Proceedings of the
8th Graz Brain-Computer Interface Conference 2019
DOI: 10.3217/978-3-85125-682-6-03

DECODING SSVEP ON TIME AND FREQUENCY DOMAIN USING CONVOLUTIONAL NEURAL NETWORK

R. Gala[1], J. Xu[#1,2], Z.Wu[1,2], A. Nürnberger[1]

[1]Data and Knowledge Engineering Group, Faculty of Computer Science, Otto von Guericke University, Magdeburg, Germany
[2]Institut für Medizinische Psychologie, Otto von Guericke University, Magdeburg, Germany

E-mail:jiahua.xu@med.ovgu.de

ABSTRACT: A brain-computer interface (BCI) is an highly interdisciplinary research topic which involved psychology, signal processing, machine learning, medical engineering etc. A steady-state visually evoked potentials (SSVEP) paradigm-based Brain Computer Interface systems are the most promising and reliable communication systems for people with disabilities on clinical purposes. This paper proposes the use of Deep Neural Network models for decoding the data acquired through visual-based Electroencephalography (EEG). in this paper, we concentrate on two points with the Convolutional Neural Network The first is decoding of the EEG data using the raw signal (time domain) and the extracted frequency features of the signal after Fourier analysis. The second point explored in this paper is the evaluation of the selection of electrodes on the performance of the Deep Neural Network models for raw signals (time domain data). We have also compared the performance of Canonical Correlation analysis of the data with the Deep Neural Network models.

INTRODUCTION

Brain-Computer Interface (BCI) systems provide direct communication between the human brain and an external device. BCI systems are applicable in different areas such as spellers [1], for operating wheelchairs [2], computer games, virtual reality and controlling home environment [3]. BCI systems make use of different types of signals comprising of slow cortical potentials, P300, sensorimotor rhythms and SSVEPs [1]. SSVEPs are recorded over the scalp from repetitive external stimulations. SSVEPs are of the same frequency as of the stimulus. In other words, looking at an image of frequency 12Hz with 0° phase angle will generate a signal of the same value and angle in the occipital region of the brain. High Signal-To-Noise (SNR) ratio, high Information Transfer Rate (ITR), little user training and less susceptibility to eye movements are the advantages leading to the usage of SSVEPs in BCIs over other types of signals [4].

Essentially, a visual BCI has four sections, signal acquisition, signal processing, signal classification, and output device which realizes the selected commands by the user of BCI indicated by Figure 1. The principal task in

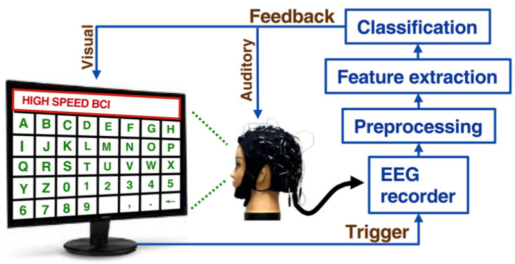

Figure 1: Working of SSVEP-based BCI [8]

this process is signal classification with reliable performance in regards to accuracy and response time. A number of methods for detection of SSVEPs have been used such as Minimum-Energy Combination (MEC), Canonical Correlation Analysis (CCA), Support Vector Machine (SVM) and Power Spectral Density Analysis (PSDA). These methods would require the data to be refined by human intervention. The use of Deep Neural Networks in BCIs has gained popularity due to its robust nature and the capability of extracting high-level features from the data without prior knowledge [7][11]. Use of Convolutional Neural Networks (CNN) for image classification and recognition provides high accuracy, automatic feature extraction with the multiple convolutional layers thus making it the perfect choice for EEG data. The EEG data received from a variety of participants having different features makes CNN an apt solution for classification of the signals. Despite of high performance of CNN, it has a necessity to have a large training dataset so that no or less over-fitting occurs in the model, which is a challenge in the EEG data, given it's relatively smaller size.

In this paper, there are two parts of the implementation, the first is the impact on performance, when giving raw signal as input and extracted frequency features as input. The second part compares the performance of the selection of electrodes and without the selection of electrodes in the input.

