A Framework for Online Inter-subjects Classification in Endogenous Brain-Computer Interfaces

Sami Dalhoumi, Gérard Dray, Jacky Montmain, Stéphane Perrey

A Framework for Online Inter-subjects Classification in Endogenous Brain-Computer Interfaces

Sami Dalhoumi[1(✉)], Gérard Dray[1], Jacky Montmain[1], and Stéphane Perrey[2]

[1] Laboratoire d'Informatique et d'Ingénierie de Production (LGI2P), Ecole des Mines d'Alès, Parc Scientifique G. Besse, 30035 Nîmes, France
{sami.dalhoumi, gerard.dray, jacky.montmain}@mines-ales.fr

[2] Movement to Health (M2H), Montpellier University, Euromov, 700 Avenue du Pic Saint-Loup, 34090 Montpellier, France
stephane.perrey@univ-montpl.fr

Abstract. Inter-subjects classification and online adaptation techniques have been actively explored in the brain-computer interfaces (BCIs) research community during the last years. However, few works tried to conceive classification models that take advantage of both techniques. In this paper we propose an online inter-subjects classification framework for endogenous BCIs. Inter-subjects classification is performed using a weighted average ensemble in which base classifiers are learned using data recorded from different subjects and weighted according to their accuracies in classifying brain signals of current BCI user. Online adaptation is performed by updating base classifiers' weights in a semi-supervised way based on ensemble predictions reinforced by interaction error-related potentials (iErrPs). The effectiveness of our approach is demonstrated using two electroencephalography (EEG) data sets and a previously proposed procedure for simulating interaction error potentials.

Keywords: Brain-computer interfaces · Inter-subjects classification · Online adaptation · Weighted average ensembles

1 Introduction

A brain-computer interface (BCI) is a communication and control technology that allows translating brain's electrical or hemodynamic activity patterns into commands for an external device [1]. This technology was originally meant to allow patients with severe neuromuscular disabilities to autonomously interact with their environment. Depending on the modality of interaction, BCIs can be classified as either exogenous or endogenous [2]. Exogenous BCIs rely on brain activity patterns that are elicited spontaneously in response to external stimuli such as visual evoked potentials (VEPs), while endogenous BCIs are based on the voluntary induction of different brain states by the user such as sensorimotor rhythms-based BCIs. Endogenous BCIs can offer a natural way of interaction for the user but they are difficult to set-up because

self-regulation of brain rhythms is not a straightforward task. For this reason, a long calibration phase is needful for user-system co-adaptation before every use of the BCI [3]. During this phase, the user interacts with the BCI in a cue-based mode which allows him to learn self-regulating his brain rhythms and the system to create a "robust" classification model. The accuracy of the system depends on the capacity of the classification model to decode brain activity patterns of the user during a feedback phase (self-paced interaction mode).

In order to bring endogenous BCIs out of the lab, many research groups have focused on conceiving new machine learning approaches that allow reducing calibration time without decreasing classification accuracy of the system. Among these approaches, inter-subjects classification has been actively explored during the last years [3–5]. It consists of incorporating labeled data recorded from different BCI users in the learning process of current BCI user. When performed correctly, inter-subjects classification allows capturing information that generalize across users and extend to new users. One way to do that is to use a weighted average ensemble technique in which base classifiers are learned using data from different BCI users and weighted according to their accuracy in classifying signals recorded from current BCI user [4]. In the absence of true class labels during feedback phase of current BCI user, these weights can be estimated in two ways: statically using a small calibration set or dynamically by recalculating these weights for each incoming sample based on its position in the feature space [4, 5]. The first approach may perform poorly because brain activity patterns of the BCI user vary between calibration and online phases and during online phase (non-stationarity). The second approach may not perform well because it does not take into consideration the stochastic dependence between time-contingent feature vectors.

