled training data to achieve superior performance than classical machine learning methods based on pre-selected features. The studies on deep learning for brain imaging use different datasets as well as different imaging modalities, making difficult to conduct a comparison on the performance of the methods.

Second, the published research works include many ad-hock choices of hyperparameters such as the depth of the network, the dimension of the convolutional filters, number of units in each layer, input dimension, etc. Future research requires well designed diagnostics of the DL - based BCI configurations to gain insight what isn't working well with the learning algorithm and how to improve its performance. Taking into account the large number of hyperparameters this may require significant amount of time and

computational resources even with the support of most powerful GPU hardware.

Third, the body of the trained DL model is like a black box, we still lack the perfect methodology to fully comprehend its deep structure. Consequently, training deep models from scratch requires large amounts of annotated data and massive computational resources.

Transfer learning has emerged as a promising solution to adapt or modify existing deep learning architectures on new applications. It uses, as a baseline, the knowledge from a deep model previously trained on a large labeled data set. The basic assumption is that layers close to the input can learn low-level descriptors not specific to a particular data set, while layers near the output can detect high-level features specific to a given task. The potential of transfer learning and fine tuning in the context of brain imaging despite differences between the original and the target data is an open research direction. One approach is to fix the weights up to a chosen layer of the network and, then, reshape and retrain the remaining higher layers for the new data set. An opposite approach to improve the performance is to fine tune the lower layers of the pretrained model. The idea is to back-propagate the errors from the new task into the first layers to fine tune the base features to the new task. Open-source tools (Caffe, Tensorflow, MatConvNet) and GPU-computing libraries (CUDA, cuCNN, OpenCL) are attractive implementation tools for the choice of a pretrained CNN model.

The capability of CNN to model highly nonlinear functions comes with high computational and memory demands both during the model training and inference. The models are usually designed with huge number of parameters to account for the complexity of large scale data that they learn from. However, when it comes to production deployment on embedded or mobile devices, the network size, speed, and power consumption become an issue. From the other side, deep learning models are often over-parameterized with significant redundancy among the weights. The concept of sparsity seems a good way to account for parameter redundancy and compressed sensing. For example training deep neural autoencoders with sparsity constraint may enforce the neurons to learn non redundant information and therefore would reduce the amount of information processed. The sparsity constraint seems to favor the extraction of common neural signatures for discriminating the same brain states (e.g. emotion valence) across different subjects. All these challenges address problems not specific only for BCIs and represent opportunities for research advancement and innovation.

REFERENCES

- [1] Goodfellow I, Bengio Y and Courville A. Deep Learning. Cambridge, MA, USA: MIT, 2016.
- [2] LeCun, Y., Bottou, L., Bengio, Y. Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324, 1998.
- [3] G. Hinton, Li Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine* 29, 82–97, 2012.
- [4] O. Georgieva, S. Milanov, P. Georgieva, I.M. Santos, A.T.Pereira, C. F. da Silva Learning to decode human emotions from ERPs, *Neural Computing and Applications*, Springer, vol. 26 (3), 2015 . DOI 10.1007/s00521-014-1653-6.
- [5] L. Bozhkov, P. Georgieva, I. Santos, A. Pereira, C. Silva. EEG-based subject independent affective computing models. *Procedia Computer Science*, Elsevier, 375–382, 2015, DOI: 10.1016/j.procs.2015.07.314.
- [6] Bozhkov, L., Koprinkova-Hristova, P., Georgieva, P. Learning to decode human emotions with Echo State Networks. *Neural Networks*, Vol. 78, 112–119, 2016.
- [7] Bozhkov, L., Koprinkova, P., Georgieva, P., Reservoir computing for emotion valence discrimination from EEG signals. *Neurocomputing*, Vol. 231, 28–40, 2017.