Neural network pruning for feature selection Application to a P300 Brain-Computer Interface

Hubert Cecotti and Axel Gräser

Institute of Automation (IAT) - University of Bremen
Otto-Hahn-Allee, NW1, 28359 Bremen, Germany

Abstract. A Brain-Computer Interface (BCI) is an interface that enables the direct communication between human and machines by analyzing brain measurements. A P300 speller is based on the oddball paradigm, which generates event-related potential (ERP), like the P300 wave, on targets selected by the user. The detection of these P300 waves allows selecting visually characters on the screen. We present a new model for the detection of P300 waves. The techniques is based on a neural network that uses convolution layers for creating channels. One challenge for improving pragmatically BCIs is to reduce the number of electrodes and to select the best electrodes in relation to the subject particularities. We propose a feature selection strategy based on salient connexions in the first hidden layer of a neural network trained with all the electrodes as input. A new classifier is created in relation to the remaining topology and the desired number of electrodes for the system. The recognition rate of the P300 speller over two subjects is 87% by considering only 8 electrodes.

1 Introduction

A Brain-computer interface (BCI) is a direct communication pathway between a human brain and an external device. BCI systems do not require any movement [1]. For this reason, BCIs are usually used for persons with severe disabilities like spinal cord injuries, who are unable to communicate through any classical devices [1]. BCIs often use EEG (electroencephalography) techniques for recording brain signals as this method is more convenient. A BCI command usually translates the EEG signal produced according to a particular stimulus (visual or not). The EEG response can correspond to event-related potentials, event-related desynchronization/synchronization (ERD/ERS) or slow cortical potentials. In relation to the paradigm, different kinds of features must be extracted. Different types of classifiers are used for classifying and detecting particular brain responses. Among these classifiers, artificial neural networks have been widely used in the field [2, 3, 4, 5, 6]. Besides other models like support vector machine (SVM) [7, 8] and hidden Markov models [9] have been proven successful for EEG classification. Neural networks can be used for classifying EEG and therefore be used for a BCI.

In this paper, we focus on P300-BCIs. P300-BCIs use visual evoked potentials as brain responses. The P300 wave is an event related potential (ERP) that can be recorded easily via EEG. The wave corresponds to a positive deflection in voltage at latency of about 300 ms in the signal after a particular visual stimulus. It means that after an event like a flashing light, a deflection in the

signal occurs after 300ms. If a P300 wave is detected 300ms after a stimulus in a specific location, it is possible to deduce that the user was paying attention to this same location. The signal is typically measured most strongly by the electrodes covering the parietal lobe. The detection of a P300 wave is equivalent to the detection of where the user was looking 300ms before its detection; Farwell and Donchin first introduced P300 potential into BCI in 1988 [10]. The location of the electrodes where the signal has a high intensity depends of the subject. For a non-experimental BCI, it is not a practical solution to cover the whole head with electrodes. The position of the electrodes and their number must be chosen wisely. The choice of the electrodes corresponds to a feature selection problem. We propose a new method for selecting the best electrodes. This method is based on the analysis of the weights of a neural network once this network has been trained with all the electrodes. Contrary to methods like the optimal brain damage [11], the goal is to prune useless weights only in the first hidden layer. The selection of the best active weights allows removing useless electrodes in the classification of P300 waves.

The classifier for the detection of P300 waves is described in the first section. The second section is dedicated to the feature selection strategy. The experiments and the database are detailed in the third section. Finally, the results and their discussion are presented in the last section.

2   Classifiers

For the detection of P300 waves in the EEG, a classifier based on a convolutional neural network (CNN) has been used. This neural network is a multi-layer perceptron (MLP), which contains more than one hidden layer. Besides, the hidden layers are not fully connected. A special topology in relation to the problem translates a particular path in the information processing. These neural networks are used for object recognition [12] and have been successfully used for handwriting character recognition [13]. They allow the automatic feature extraction within layers and keep as input the raw information without specific normalization, except for scaling and centering the input vector. This kind of model has many advantages when the input data contain an inner structure like for images and where invariant features must be discovered, i.e. when a kernel based method cannot catch the complexity of the problem. One advantage of convolutional neural network is the possibility to include knowledge inside the network, contrary to kernel based methods [14]. One other interest is to avoid hand designed input features, which are not derived by the general problem.

The network is composed of 5 layers. Each layer has a specific semantic: the first hidden layer represents the creation of channels; the second hidden layer subsamples and filters the signal. The input layer, L_0, represents the raw EEG signal $I_{i,j}$ with $0 \leq i < N_{elec}$ and $0 \leq j < N_t$ where N_{elec} is the number of electrodes considered in the experiment, N_t is number of sampled points in the signal. In the experiments, $N_t = 78$ it corresponds to 650ms of the recorded signal after a flashing light. The first hidden layer, L_1, is composed of N_s maps.

