ORIGINAL PAPER

Towards a Cure for BCI Illiteracy

Carmen Vidaurre · Benjamin Blankertz

Abstract Brain–Computer Interfaces (BCIs) allow a user to control a computer application by brain activity as acquired, e.g., by EEG. One of the biggest challenges in BCI research is to understand and solve the problem of "BCI Illiteracy", which is that BCI control does not work for a non-negligible portion of users (estimated 15 to 30%). Here, we investigate the illiteracy problem in BCI systems which are based on the modulation of sensorimotor rhythms. In this paper, a sophisticated adaptation scheme is presented which guides the user from an initial subject-independent classifier that operates on simple features to a subject-optimized state-of-the-art classifier within one session while the user interacts the whole time with the same feedback application. While initial runs use supervised adaptation methods for robust co-adaptive learning of user and machine, final runs use unsupervised adaptation and therefore provide an unbiased measure of BCI performance. Using this approach, which does not involve any offline calibration measurement, good performance was obtained by good BCI participants (also one novice) after 3–6 min of adaptation. More importantly, the use of machine learning techniques allowed users who were unable to achieve successful feedback before to gain

C. Vidaurre (✉) · B. Blankertz
Machine Learning Dp, Berlin Institute of Technology,
Franklinstr. 28/29, 10587 Berlin, Germany
e-mail: vidcar@cs.tu-berlin.de

B. Blankertz
e-mail: blanker@cs.tu-berlin.de

B. Blankertz
IDA Group, Fraunhofer FIRST, Kekulestr. 7,
12489 Berlin, Germany

significant control over the BCI system. In particular, one participant had no peak of the sensory motor idle rhythm in the beginning of the experiment, but could develop such peak during the course of the session (and use voluntary modulation of its amplitude to control the feedback application).

Introduction

Brain–Computer Interfaces (BCIs) systems aim to provide users control over a computer application by their brain activity (see Dornhege et al. 2007; Kübler et al. 2001; Millán et al. 2004; Pfurtscheller et al. 2005; Wolpaw et al. 2002). In EEG-based BCIs, one of the biggest research challenges is to understand and solve the problem of "BCI Illiteracy", which is that BCI control does not work for a non-negligible portion of users (estimated 15 to 30%), (c.f. Dickhaus et al. 2009). In a screening study, $N = 80$ participants performed motor imagery first in a calibration (i.e., without feedback) measurement and then in a feedback measurement in which they could control a 1D cursor application. Coarsely, we observed three categories of users: participants for whom (I) a classifier could be successfully trained and who performed feedback with good accuracy; (II) a classifier could be successfully trained, but feedback did not work well. It is known that there are changes between the calibration and the feedback step that can affect the EEG signals, making the feedback fail. In the study with 80 users, the bias of the classifier was supervisedly updated using the first 20 feedback trials (as in Shenoy et al. 2006), but this strategy revealed not to be sufficient for some of the participants; (III) no classifier with

acceptable accuracy could be trained. Whereas participants of Cat. II had obviously difficulties with the transition from offline on online operation, users of Cat. III did not show the expected modulation of sensorimotor rhythms (SMRs): either no SMR idle rhythm was observed over motor areas, or this idle rhythm was not attenuated during motor imagery. Here we present preliminary results of a one-session pilot study in which it was investigated, whether co-adaptive learning using machine-learning techniques could help users of Cat. II and III to achieve successful feedback. Our results show that adaptive machine learning methods successfully helped participants who suffered from the BCI illiteracy problem before, to gain control of the system.

Materials and Methods

Experimental Setup

The study consisted of a one-day session that immediately started with BCI feedback using a pre-trained subject-independent classifier, as in Vidaurre et al. (2007). Using supervised and unsupervised techniques, the classifier was adapted to the specific brain signals of the experimental user during the session. Adaptation was performed during three levels. While the feedback application itself stayed the same for the whole experiment, the features on which the classifier operated and the adaptation methods changed from level to level as described below.

Methods

Eleven participants took part in the study. Six of them belonged to Cat. I (for one novice user, no prior data was available, but she turned out to be a Cat. I user), two further participants belonged to Cat. II and three to Cat. III. All users performed 8 feedback runs, each of them consisting of 100 trials (50 trials of each class). The timing of the trials was as follows: at time 0, the cue was provided in the form of a small arrow over a cross placed in the middle of the screen, one second later, the cross started to move to provide feedback. Its speed was determined by the classification output (similar to Blankertz et al. (2007, 2008a)). The task of the participant was to use motor imagery to make the cursor move into a previously indicated target direction. The feedback lasted for 3 s and was followed by a short pause. Two different types of motor imagery, chosen out of three possibilities (motor imagery of left hand, right hand or foot) were selected in advance. For seven users, previous data with motor imagery performance was available which revealed which two motor imagery tasks should be used. For the other four participants (three of Cat. III and one novice) no prior