MATERIALS AND METHODS

EEG Dataset: This paper uses data published in [5], which is recorded using a 40-target BCI speller. The BCI

 DOI: 10.3217/978-3-85125-682-6-03

speller consists of a 5x8 matrix of 40 characters, out of which 26 are English alphabets, 10 digits, and 4 symbols. A total number of 35 participants (27 naive, 8 experienced) participated in the experiment. The experiment had 6 blocks, each comprising of 40 trials corresponding to the 40 symbols on the screen. One trial total length is 6 s, with 0.5 s target cue and last 0.5 s was blank. There was a gap of several minutes between two consecutive blocks for the participant to rest. The 40 frequencies recorded are from 8 Hz to 15.8 Hz with an interval of 0.2 Hz. The whole-head EEG data is recorded using 64 electrodes according to the 10-20 system at a sampling rate of 1000Hz as shown in Figure 2. The recorded data epochs were downsampled to 250 Hz as the upper limit of the SSVEP range is 90 Hz. No pre-processing has been performed on the data. At the end of the experiment, the data available for each participant is 240 trials i.e. 6 blocks x 40 trials. Each trial consisted of 64 channels x 1500 time points.

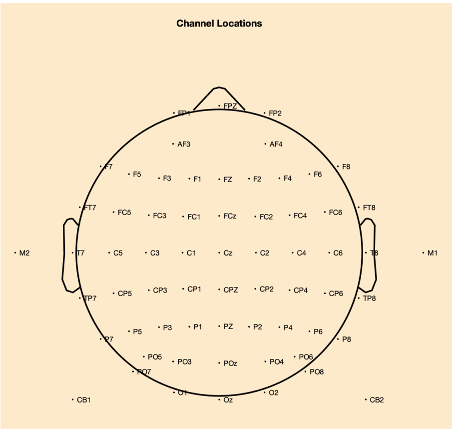

Figure 2: Electrode placement

Canonical Correlation Analysis: Canonical Correlation Analysis is a multivariable statistical method which finds a correlation between two sets of variables. It was introduced in the field of EEG analysis by Lin et. al in [9]. In the case of SSVEP-based EEG signals, there are two variables, X which is the recorded multi-channel EEG signal and Y refers to the reference signals. The frequency recognition is obtained by calculating the canonical correlation between multi-channel SSVEP and the reference signals. The frequency which has the maximum correlation value in the reference signals is the same as the frequency of the multi-channel SSVEP signal. CCA helps in reducing a large amount of information into useful information by maximizing the correlation.

Convolutional Neural Network: Recently CNNs have attained success and popularity in so many different fields. The different layers of CNN help in dimensionality reduction, in turn, reducing the number of training parameters which will increase the training speed and improve performance. A CNN consists of an input layer, convolutional layer, activation, pooling layer, and fully connected layer. Various parameters such as Dropout and Batch Normalization can be used for further optimization in the CNN model. The convolution layer slides a filter of a particular size over the given input to produce a feature map. As done in any Neural Network, we use activation function on the output of the convolution layer after which we have a non-linear output. The dimensionality reduction is performed by the pooling layers, which extracts useful parameters and thus reduce the number of parameters to compute and to be learned. This will also prevent overfitting of the model. There can be such multiple layers, each comprising of convolution layer, activation function, and pooling layer. The last few layers are fully connected layers which are similar to the regular neural networks. But before we pass the data from convolutional layers to the fully connected layers, we need to flatten the data as fully connected layers understand only 1-dimensional data. The main part of the training is the configuration of the above-described layers.

Table 1: CNN for Time Domain

Layer / Parameter	Number / Size
Conv1D	32 filters
MaxPooling1D	2
Dropout	0.5
BatchNormalization	-
Conv1D	64 filters
MaxPooling1D	2
Dropout	0.5
BatchNormalization	-
Flatten	-
Dense	256 filters
Dense	512 filters, 40 output

In the past few years, the use of CNN in the area of SSVEP has increased. In [6], the accuracy of CNN was 69.03% 256 channel SSVEP recordings of 11 subjects, whereas in [7] the accuracy of CNN classification is as high as 99.27% for static and 94.03% for ambulatory SSVEP data of 7 subjects. In these papers, a power spectral density analysis has been performed to extract frequency features and then feed it to the model for classification. Here in this paper, we will compare the performance of the model between such extracted features and raw signal.