We define $L_1 M_m$, the map number m. Each map of L_1 has the size N_t. Each neuron j of this layer is connected to N_{elec} corresponding neurons in the input. Furthermore, for one map, each neuron shares the same set of weights. It assures the independence of the weights over time within one pattern. The second hidden layer, L_2, is composed of $5N_s$ maps. Each map of L_2 has 6 neurons. Each neuron of this map is connected to 13 neurons of the previous layer, without overlapping. The third hidden layer, L_3, contains 100 neurons. This layer is fully connected to the different maps of L_2. Finally, the output layer, L_4, contains 2 neurons that represent the 2 classes of the problem (P300 and not P300). This layer is fully connected to L_3.

The weights are corrected during learning thanks to a gradient descent by minimizing the least mean square error of the validation database. 95% of the training database is used effectively for learning whereas the remaining 5% is used as the validation database.

3 Feature selection strategy

As the first hidden layer corresponds to the creation of the channel, it is possible to extract information about the most relevant electrodes once the network is trained. When a weight is close to 0 then it means that its discriminant power is very low. We define the power of the electrode i by:

$$\xi_i = \sum_{j=0}^{j=N_s} |w(i,j)|$$

where $0 \leq i < N_{elec}$ and $w(i,j)$ represents the weight of a link between any neuron of the map j to the electrode i at any time. ξ_i is the combination of the different maps that compose the network. Therefore, it is possible to create a new classifier with a pre-fixed number of x electrodes by selecting x electrodes, which correspond to the x higher ξ values. In this case, the input size is reduced. It can also be seen as pruning the useless weights in the initial network. We define several CNN that will be used in the next sections. CNN-T is the classifier that uses all the electrodes. CNN-8-FS considers the 8 best relevant electrodes. The classifier CNN-8 corresponds of the arbitrary selection of 8 electrodes: F_Z, C_Z, P_Z, P_3, P_4, PO_7, PO_8 and O_Z. These electrodes were chosen in relation to the guideline provided during a BCI tutorial in Utrecht, Holland, 2008.

4 Experiments

The considered database is the data set II from the third BCI competition [15]. In these experiments, the subject was presented with a matrix of size 6 by 6, that contains characters: [A-Z], [1-9] and [_]. The main classification problem has therefore 36 classes. The subject's task was to sequentially focus attention on characters from a pre-defined word. The 6 rows and 6 columns of this matrix were successively and randomly intensified at a rate of 5.7Hz. The character to

select is defined by a row and a column. Therefore, 2 out of 12 intensifications of rows/columns highlighted the expected character, i.e. 2 of the 12 intensifications should produce a P300 response. Row/column intensifications were block randomized in blocks of 12. The sets of 12 intensifications were repeated 15 times for each character epoch. All the rows/columns were intensified 15 times. Thus, 30 P300 responses should be detected for each character.

Signals have been recorded from two subjects in five sessions with the BCI2000 framework [16]. Signals were collected from 64 ear-referenced channels. The signal was bandpass filtered between 0.1 and 60Hz and digitized at 240 Hz [17]. The training database is composed of 85 characters while the test database contains 100 characters. Each character epoch is supposed to contain 2 P300 signals, one for a row flash and one for the column flash. For the training database, the number of P300 to detect is 85*2*15.

5   Results

The evaluation of the P300 speller is divided into two steps: the results of the P300 detection and their impact in the character recognition. Table 2 presents the recall and precision obtained for the detection of P300 waves. The recall and precision are defined by Recall=TP/(TP+TN) and Precision=TP/(TP+FP), where TP, TN and FP represents respectively the number of true positive, true negative and false positive. The ξ for all electrodes are displayed in figure 1. The gray-level represents the values of ξ (dark for high values, white for low values). We can observe the precise location of the relevant electrodes on subject A whereas the information is more elusive with subject B. The character recognition rate (in %) is presented for several epochs in table 2. The best accuracy is achieved with CNN-T (94.5%). The accuracy reaches 87% and 87.5% for CNN-8-FS and CNN-8 respectively. These results highlight the low difference between the feature set created by using the connections with the highest salience and the fixed feature set given from the neuroscience field when the information is spread.

As expected and like for the P300 detection, the results are lower than when all the 64 electrodes are used. The precision of the P300 response is better when the electrodes are selected thanks to the proposed strategy. The recall is also better for subject B. However, these improvements in the detection are only translated for subject A, who provides better results in character recognition compared to the fixed choice of the electrodes. It can be explained by the concentration of the relevant electrodes in particular locations (around P_Z). As the information is more dispersed and homogeneous with subject B, the impact of the electrode selection is less important. Nevertheless, half of the selected electrodes are common between both subjects and the set of pre-defined electrodes. Noteworthily, these two subjects are not representative of the whole population and possess an average P300 response quality. In addition to the high recognition rate of the initial model (CNN-T), artificial neural networks can be considered as tools for analyzing the topology of particular brain activities.

Table 1: Electrode ranking.

Subject	Best electrodes							
	1	2	3	4	5	6	7	8
A	P_Z	PO_7	C_1	PO_Z	C_5	CP_Z	PO_8	C_Z
B	PO_8	O_1	PO_7	C_Z	PO_3	P_Z	CP_Z	PO_4

Table 2: Results of the P300 speller.

