DOI: 10.3217/978-3-85125-467-9-155

Predicting Serial Visual Presentation Events from EEG Using Spatial-temporal Convolution Neural Network

Zijing Mao[1], Yufei Huang[1]
[1]The University of Texas at San Antonio, San Antonio, Texas, USA

Introduction: An important problem concerning BCI applications is event prediction based on EEG data. Past research has identified a host of event-related potentials (ERPs) such as P300 and N1 that are indicative of different basic sensory, cognitive, and motor events. However, the ERPs can change in both magnitude and timing with subjects and experiments, making cross-subject prediction based on ERPs less reliable. Additionally, significant improvement in spatial and temporal resolution of EEG has tempted us to predict much more complex cognitive events that can produce a variety of EEG patterns highly convoluted in space, time, and frequency. To address these issues, we investigate deep learning (DL) solutions in this paper. Recently DL has shown more and more widely applications on BCI tasks (H Cecotti, et al., 2011) and the key to DL's success is its ability to automatically discover discriminate feature representations (Y Bengio, et al., 2013) that are essential for accurate prediction from raw signals. However, designing an appropriate DL model is most of the time hinders further usage. The key factor complicating this process is there are numerous nuisance hyper-parameters from DL model require fine-tuning. Therefore, in order to take full advantage of the learning ability from deep learners for RSVP signals and lighten DL model designation workload, we have investigated several architectures of DL and tested their performances on a time-locked rapid serial visual presentation (RSVP) dataset for comparison.

Material, Methods and Results: In this study, we considered an RSVP experiment called the Cognitive Technology Threat Warning System (CT2WS) [1-2]. We performed leave-one-subject-out test involving 15 subjects with in total about 10,400 epochs (~700 epochs per subject). Considering different DL architectures, we investigated deep neural network (DNN) modules, which are consists of several fully-connected layers; a proposed hierarchical DNN (HDNN) modules which tries to capture the local temporal correlations in raw EEG signals; and the proposed spatial-temporal convolution neural networks (STCNN) which is designed specifically to capture both spatial and local temporal correlations using multiple filters or kernels. Additionally, we also performed tests on three existing DL architectures designed for EEG recognition, named as CNN with 2 temporal filters (CNN2TF, P. W. Mirowski, et al., 2008); CNN with 1 spatial filter (CNN1SF, H Cecotti, et al., 2013); and CNN with 1 spatial filter and 1 temporal filter (CNN1SF1TF, H Cecotti, et al., 2011). Moreover, inspired from image recognition, we also tested our RSVP dataset on classical CNN for computer vision: ImageNet (A Krizhevsky et al., 2012) and GoogleNet (C Szegedy et al., 2014) which also surprising can achieve performable accuracy even better than some of the specific architecture designed for EEG signal processing. Overall, our proposed STCNN achieved the best performance with 89% AUC of ROC, followed by CNN1SF1TF with 88% and then ImageNet with 86% of AUC. DNN has the worst performance (84% of AUC) among all DL algorithms, similar to xDAWN (84% of AUC), which is the state-of-art algorithm for RSVP.

Discussion: In this paper we have tested 8 different DL architectures, including two proposed DL models for EEG based event classification, which are designed to improve the representation of spatial and local temporal correlations in EEG. We showed almost all DL have a significant improvement over other shallow learning algorithms, which is consistence with other works conclusion that deep learners have better capability on feature representations. Besides, models designed to capturing both spatial and temporal features outperforms other DL algorithms which is also consistence with our hypothesis that a specify design in the convolutional layer to exploit local correlations can significantly improve the performance. Finally, to explain the performance of ImageNet and GoogleNet, we believe on the one hand, these networks also captures the spatial and temporal correlation of EEG signals therefore they are better than DL models without using such designation; and on the other hand, since EEG epochs arrange all channels on a 1D vector which smears the spatial filtering from capturing the correct local correlation, they are not the ideal model to use either.

Significance: We have achieved several unique features specifically tailored for RSVP BCI tasks. First, we have investigated the existing DL models for EEG signals and our proposed STCNN which have achieved the best performance in the RSVP dataset. Second, we showed that DL models designed specifically to capture spatial and temporal features will benefits RSVP classification performance. Finally, our trained STCNN model on RSVP dataset can be extended as a feature extractor for other RSVP tasks. Since the architecture of STCNN is applicable for general EEG classification, additional effort in the future to investigate its performance for various BCI tasks is desirable.

References

[1] "U.S Department of Defense Office of the Secretary of Defense," Code of federal regulations, protection of human subjects. 32 CFR 219, 1999.
[2] "U.S. Department of the Army. Use of volunteers as subjects of research. AR 70-25," Washington, DC: Government Printing Office, 1990.