In a preliminary work [6], we found that a static classifiers weighting approach using a small calibration set outperforms dynamic classifiers weighting approaches and we showed that online adaptation of base classifiers' weights using ensemble predictions during feedback phase may increase classification accuracy in comparison to the static approach. However, the update coefficient used for adjusting adaptation speed was subject-dependent which presented a limitation to the proposed approach. In this work, we propose to use interaction error-related potentials (iErrPs) as an additional source of information to reduce uncertainty about ensemble predictions during feedback phase. iErrPs are a type of event-related potentials that occur immediately after the user perceives that the feedback provided by the BCI is in contradiction to his intent [7]. The physiological background of iErrPs has been well established [7] and they have been successfully used to improve BCIs accuracy [8, 9]. However, iErrPs are also subject to a degree of uncertainty because their detection is not perfect. Ferrez and Del R. Millán [7] reported an average recognition rate of iErrPs with 16.5 % of false positives (i.e., the correct predictions are considered erroneous) and 20.8 % of false negatives (i.e., the erroneous predictions are considered correct).

This paper is organized as follows: in Sect. 2, we describe our adaptive weighted average ensemble method. In Sect. 3, we present material used to evaluate this method and the experimental results. Section 4 concludes the paper and gives future directions of this work.

2 Methods

In this section, we describe different steps of our adaptive weighted average ensemble method for binary classification tasks. Base classifiers trained on brain signals recorded from different subjects are weighted according to their accuracies in classifying a small calibration set from current BCI user. These weights are updated during feedback phase using ensemble predictions reinforced by iErrPs. In the absence of an iErrP, the label predicted by the ensemble is considered correct and base classifiers' weights are updated based on their disagreement with the ensemble. When an iErrP is detected, the prediction of the ensemble is considered to be wrong and base classifiers' weights are updated using the opposite label.

2.1 Base Classifiers' Weights Initialization

Let $\{h^1, h^2, \ldots, h^K\}$ be K classification models learned using data from different BCI users (many classification models may be learned using data recorded from the same user and preprocessed in different ways). For each incoming feature vector x and each class label y, the classifier $h^k, k = 1 \ldots K$ outputs the value $h^k_y(x) \in [01]$ which is an estimation of the posterior probability $p(y/x)$. Given a small calibration set $L = \{(x_t, y_t), x_t \in \mathbb{R}^d, y_t \in \{0, 1\}, t = 1 \ldots T\}$ recorded from current BCI user, each classifier is assigned a weight w^k inversely proportional to its error in classifying this labeled set:

$$w^k = \max\left(0, MSE^r - MSE^k\right), k = 1..K \tag{1}$$

where, MSE^r is the mean squared error of a random classifier and MSE^k is the mean squared error of the classifier h^k given below.

$$MSE^r = \sum_y p(y).(1 - p(y))^2 \tag{2}$$

$$MSE^k = \frac{1}{T}.\sum_{t=1}^{T}\left(1 - h^k_{y_t}(x_t)\right)^2 \tag{3}$$

For binary classification with equal class priors, $MSE^r = 0.25$.

This weighting scheme allows removing classifiers performing less or equal than a random classifier from the ensemble and assigning weights to the rest of classifiers inversely proportional to their error in classifying calibration data.

2.2 Base Classifiers' Weights Adaptation Using Ensemble Predictions

Given a new labeled sample (x_{t+1}, y_{t+1}), the mean squared errors of base classifiers up to the time step $(t + 1)$ can be updated in the following way:

$$MSE^k(t+1) = \frac{1}{t+1}.\left[t.MSE^k(t) + \left(1 - h^k_{y_{t+1}}(x_{t+1})\right)^2\right], k = 1...K \tag{4}$$

where, $MSE^k(t)$ is the mean squared error of the classifier h^k up to the time step t.

Base classifiers' weights can then be updated using the adaptive version of Eq. (1):

$$w^k(t+1) = \max\left(0, MSE^r - MSE^k(t+1)\right), k = 1..K \tag{5}$$

In order to take into consideration different types of data shift, we add an update coefficient $UC \in [01]$ to Eq. (4) that becomes:

$$\begin{aligned} MSE^k(t+1) &= \frac{1}{(1-UC).t + UC} \times \\ &\left[(1-UC).t.MSE^k(t) + UC.\left(1 - h^k_{y_{t+1}}(x_{t+1})\right)^2\right], k = 1...K \end{aligned} \tag{6}$$

For $UC = 0$, there is no update, for $UC = 1$, only the new data sample is used for calculating error and when $UC = 0.5$, we retrieve exactly the update Eq. (4).

