 DOI:10.3217/978-3-85125-378-8-10

Multi-state driven decoding for brain-computer interfacing

Minkyu Ahn[1], Hohyun Cho[1], Sangtae Ahn[1] and Sung Chan Jun[1*]
[1] Gwangju Institute of Science and Technology, Gwangju, South Korea
minkyuahn@gmail.com, augustcho@gist.ac.kr, stahn@gist.ac.kr, and scjun@gist.ac.kr

Abstract

Performance variation is one of the critical issues to be resolved in brain-computer interface field. Subjects exhibit different performances from session to session and even across trials. To overcome this issue, three strategies have been commonly proposed, including extraction of robust features, adaptation methods, and monitoring and rejection of bad trials. In this work, we suggest a new multi-classifier strategy using trial data shuffling. This strategy generates different classifiers according to various noise states. Our proposed strategy showed an improvement in classification performance of approximately 3 percent, and a trial-wise quality measure facilitated to monitor bad trials. This seems a promising method to improve the reliability of the BCI system.

1 Introduction

Brain-Computer Interface (BCI) suffers from performance variation. Within subjects, the non-stationary properties of brain signals are considered to be a major cause of this problem. Existing strategies can be categorized as follows: 1) extraction of robust features (Cho et al., 2012), 2) adaptation of feature or classifier (Krusienski et al., 2011; Shenoy et al., 2006; Satti et al., 2010), and 3) monitoring and rejecting bad trials (Ferrez and del R Millan, 2008). Brain signals yield both meaningful (which is used to decode a user's intention or thought) and non-relevant information (background noise). We focused primarily on the first type of information under the assumption that meaningful information is the same over trials. Therefore, most studies have assumed that features are extracted well and reflect task-related information. In practice, however, noise varies over time; this variation influences the overall signal property, which degrades the performance of the feature extractor. Thus, noise information should be considered carefully. Moreover, a user may change the way s/he conducts a mental task or may not be in a proper mental state. Therefore, these trials may violate the assumption above, thus calling into question the use of the methods in Strategies 1 and 2. Grosse-Wentrup and Schölkopf (2012) demonstrated that the prediction of poor mental state, in which a user is not likely to generate a motor imagery-related signal, is possible before the user begins the imagination task. It is remarkable that BCI can predict a user's condition in advance; therefore, the system enables us to deal more readily with the potentially unreliable state of a user. In this study, we propose a new multi-state driven (MSD) method that deals with the problem of fluctuations in BCI. The concept of this MSD strategy is described in Section 2. Materials, evaluation and results are presented in the subsequent sections. Finally, we discuss related ideas and our future work.

*Corresponding author, scjun@gist.ac.kr: Research supported by National Research Foundation of Korea (Grant No. NRF-2013R1A1A2009029).

2 Methods and materials

2.1 Multi-State Driven method

Typically, every trial obtained in the calibration phase is used to construct one classifier. However, this is likely to overlook two points. First, why is the entire trial dataset used simultaneously in order to construct a classifier? In general, a classifier depends heavily on the given dataset. Thus, performance will vary over the datasets used. Second, why is only one classifier used? As reported in (Lotte et al., 2007), a multi-classifier method gives better and more stable results. The use of one classifier is risky, as its performance may vary greatly over trials. To reduce this risk, we may separate one dataset into multiple sub-datasets containing a small number of trials and use those to generate multiple classifiers (one per sub-dataset). Finally, a decoder is constructed with a reasonable combination of these classifiers. Given trial dataset *T*, it is assumed that each trial of dataset *T* consists of the sum of *s* (task-involved information) and noise *n*. With just several trials, a classifier can be constructed through a function associated with preprocessing, and a variety of filtering and feature selections. Therefore, Strategy 1 focuses on extracting the signal *s* only, while Strategy 2 considers $s + n$ simultaneously. If noise variability over trials is high and sub-dataset $D \subset T$ has its own unique noise state $N\ (= \sum_i n_i)$, these will influence classifier construction and classification is likely to work poorly for trials not contained in *D*. We defined this unique noise property as state *N*. This noise state *N* definitely varies over sub-datasets and some sub-datasets may contain noise, as well as information unrelated to the task that is caused by a user. Thus, we expect that such a variable noise state may become involved in the construction of various classifiers; we call this idea a MSD strategy. Any type of pre-processing and feature selection can be applicable to this MSD strategy. Many ideas may exist with respect to generating a decoder that considers outputs from these multiple classifiers. In this study, a "voting" method (Lotte et al., 2007) was introduced and a single value was used as a quality score for each trial. By introducing MSD, multiple outputs (one output per classifier) coming from many classifiers are generated when a trial is given as input. These multiple outputs may be used to evaluate the measure of quality for a given input trial. The number of class labels classified is counted from multiple outputs and the probability for each trial is estimated. This probability shows how well each trial is discriminated and thus we used this trial discriminant score (DS) to monitor bad trials in order to confirm the improvement in accuracy.