information could be used and they were asked to select two out of the three possible motor imagery tasks. Throughout the whole session, all classifiers were based on Linear Discriminant Analysis (LDA). When advisable due to high dimensionality of features, the estimation of the covariance matrix that is needed for LDA was corrected by shrinkage (Ledoit and Wolf 2004; Vidaurre et al. 2009). In order to define the adaptation schemes for LDA we use a specific variant that is introduced here. For LDA the covariance matrices of both classes are assumed to be equal (assumption of linear separability) and it will be denote by Σ here. Furthermore we denote the means of the two classes by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, an arbitrary feature vector by $\boldsymbol{x}$ and define:

$$D(\boldsymbol{x}) = [b; \boldsymbol{w}]^\top \cdot [1; \boldsymbol{x}] \tag{1}$$

$$\boldsymbol{w} = \Sigma^{-1} \cdot (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \tag{2}$$

$$b = -\boldsymbol{w}^\top \cdot \boldsymbol{\mu} \tag{3}$$

$$\boldsymbol{\mu} = \frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2}{2} \tag{4}$$

where $D(\boldsymbol{x})$ is the difference in the distance of the feature vector $\boldsymbol{x}$ to the separating hyperplane, which is described by its normal vector $\boldsymbol{w}$ and bias b. Note that the covariance matrices and mean values used in this paper are sample covariance matrices and sample means, estimated from the data. In order to simplify the notation and the description of the methods, we will in the following use covariance matrix instead of sample covariance matrix and mean instead of sample mean. Usually, the covariance matrix used in Eq. 2 is the class-average covariance matrix. But it can be shown that using the pooled covariance matrix (which can be estimated without using label information, just by aggregating the features of all classes) yields the same separating hyperplane. In this study we used the pooled covariance matrix in Eq. 2. Similarly, the class-average mean (calculated in Eq. 4) can be replaced by the pooled mean (average over all feature vectors of all classes). This implies that the bias of the separating hyperplane can be estimated (and adapted) in an unsupervised manner (without label information). The restriction of the method is to have an estimate of the prior probabilities of the 2 classes. If LDA is to be used as a classifier, observation $\boldsymbol{x}$ is classified as class 1, if $D(\boldsymbol{x})$ is less than 0, and otherwise as class 2. But in the cursor control application we use the classifier output $D(\boldsymbol{x})$ as real number to determine the speed of the cursor. Finally, we introduce the features and classifiers that have been used in the three levels of the experiment, including three on-line adaptation schemes: the first two are supervised, i.e., they require information about the class label (type of motor imagery task) of the past trial in order to update the classifier. The last method

updates the classifier without knowing the task of the past trial (unsupervised adaptation).

Methods for Level 1 (runs 1–3)

The first run started with a pre-trained subject-independent classifier on simple features: band-power in alpha (8–15 Hz) and beta (16–32 Hz) frequency range in three Laplacian channels at C3, Cz, C4. During these runs, the LDA classifier was adapted to the user after each trial. The inverse of the pooled covariance matrix (see Eq. 2) was updated for observation $x(t)$ using a recursive-least-square algorithm, (see Vidaurre et al. 2006 for more information):

$$\Sigma(t)^{-1} = \frac{1}{1-UC}\left(\Sigma(t-1)^{-1} - \frac{v(t) \cdot v^{\top}(t)}{\frac{1-UC}{UC} + x^{\top}(t) \cdot v(t)}\right) \tag{5}$$

where $v(t) = \Sigma^{-1}(t-1) \cdot x(t)$. Note, the term $x^{\top}(t) \cdot v(t)$ is a scalar and no costly matrix inversion is needed. In Eq. 5, UC stands for update coefficient and is a small number between 0 and 1. For the present study, we chose UC = 0.015 based on a simulation using the data of the screening study. To estimate the class-specific adaptive mean $\mu_1(t)$ and $\mu_2(t)$ one can use an exponential moving average:

$$\mu_i(t) = (1-UC) \cdot \mu_i(t-1) + UC \cdot x(t) \tag{6}$$

where i is the class of $x(t)$ and UC was chosen to be 0.05. Note that the class-mean estimation is done in a supervised manner.