CNN Model for raw signal: The raw signal is in form of time domain as indicated by 64 channel x 1500 time points is given as input to the CNN model. To evaluate the performance of the CNN model on the raw signal as input, we make use of Convolutional Neural Network in 1 dimension. The structure is described in Table 1, 1 Conv1D layer with 32 filters having 'ReLU' as activation function and a Max Pooling layer with size 2. The second layer comprises of 1 Conv1D layer with 64 filters having 'ReLU' as activation function and a Max Pooling layer with size 2. The Dropout value is taken as 0.5 and BatchNormalization has been applied. Fully connected layers of size 256 and 512 have been used. The last fully connected layer uses Softmax as the activation function for giving the output. Adam optimization algorithm is used for updating the network weights in training data. The training and validation accuracy indicated in Figure 3

Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-03

Table 2: CNN for Frequency Domain

Layer / Parameter	Number / Size
Conv2D	32 filters
MaxPooling2D	3
Dropout	0.5
BatchNormalization	-
Conv2D	64 filters
MaxPooling2D	3
Dropout	0.5
BatchNormalization	-
Flatten	-
Dense	512 filters, 40 output

shows the learning of the model.

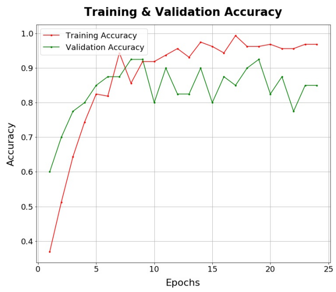

Figure 3: Training & Validation Accuracy - Time Domain

CNN Model for extracted frequency features: Power Spectral Density Analysis (PSDA) is performed on the signal before giving input to the CNN model. Here we use Convolutional Neural Network in a 2 dimensions matrix. The structure is described in Table 2, 1 Conv2D layer with 32 filters having 'ReLU' as activation function and Max Pooling layer with size 3x3. The second layer consists of 64 filters with 'ReLU' as activation function and Max Pooling layer with a filter of size 3. The Dropout value is taken as 0.5 and BatchNormalization has been applied. The last layer which is fully connected layer is of size 512 with Softmax as an activation function. Adam optimization algorithm is used just like the previous for updating the network weights in training data. The training and validation accuracy indicated in Figure 4 shows the learning of the model.

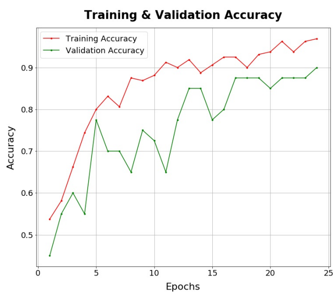

Figure 4: Training & Validation Accuracy - Frequency Domain

Selection of electrodes: The SSVEP signals decoding have a much better quality when the electrodes placed

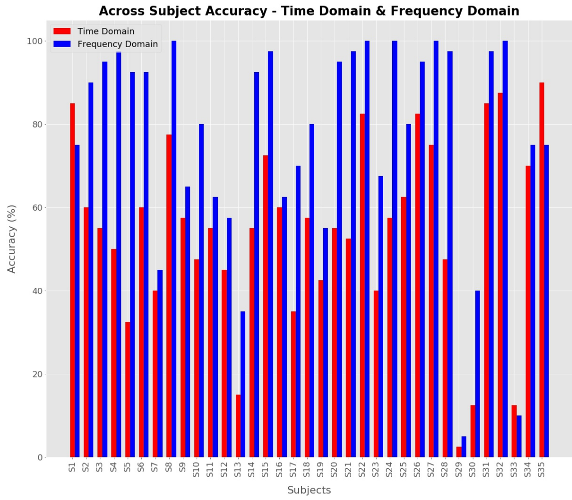

Figure 5: Overall Accuracy of Subjects in both domains

over the occipital and parietal areas are thus removing other background activities [5]. From the 64 electrodes shown in Figure 2, the electrodes selected in this paper are 9 i.e. Pz, PO5, PO3, POz, PO4, PO6, O1, Oz and O2 which give a better performance in classification of the signal [5]. The data from those electrodes is selected for classification. The same model is given two different inputs of raw signal (time domain) and the result is compared to check the performance of the model in both scenarios. The structure of the model is just the same as the one described in Table 1. The channel selection is tested in the case of the time domain.

RESULTS

Comparison of CNN models for raw signal and extracted frequency features: In our models, we have attained an accuracy of 76.5% in the time domain, 80% accuracy in the frequency domain. The CNN model with time domain performs almost as well as the model with frequency extracted features. The accuracy of the model in time domain is quite higher in certain subjects than in frequency domain as shown in Figure 5 indicating that the extraction of the frequency component is alternative. The idea behind using CNN models is an automatic feature extraction and learn the signal specific oscillation in the hidden layers which is achieved in the model used for raw signal. From the given data, we have used has 240 trials per subject, the number of samples is quite less for the model to train in the time domain.