2.2 Motor imagery experiment

The following experiment was approved by the Institutional Review Board at Gwangju Institute of Science and Technology. All subjects were informed of the experimental process and purpose before the experiment and signed letters of consent were collected from every subject. With four subjects, we recorded electroencephalographs (EEG) through a Biosemi active2 (64 channels, sampling rate: 512 Hz). Three sessions were conducted with each subject and each session consisted of offline and online experiments. In the offline experiments, a trial began with a blank screen; the instruction bar appeared on the left, right or bottom side of the screen after 2 sec. Subjects were instructed to imagine movement of a body part for 2 sec when a ball was presented. After another 2 sec, the ball moved in the instructed direction. 1st and 2nd runs consisted of 30 trials, as these two runs were used to identify the best pair of motor imagery (i.e., left and right hand, and foot movement imagination). At the end of the 2nd run, we estimated cross-validated accuracy from 60 trials for three conditions through Common Spatial Pattern (CSP) and Fisher Linear Discriminant Analysis (FLDA). Here we applied bandpass filtering (8-30 Hz) and the temporal interval was selected manually after examining event-related (de)synchronization patterns. For the 3rd and 4th runs, subjects conducted twenty trials per condition. Therefore, this yielded forty more trials per condition. Finally, we collected sixty trials for each condition in a two class motor imagery experiment. From these trials, we constructed a classifier

using invariant Common Spatio-Spectral Pattern (iCSSP) and FLDA, as in (Cho et al., 2012). This classifier was applied in the online experiment. Subjects conducted 150 trials for two conditions (75 trials for each) over 3 runs. A feedback trial similar to the trials in the offline experiment was designed. However, in the feedback trials, we gave a classified result and moved the ball in that direction.

2.3 Evaluation

The purpose of this study was to evaluate our proposed strategy and compare it to the conventional strategy that uses whole trials for one classifier. For this evaluation, we fixed informative intervals as 8-30 Hz and 0.4-2.4 sec. for spectral and temporal filtering (Ahn et al., 2012). The evaluation was conducted through 'in-phase' and 'phase to phase' performances. To estimate performance in the calibration/feedback phase, we first applied cross-validation to check in-phase performance. Whole trials were divided into 10 groups and 7 groups were chosen as a training set, while the remaining 3 groups were used as a testing set. CSP and FLDA were applied, such that 120 iterations yielded 120 accuracy estimates; we assigned the mean accuracy to conventional (CV) performance. We also estimated MSD accuracy. In 10 groups of trials, we chose nine groups as training data and the remaining group was used as testing. From 9 groups in the training set, 7 groups were selected to generate a classifier; therefore, we were able to construct an MSD decoder consisting of 36 classifiers. This decoder evaluated each trial in the test data, and the process was repeated through 10-fold validation; thereafter, a trial-wise discriminant value was estimated and we calculated MSD accuracy. Finally, we obtained both CV and MSD accuracy for each calibration and feedback phase. Conventional phase-to-phase performance is designed so that a classifier constructed by calibration data evaluates the trials in the feedback phase. We applied this conventional method for our calibration and feedback data through CSP and FLDA. In addition, MSD was implemented and applied. The MSD decoder constructed from calibration data was applied to discriminate the trials in the feedback phase.

MSD produces trial-wise discriminability from classifiers, which facilitates the evaluation of trial quality. It is possible to reject a bad trial that falls beneath a certain threshold. This was evaluated for phase to phase performance. The MSD decoder constructed using calibration data evaluated trials in the feedback phase and bad trials were rejected. For this evaluation, we used a 50% threshold, which means that the MSD decoder identified a certain class that received at least more than one evaluation. However, we applied different criteria to examine how the accuracy changed. Based on this criterion, we accepted trials showing higher probability than the threshold and the hit rate was calculated.