Methods for Level 2 (runs 4–6)

For the subsequent 3 runs, a classifier was trained on a more complex composed band-power feature. On the data of run 1–3, a subject-specific narrow band was chosen automatically (Blankertz et al. 2008b). For this frequency band, optimized spatial filters have been determined by Common Spatial Pattern (CSP) analysis (Blankertz et al. 2008b). Furthermore, six Laplacian channels have been selected according to their discriminability, which was quantified by a robust variant of the Fisher score (mean replaced by median). The selection of the positions was constraint such that two positions have been selected from each of the areas over left hand, right hand and foot. While CSP filters were static, the position of the Laplacians was reselected based on the Fisher score of the channels. Channel selection and classifier were recalculated after each trial using the last 100 trials. The classifier used here was regularized version of LDC, with automatic shrinkage, to account for the higher dimensionality of the features, as

in Vidaurre et al. (2009). The feature vector was the concatenation of log band-power in the CSP channels and the selected Laplacians channels. The addition of the repeatedly selected Laplacian channels was included in order to provide flexibility with respect to spatial location of modulated brain activity. During these three runs the adaptation to the user was done again in a supervised way.

Methods for Level 3 (runs 7–8)

Finally for the last 2 runs, CSP filters have been calculated on the data of runs 4–6 and a classifier was trained on the resulting log band-power features. The bias of the classifier in Eq. 3 was adapted by updating the pooled mean μ after each trial with UC = 0.05. The update rule for the pooled mean was analogue to Eq. 6, but without distinction by class labels. Note that this adaptation scheme is unsupervised. For more information about unsupervised methods, see Vidaurre et al. (2008).

Results

As a verification of the novel experimental design, we first discuss the results for the six participants of Cat. I. Here, very good feedback performance was obtained within the first run after 20 to 40 trials (i.e., after 3–6 min) of adaptation and hit rates increased further in runs 2 and 3 and stayed on that level in subsequent runs. This can be seen in Fig. 1, where the grand average of feedback performance within each run is displayed, according to the Category of the participants. Note that all runs of one volunteer have been recorded within one session. The challenge of the experiment was the performance with the two participants of Cat. II and three users of Cat. III. All those five participants did not have control in the first three runs, but they were able to gain it when the machine learning based techniques came into play in runs 4–6: in the average performance for Cat. II a sudden jump occurred from run 3 to run 4, and for Cat. III a continuous increase in runs –6. According to Kübler et al. (2004) an accuracy of 70% is assumed to be a threshold required for BCI applications related to communication, such as cursor control.

Conclusion

Machine Learning based BCIs use EEG features of larger complexity that can be fitted better to the individual characteristics of brain patterns of each user (see Blankertz et al. 2007, 2008b; Dornhege et al. 2004, 2007; Müller et al. 2003, 2008). The down side of this approach is the need for
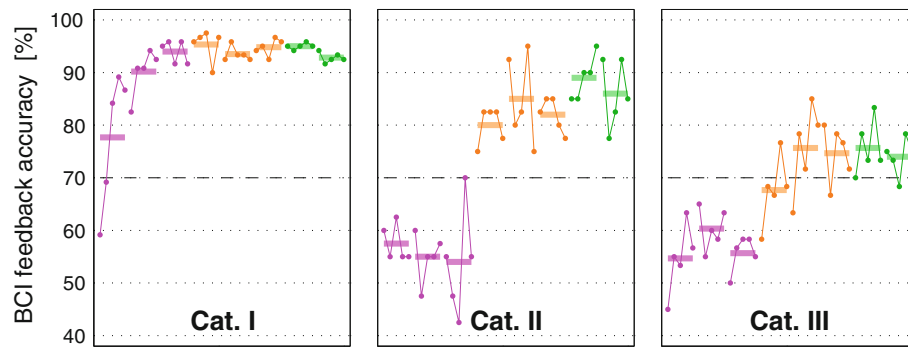
Fig. 1 Grand average of feedback performance within each run (horizontal bars and dots for each group of 20 trials) for participants of Cat. I (*N* = 6), Cat. II (*N* = 2) and Cat. III (*N* = 3). An accuracy of 70% is assumed to be a threshold required for BCI applications related to communication such as cursor control. Note that all runs of one volunteer have been recorded within one session

an initial offline calibration. Furthermore, users are in a different mental state during offline calibration than during online feedback (c.f. Shenoy et al. 2006), which renders the classifier that is optimized on the data of the calibration suboptimal and sometimes even non-functional for feedback (see Sugiyama et al. 2007; von Bünau et al. 2009) for a discussion of non-stationarities in BCI). Moreover, some users have difficulties to properly perform motor imagery for calibration due to the lack of feedback. Here, we have presented a novel method for Machine Learning based brain–computer interfacing which overcomes these problems. It replaces the offline calibration by a 'coadaptive calibration', in which the mental strategy of the user and the algorithm of the BCI system are jointly optimized. This approach leads some users very quickly (3–6 mins) to accurate BCI control. Other users, who could not gain BCI control in the classic Machine Learning approach (i.e., belonging to Cat. II or II), could gain BCI control within one session, see Fig. 1. In particular, one participant who had no peak of the SMR idle rhythm in the beginning of the measurement could develop such with our adaptive feedback training, (Vidaurre et al. in prep) This important finding gives rise to the development of neurofeedback training procedures that might help to cure BCI illiteracy. Further studies with a larger number of participants will be required in order to confirm these initial findings.