Channel Selection results: The model achieves 76.5% accuracy when data from only 9 channels is taken into consideration and achieves 22.5% accuracy without selection of 9 channels. In [5], it is mentioned that utilization of electrodes positioned over occipital and parietal will give better accuracy in classification of the given signals. Our CNN model verifies that the channel selection proves beneficial in the classification as shown in Fig-

 DOI: 10.3217/978-3-85125-682-6-03

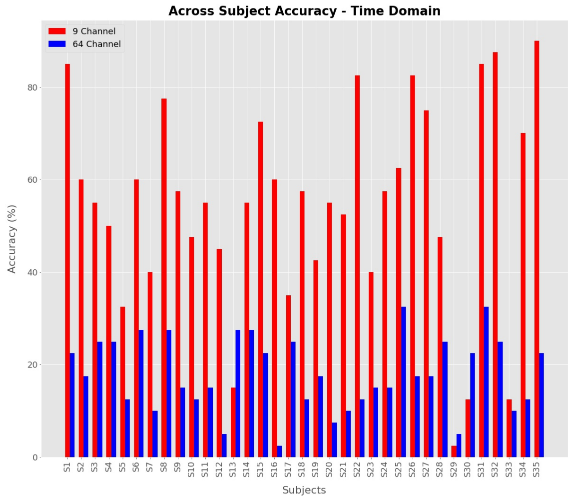

Figure 6: Overall Accuracy of Subjects with 9-channel and 64-channel data

ure 6.

Comparison of CCA with CNN models: The accuracy of CCA for all the subjects can be compared with the accuracy of CNN models in various conditions. The CCA model requires hand-crafted data for processing and getting the maximum correlation values. While the CNN models automatically detect the high-level features with the help of hidden layers. It can also be observed that the CCA and CNN give quite similar results on certain subjects, however as the most successful algorithms in SSVEP decoding for many years,CCA generated the overall higher accuracy than the time and frequency domain. The CCA and CNN results are shown in Figure 7 and Table 3.

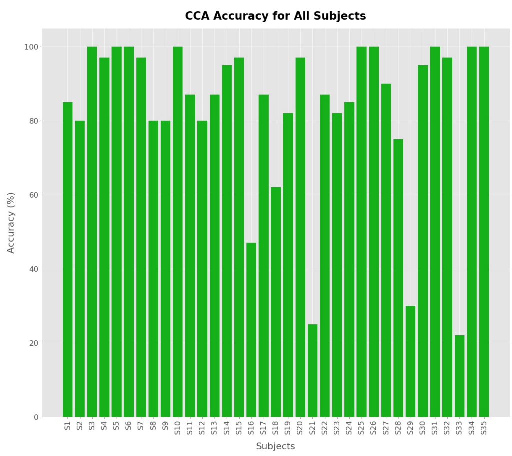

Figure 7: Overall Accuracy of Subjects using CCA

DISCUSSION

Table 3: CCA,time and frequency domain accuracy

Sub\Acc	**CCA**	**Time Domain**	**Frequency Domain**
Subject 1	85	85	77.5
Subject 3	100	57.5	95
Subject 5	100	52.5	92.5
Subject 6	100	60	92.5
Subject 7	97	40	42.5
Subject 9	80	58.5	62.5
Subject 10	100	58.5	80
Subject 11	87	55	62
Subject 12	80	47.5	58.5
Subject 13	87	17.50	35
Subject 14	95	60	90
Subject 17	87	37.5	70
Subject 19	82	45	50
Subject 20	97	58.5	90
Subject 23	82	40	70
Subject 25	100	65	80
Subject 26	100	82.5	85
Subject 29	30	2.5	5
Subject 30	95	12.5	40
Subject 31	100	85	97.5
Subject 33	22	12.5	10
Subject 34	100	70	77.5
Subject 35	100	90	77.5
Mean	87.21	51.02	65.55
Sub\Acc	**CCA**	**Time Domain**	**Frequency Domain**
Subject 2	80	60	90
Subject 4	97	50	97.5
Subject 8	80	78.5	100
Subject 15	97	72.5	98
Subject 16	47	60	62
Subject 18	62	58.5	80
Subject 21	25	60	95
Subject 22	87	82	100
Subject 24	85	60	100
Subject 27	90	77.5	98.5
Subject 28	75	50	98.5
Subject 32	97	90	100
Mean	76.83	65.5	92.90
Sub\Acc	**CCA**	**Time Domain**	**Frequency Domain**
Overall	83.65	56.51	76.30