In self-paced interaction mode, the true class labels are unknown for the classification model. One way to alleviate this problem is to use ensemble predictions for online adaptation. For each incoming feature vector, the label predicted by the ensemble is considered to be the true class label and each base classifier's weight is updated according to its disagreement with the ensemble. So, formula (6) becomes:

$$\begin{aligned} MSE^k(t+1) &= \frac{1}{(1-UC).t + UC} \times \\ &\left[(1-UC).t.MSE^k(t) + UC.\left(1 - h^k_{\tilde{y}_{t+1}}(x_{t+1})\right)^2\right], k = 1...K \end{aligned} \tag{7}$$

Where $\tilde{y}_{t+1}$ is the label predicted by the ensemble:

$$\tilde{y}_{t+1} = argmax_y \left(\sum_{k=1}^{K} w^k(t).h^k_y(x_{t+1})\right) \tag{8}$$

As ensemble's decisions are subject to a high degree of uncertainty, using them for adaptation may lead to error accumulation and consequently degrades the accuracy of the BCI. Thus, we should use and additional information to minimize uncertainty. In BCIs, such information could come from interaction error-related potentials.

2.3 Base Classifiers' Weights Adaptation Using Ensemble Predictions Reinforced by Interaction Error-Related Potentials

Let $E \in \{0, 1\}$ be the true absence or presence of an iErrP following the output of the BCI. $E = 0$, when the decision of the ensemble $\tilde{y}_{t+1}$ corresponds to the intent of the

user y_{t+1} and $E = 1$, in the opposite case. The iErrPs classifier outputs a value $\tilde{E} \in \{0, 1\}$ which is a prediction of E. The predicted value $\tilde{E}$ may or may not correspond to the real value E depending on the accuracy of the iErrPs classifier. This iErrPs classifier can be used to assess the reliability of the predicted labels as follows:

$$MSE^k(t+1) = \frac{1}{(1-UC).t + UC} \times \left[(1-UC).t.MSE^k(t) + UC.\left((1-\tilde{E}) - h^k_{\tilde{y}_{t+1}}(x_{t+1})\right)^2\right], k = 1 \ldots K \tag{9}$$

When $\tilde{E} = 0$, the predicted label is considered correct and the update is the same as in Eq. (7). When $\tilde{E} = 1$, the opposite class label is used for update because $\left(h^k_{\tilde{y}_{t+1}}(x_{t+1})\right)^2 = \left(1 - h^k_{(1-\tilde{y}_{t+1})}(x_{t+1})\right)^2$.

3 Experiments

In this section we evaluate our adaptive ensemble approach using two EEG data sets and the procedure for simulating iErrPs used in [8, 9].

3.1 EEG Data Sets

Data set 2A in BCI Competition IV. This data set comprises electroencephalography (EEG) signals recorded from 9 subjects using 22 Ag/AgCl electrodes at 250 Hz sampling rate [10]. Subjects performed left hand, right hand, foot and tongue motor imagery tasks. For the purpose of this study, only EEG signals corresponding to the left hand and right hand motor imagery tasks were used. For each subject, two sessions on different days, each of which comprises 72 trials of duration 7 s, were collected. At the beginning of each trial, a fixating point appeared on a computer screen. After two seconds, a cue appeared informing the subject which motor imagery task to perform until the cue disappeared.

EEG measurements were band-pass filtered using a 5^{th} order Butterworth filter in the frequency bands of 4 Hz width ranging from 8 Hz to 30 Hz with step size of 2 Hz and an additional wide band from 8 Hz to 30 Hz. Time segments 3–5 s after the beginning of each trial were extracted. The common spatial pattern (CSP) algorithm and the logarithmic variance features are used for spatial filtering and feature extraction (the three most discriminative CSP filters for each class are used) [11].