References

Blankertz B, Dornhege G, Krauledat M, Müller KR, Curio G (2007) The non-invasive Berlin Brain–Computer Interface: fast acquisition of effective performance in untrained subjects. NeuroImage 37(2):539–550

Blankertz B, Losch F, Krauledat M, Dornhege G, Curio G, Müller KR (2008a) The Berlin Brain–Computer Interface: accurate performance from first-session in BCI-naive subjects. IEEE Trans Biomed Eng 55(10):2452–2462

Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR (2008b) Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process Mag 25(1):41–56

Blankertz B, Tangermann M, Vidaurre C, Dickhaus T, Sannelli C, Popescu F, Fazli S, Danóczy M, Curio G, Müller KR (2009) The Berlin Brain–Computer Interface. In: Allison B, Graimann B, Pfurtscheller G (eds) Non-invasive and invasive brain–computer interfaces. Springer, The Frontiers Collection (in press)

Dickhaus T, Sannelli C, Müller KR, Curio G, Blankertz B (2009) Predicting BCI performance to study BCI illiteracy. BMC Neurosci 10(Suppl 1): P84

Dornhege G, Blankertz B, Curio G, Müller KR (2004) Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multi-class paradigms. IEEE Trans Biomed Eng 51(6):993–1002

Dornhege G, Millán J del R, Hinterberger T, McFarland D, Müller KR (eds) (2007) Toward brain–computer interfacing. MIT Press, Cambridge, MA

Kübler A, Kotchoubey B, Kaiser J, Wolpaw J, Birbaumer N (2001) Brain–computer communication: unlocking the locked in. Psychol Bull 127(3):358–375

Kübler A, Neumann N, Wilhelm B, Hinterberger T, Birbaumer N (2004) Predictability of brain–computer communication. J Psychophysiol 18:121–129

Ledoit O, Wolf M (2004) Honey, I shrunk the sample covariance matrix. J Portfolio Manag 30:110–119

Millán J del R, Renkens F, Mouriño J, Gerstner W (2004) Brain-actuated interaction. Artif Intell 159:241–259

Müller KR, Anderson CW, Birch GE (2003) Linear and non-linear methods for brain–computer interfaces. IEEE Trans Neural Sys Rehab Eng 11(2):165–169

Müller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B (2008) Machine learning for real-time single-trial analysis: from brain–computer interfacing mental state monitoring. J Neurosci Methods 167:82–90

Pfurtscheller G, Neuper C, Birbaumer N (2005) Human brain–computer interface. In: Riehle A, Vaadia E (eds) Motor cortex in voluntary movements, Chap. 14. CRC Press, New York, pp 367–401

Shenoy P, Krauledat M, Blankertz B, Rao RPN, Müller KR (2006) Towards adaptive classification for BCI. J Neural Eng 3(1): R13–R23

Sugiyama M, Krauledat M, Müller KR (2007) Covariate shift adaptation by importance weighted cross validation. J Mach Learn Res 8:1027–1061

Vidaurre C, Schlögl A, Scherer R, Cabeza R, Pfurtscheller G (2006) A fully on-line adaptive BCI. IEEE Trans on Biomed Eng 53:1214–1219

Vidaurre C, Schlögl A, Scherer R, Cabeza R, Pfurtscheller G (2007) Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. IEEE Trans Biomed Eng 54:550–556

Vidaurre C, Schlögl A, Blankertz B, Kawanabe M, Müller KR (2008) Unsupervised adaptation of the LDA classifier for brain–computer interfaces. In: Proceedings of the 4th international brain–computer interface workshop and training course 2008, Verlag der Technischen Universität Graz, pp 122–127

Vidaurre C, Krämer N, Blankertz B, Schlögl A (2009) Time domain parameters as a feature for EEG-based brain–computer interfaces. Neural Netw 22:1313–1319

Vidaurre C, Sannelli C, Müller KR, Blankertz B (in prep) Machine learning co-adaptive learning: towards a cure for BCI illiteracy

von Bünau P, Meinecke FC, Kiraly FC, Müller KR (2009) Finding stationary subspaces in multivariate time series. Phys Rev Lett 103(21):214101

Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain–computer interfaces for communication and control. Clin Neurophysiol 113:767–791