The evaluation of the CNN model on time and frequency domain with different architectures were performed. from the result, we reconfirmed the method of electrode selection gave a lot higher accuracy than the usual data without selection with deep neural networks. In the time domain data, the CNN model itself learns the low-level and the high-level features and makes the decision of discarding the unnecessary data unlike in the extracted frequency features. the CCA was implemented to compare the traditional algorithm and deep neural networks. We conclude two points from the above experiments. The first being that given the raw signal and adjusted parameters, CNN can classify the input data without prior feature extraction. Initially, the classification accuracy did not increase as the number of parameters like Dropout, BatchNormalization, Regularizers were continuously updated and the model's accuracy was checked.

 DOI: 10.3217/978-3-85125-682-6-03

The second conclusion is that in a certain kind of data, the data taken from selected electrodes would assist the model in better classification than the whole data.

CONCLUSION

Two different deep neural network structure on time and frequency domain have been put forward and tested in this paper. The accuracy of the model in the time domain is quite higher in certain subjects than in the frequency domain which may indicate that the extraction of the frequency component is alternative. Generally on all subjects, frequency domain show the overall higher accuracy than the time domain. The traditional CCA algorithm still achieved the best accuracy in the case of small data on time domain which is not unexpected. The selection of channels play an important role in identifying the brain rhymes. Despite having good accuracy, each of the model can be further tested and configured to achieve even more accuracy given more amount of data. Data augmentation method will be applied for the next round test. 240 trials per subject is not enough for CNN to learn the features and also might generate the overfitting problems in the model. Hence, a reiteration was performed and then selected values for the parameters of regularization to avoid overfitting. Furthermore, CNNs can be combined with Recurrent neural network (RNN)to explore Brain physiological data on time domain data as well as on frequency domain.

ACKNOWLEDGMENT

Jiahua.Xu and Riddhi.Gala contributed this paper equally, Jiahua.Xu proposed the framework of methods,scripted the core code and revised the draft, Riddhi.Gala implemented the code and drafted the paper, thanks Andreas.Nürnberger and Zheng.wu for feedbacks and comments. The authors gratefully acknowledge financial support from China Scholarship Council.

REFERENCES

[1] Chen X, Chen Z, Gao S, Gao X. A high-ITR SSVEP-based BCI speller. Brain-Computer Interfaces. 2014;1:3-4, 181-191.
[2] Leeb R, Friedman D, Müller-Putz GR, Scherer R, Slater M, Pfurtscheller G. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with tetraplegics. Computational Intelligence and Neuroscience. 2007; 79642:1-8.
[3] Gao X, Xu D, Cheng M, Gao S. A BCI-based environmental controller for the motion-disabled. IEEE Trans Neural Syst Rehabil Eng. 2003; 11:137-140.
[4] Wu Z, Lai Y, Xia Y, Wu Y, and Yao D. Stimulator selection in SSVEP-based BCI. Medical Engineering & Physics. 2008;30:8,1079-1088.
[5] Wang Y, Chen X, Gao X, and Gao S. A Benchmark Dataset for SSVEP-Based Brain-Computer Interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. 2016; 25:10, 1746?1752.
[6] Thomas J, Maszczyk T, Sinha N, Kluge T, Dauwels J. Deep Learning-based Classification for Brain-Computer Interfaces, in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, October 5-8, 2017. 234-239.
[7] Kwak N, Müller K, Lee S. A convolutional neural network for stead state visual evoked potential classification under ambulatory environment. PLoS One. 2017; 0172578.
[8] Chen X, Wang Y, Nakanishi M, Gao X, Jung T, Gao S. High-speed spelling with a noninvasive brain-computer interface, in Proceedings of National Academy of Sciences, 2015;112:44, E6058-67.
[9] Lin Z, Zhang C, Wu W and Gao X 2006 Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Transactions Biomedical Engineering, 2006; 53:6, 1172-1176.
[10] Rezeika A, Benda M, Stawicki P, Gembler F, Saboor A and Volosyak I. Brain-Computer Interface Spellers: A Review. Brain Sciences. March 30 2018; 8:57.
[11] Lee H, Choi Y. A Convolution Neural Network Scheme for Classification of Motor Imagery EEG based on Wavelet Time-Frequency Image.IEEE 2018 International Conference on Information Networking. 2018. 906-909.