On the Selection of Dataset for Online Classifier Training

Cota Navin Gupta, Ramaswamy Palaniappan
School of Computer Science and Electronics Engineering, University of Essex, Colchester, CO4 3SQ, UK
E-mail: ngcota@essex.ac.uk

Very little research has been done on selection of datasets for online brain computer interface (BCI) classifier training, especially participant concentration during data collection. This work presents a study, which highlights the importance of concentration by the participant to achieve good classification accuracies, by considering a simple oddball paradigm.

A four-class BCI oddball paradigm was designed, during which the participant perceived different colour flashes (red, blue, green and black in a randomized way) on white background. A flash of red, blue, green and black was considered a block. Flashtime was set to 100 ms, while ISI was set to 750 ms to prevent perceptual errors, as all participants were naive first time BCI users. Each participant was instructed to focus and keep a count of the cue (target) colour during the first two sessions (S1 and S2). The participant was asked to gaze/relax, thereby refraining from doing the task, during the third session (S3). To prevent habituation, the number of recorded blocks during each session was varied without the knowledge of participants. The number of blocks recorded during the three sessions was 36, 40 and 40 respectively.

Electroencephalogram (EEG) data were collected with a Biosemi ActiveTwo system at a sampling rate of 256 Hz and the experiments were approved by the University of Essex Ethical Committee. Eight optimum channels reported in [1], referenced to the average of mastoids channels were recorded. A forward-reverse band-pass filter with cut off frequencies (1 Hz and 12 Hz) was used to filter data from each channel, to obtain the signals in P300 spectral range, using Matlab commands: *ellip* and *filtfilt*. The designed filter lost no more than 1 dB in the pass band and had at least 40 dB attenuation in stopband. To remove eye-blinks and artifact activity, windsorising as described in [1] was implemented, due to its simplicity and effectiveness.

Two Bayes LDA classifiers with the same parameters as reported in [1] were trained using S1 session (participant concentrates) and S3 session (participant does not concentrate) datasets. The classification accuracies obtained when tested with S2 session are tabulated in Table 1. The results clearly indicate the poor classification accuracies obtained when a dataset where participant is not concentrating is selected for online classifier training.

TABLE 1: Classification accuracies (%) for various combinations of training and testing datasets.

Participant	S1-Training Dataset (Concentrating), S2-Testing Dataset (Concentrating)	S3-Training Dataset, (No Concentration) S2-Testing Data (Concentrating)
1	73.75	27.50
2	48.70	25.00
3	62.50	32.50
4	56.25	30.00

Conclusions

This work justifies the fact that online classification performance of BCI systems depends entirely on the selected training dataset used to train a classifier. Hence it becomes imperative to monitor participant concentration for reliable accuracies. Our future work would therefore revolve around solving this issue by possibly integrating EEG-NIRS (Near Infrared Spectroscopy), to gauge participant concentration.

Acknowledgements

CNG acknowledges support from Overseas Research and University of Essex Scholarships for this work.

References

1. Hoffmann U, Vesin JM, Ebrahimi T and Diserens K: **An efficient P300-based brain-computer interface for disabled subjects.** *J. Neurosci. Methods* 2007 **167(1):**115-125.