Two Class Motor Imagery Data Set from BNCI Horizon 2020 Project. This data set was provided by the Graz group [12]. 14 subjects performed sustained kinesthetic motor imagery of the right hand and feet. 5 subjects had previously performed BCI experiments and 9 subjects were naïve to the task. Each subject performed a training phase composed of 50 trials per class and a validation phase composed of 30 trials per class. EEG signals were recorded using 15 Ag/AgCl electrodes at 512 Hz sampling

rate. Time segments of length 3 s, starting at 3 s after the beginning of each trial were preprocessed in the same way as in the previous data set. CSP algorithm and logarithmic variance features were used to extract relevant features from this data set.

3.2 Procedure for Simulating IErrPs

Llera et al. [8] proposed a simple procedure for simulating iErrPs that allows understanding the relation between the accuracy of the iErrPs classifier and the accuracy of the task classifier. Below we describe it in case of our adaptive ensemble method.

Let α_1 and α_2 be the false positive and false negative rates of the iErrPs classifier, respectively. Given the output of the ensemble classifier $\tilde{y}_t$ and the true class label y_t at time step t, the procedure is performed as follows:

- If $\tilde{y}_t = y_t$, we draw $\tilde{E} = 1$ with probability α_1 and $\tilde{E} = 0$ with probability $1 - \alpha_1$ and apply Eqs. (5) and (9).
- If $\tilde{y}_t \neq y_t$, we draw $\tilde{E} = 1$ with probability $1 - \alpha_2$ and $\tilde{E} = 0$ with probability α_2 and apply Eqs. (5) and (9).

3.3 Results

Evaluation was performed offline using leave-one-subject-out cross-validation. In each step, training data from N-1 subjects (from now called source subjects) were used for learning spatial filters and base classifiers, the calibration set extracted from training data of the N^{th} subject (from now called target subject) was used to initialize base classifiers' weights and test data of the same subject was used for evaluation (N = 9 in the first data set and N = 14 in the second data set). During training phase, CSP filters and corresponding linear discriminant analysis (LDA) classifiers are learned using EEG signals recorded from each subject and filtered in different frequency bands, resulting in 88 base classifiers in the first data set and 143 base classifiers in the second data set. Calibration set of the target subject is filtered in different frequency bands and projected into the subspaces spanned by the previously learned CSP filter banks. The initial mean-squared error of each base classifier is calculated using the corresponding projection. For evaluation, each trial in the test set of target subject is filtered in different frequency bands and projected into the subspaces spanned by the CSP filter banks.

Figure 1 illustrates the average classification accuracies of a static accuracy-weighted ensemble (AWE) learned using data from source subjects and a baseline LDA classifier learned using only calibration data of target subject filtered in the 8–30 Hz frequency band (traditional approach) for the first data set. As we can see, learning from other users allows increasing classification accuracy when the size of calibration set is small because the subject-independent information captured from large data sets is more robust than subject-specific information learned from a small data set. As the size of calibration set increases, the accuracy of the baseline classifier increases while the accuracy of the inter-subject classification model remains relatively constant. This shows that, in contrary to the traditional classification approach, the performance of the inter-subjects classification approach is not much dependent on the size of calibration set.

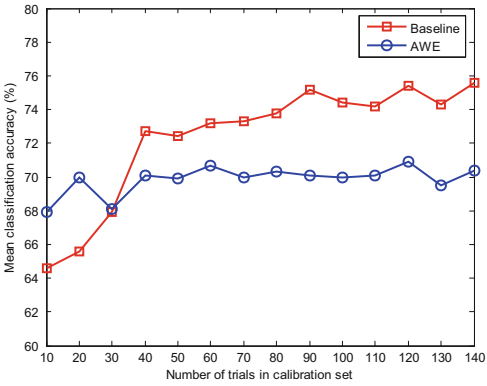

Fig. 1. Average classification accuracy of the standard classification approach (baseline) and the static inter-subjects classification approach (AWE) for different sizes of calibration set in the first data set

In order to assess whether online adaptation of base classifiers' weights allows increasing performance of our inter-subjects classification approach, we performed a comparison between the static accuracy-weighted ensemble and the adaptive accuracy-weighted ensemble. To do so, we evaluated three scenarios for online adaptation of base classifiers' weights:

- Guided: adaptation is performed using only ensemble predictions.
- Realistic iErrPs detection: adaptation is performed using ensemble predictions reinforced by an iErrPs classifier with false positive rate α_1 of 16.5 % and false negative rate α_2 of 20.8 % as found in [7].
- Perfect iErrPs detection: adaptation is performed using ensemble's predictions reinforced by a perfect iErrPs classifier ($\alpha_1 = \alpha_2 = 0$).