Method	Subject	P300 detection		Epoch			
		Recall	Precision	1	5	10	15
CNN-T	A	0.674	0.317	16	61	86	97
CNN-T	B	0.678	0.407	35	79	91	92
CNN-8-FS	A	0.612	0.292	12	48	67	87
CNN-8-FS	B	0.665	0.366	32	72	82	87
CNN-8	A	0.617	0.287	14	46	64	84
CNN-8	B	0.639	0.355	28	71	85	91

Fig. 1: Discriminant power (ξ) for each electrode.

Subject A   Subject B
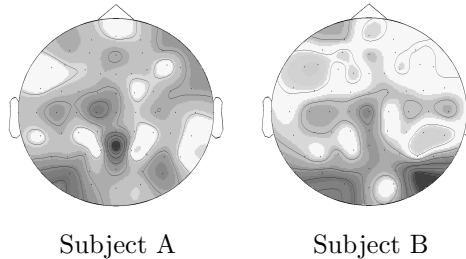
6   Conclusion

A new model has been presented for the detection of P300 waves and its use in a BCI P300 speller. This model was tested on the database 2 on the third BCI competition and provided excellent results (94.5%). Thanks to its particular topology, this network allows further analysis for discovering the best active electrodes in the classification. The weight analysis of the network is consistent to neuroscience knowledge. By prunning the network, it is possible to select a relevant subset of electrodes. This strategy allows a recognition rate of 87% by using 8 electrodes. Further works will deal with the selection of the optimal number of electrodes in relation to a desired accuracy.

Acknowledgment

References

[1] N. Birbaumer and L. G. Cohen. Brain-computer interfaces: communication and restoration of movement in paralysis. *Journal of Physiology-London*, 579(3):621–636, 2007.

[2] H. Cecotti and A. Gräser. Time delay neural network with Fourier Transform for multiple channel detection of steady-state visual evoked potential for brain-computer interfaces. *In Proc. of European Signal Processing Conference*, 2008.

[3] T. Felzer and B. Freisieben. Analyzing EEG signals using the probability estimating guarded neural classifier. *IEEE Trans. on Neural Systems and Rehab. Eng.*, 11(4), 2003.

[4] E. Haselsteiner and G. Pfurtscheller. Using time dependent neural networks for EEG classification. *IEEE Trans. Rehab. Eng.*, 8(4):457–463, 2000.

[5] Nikola Masic and Gert Pfurtscheller. Neural network based classification of single-trial EEG data. *Artificial Intelligence in Medicine*, 5(6):503–513, 1993.

[6] Nikola Masic, Gert Pfurtscheller, and Doris Flotzinger. Neural network-based predictions of hand movements using simulated and real EEG data. *Neurocomputing*, 7(3):259–274, 1995.

[7] B. Blankertz, G. Curio, and Klaus-Robert Müller. Classifying single trial EEG: Towards brain computer interfacing. *In T. G. Diettrich, S. Becker, and Z. Ghahramani, editors, Advances in Neural Inf. Proc. Systems (NIPS 01)*, 14:157–164, 2002.

[8] A. Rakotomamonjy and V. Guigue. BCI competition iii : Dataset ii - ensemble of SVMs for BCI p300 speller. *IEEE Trans. Biomedical Engineering*, 55(3):1147–1154, 2008.

[9] S. Zhong and Joydeep Gosh. HMMs and coupled HMMs for multi-channel EEG classification. *In Proc. IEEE Int. Joint. Conf. on Neural Networks*, 2:1154–1159, 2002.

[10] L. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.*, 70:510–523, 1988.

[11] Yann Le Cun, John S. Denker, and Sara A. Solla. Optimal brain damage. In *Advances in Neural Information Processing Systems*, pages 598–605. Morgan Kaufmann, 1990.

[12] Yann LeCun, Fu-Jie Huang, and Leon Bottou. Learning methods for generic object recognition with invariance to pose and lighting. *Proc. of CVPR'04, IEEE Press*, 2004.

[13] P. Y. Simard, D. Steinkraus, and J. C. Platt. Best practices for convolutional neural networks applied to visual document analysis. *7th International Conference on Document Analysis and Recognition*, pages 958–962, 2003.

[14] Yoshua Bengio and Yann LeCun. Scaling learning algorithms towards AI. *in Bottou, L. and Chapelle, O. and DeCoste, D. and Weston, J. (Eds), Large-Scale Kernel Machines, MIT Press*, 2007.

[15] B. Blankertz, Klaus-Robert Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J.d.e.l. R. Millán, M. Schröder, and N Birbaumer. The BCI competition. iii: Validating alternative approaches to actual BCI problems, 2006.

[16] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.*, 51(6):1034–1043, 2004.

[17] G-E. Sharbrough, R. P. Chatrian, H. Lesser, M. Luders, T. W. Nuwer, and T. W. Picton. American electroencephalographic society guidelines for standard electrode position nomenclature. *J. Clin. Neurophysiol*, 8:200–202, 1991.