Table 1: Performance (%) comparison between conventional approach (CV) and MSD

Datasets	**In-phase (Calibration)**		**In-phase (Feedback)**		**Phase to phase**	
	CV	***MSD***	***CV***	***MSD***	***CV***	***MSD***
A1	85.0	87.5	72.9	80.7	75.3	78.7
A2	92.8	95.8	58.9	62.0	75.3	75.3
A3	86.6	90.8	65.8	71.4	70.7	75.7
B1	80.6	84.2	94.1	95.3	70.0	71.3
B2	88.5	89.2	80.5	83.3	80.7	80.7
B3	80.9	85.0	84.8	87.3	69.3	75.3
C1	89.4	91.7	76.5	79.3	77.3	81.3
C2	81.8	86.7	62.9	66.0	58.0	61.3
C3	60.2	60.0	51.6	51.3	51.3	56.0
D1	60.8	61.7	61.7	62.0	50.0	51.3
D2	75.3	78.3	87.1	87.3	72.0	76.0
D3	85.7	87.5	94.7	94.0	92.0	93.3
Mean	80.6	83.2	74.3	76.7	70.2	73.0
Standard Deviation	10.5	11.3	14.3	14.0	12.1	11.7
p value (*:<.01, **:<.005, ***:<.001)	0.00097656 (***)		0.0068359 (*)		0.0019531 (**)	

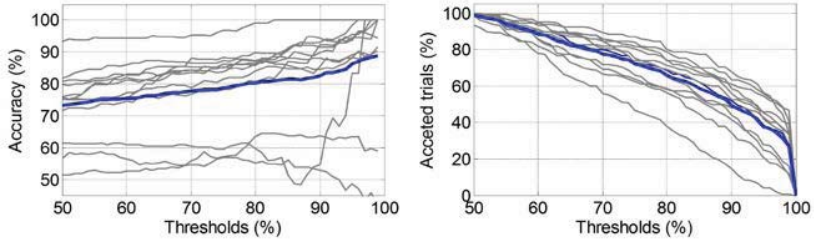

Figure 1: Accuracy (left) and accepted trials (right) in % over varying threshold levels (50% to 100%). Each thin, gray line denotes the results for each dataset and the thick, blue line indicates the mean.

3 Results and conclusions

As shown in Table 1, MSD resulted in higher performance than CV accuracy showing the improvement 2.6% (calibration), and 2.4% (feedback). Wilcoxon signed rank tests revealed the significance at $p < 0.001$ and $p < 0.01$, respectively. In phase to phase performance, the MSD was also superior. The mean accuracy increased from 70.17% to 73.03% ($p < 0.005$). In the test of the applicability of the trial rejection method, we observed that, for most datasets, the accuracy increased as the criterion became stronger (Figure 1). Meanwhile, the number of trials accepted decreased steadily. This demonstrated that most datasets, except for C2, C3 and D1, generated reasonably good task-related information. Thus, trial rejection might be a more efficient technique to apply. In this study, we proposed a new strategy to make the BCI system more reliable. Our results showed notable improvement in performance. To overcome performance variation, further investigations of trial-wise and session-wise data are required. Our future studies will investigate this issue.

References

Ahn, M., Hong, J., & Jun, S. (2012). Feasibility of approaches combining sensor and source features in brain–computer interface. *Journal of Neuroscience Methods, 204*(1), 168 - 178.

Cho, H., Ahn, M., Sangtae, A., & Jun, S. (2012). Invariant Common Spatio-Spectral Patterns. *3rd TOBI Workshop*, (pp. 31-32).

Ferrez, P., & del R Millan, J. (2008). Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE transactions on bio-medical engineering, 55*(3), 923-929.

Grosse-Wentrup, M., & Schölkopf, B. (2012). High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces. *Journal of Neural engineering, 9*(4), 046001.

Krusienski D. J., Grosse-Wentrup M., Galan F., Coyle D., Miller K. J., Forney E., Anderson C. W. (2011) Critical issues in state-of-the-art brain-computer interface signal processing. *Journal of Neural engineering, 8*(2), 025002.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, R1.

Satti, A., Guan, C., Coyle, D., Prasad, G. (2010) A Covariate Shift Minimisation Method to Alleviate Non-stationarity Effects for an Adaptive Brain-Computer Interface. *20th International Conference on Pattern Recognition*, (pp. 105–108).

Shenoy, P., Krauledat, M., Blankertz, B., Rao, R., & Müller, K.-R. (2006). Towards adaptive classification for BCI. *Journal of neural engineering*, R13-23.