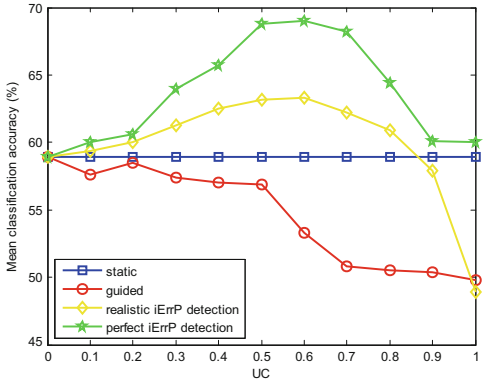

Fig. 2. Average classification accuracies of the static weighted average ensemble and different scenarios of the adaptive weighted average ensemble when the size of calibration set is equal to 10 trials in the second data set

Figure 2 illustrates the comparative results for the second data set when the size of calibration set is equal to 10 trials. The x-axis corresponds to different values of the update coefficient UC and the y-axis to the average classification accuracy over all subjects. The results of the adaptive ensemble method using realistic iErrPs classifier are the average over 100 tests for each subject and each value of UC. This figure shows that using iErrPs for assessing the reliability of the ensemble predictions allows preventing error accumulation and increases classification accuracy especially for values of the update coefficient between 0.5 and 0.7.

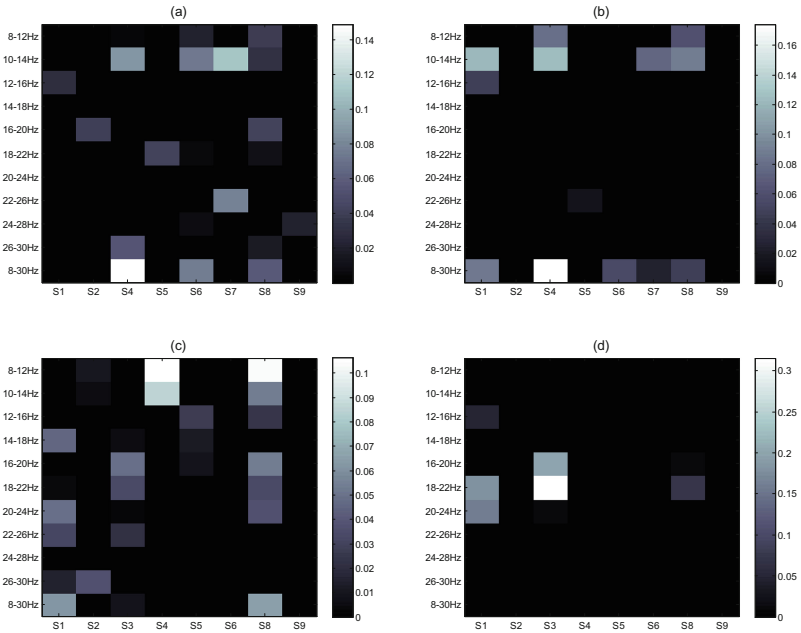

Fig. 3. The evolution of base classifiers' weights during the test session for two different subjects in data set 1. (a) and (b) correspond to base classifiers' weights at the beginning and the end of test session of subject 3. (c) and (d) correspond to base classifiers' weights at the beginning and the end of test session of subject 7

We performed the same comparison for the first data set with predefined update coefficient $UC = 0.5$. Table 1 shows the accuracies of the static ensemble method and the three adaptive methods for different subjects. For most of subjects, the adaptive ensemble method using a realistic iErrPs classifier allows increasing classification accuracy compared to the static method which shows its applicability in online settings.

For further investigation of the behavior of our adaptive ensemble method, Fig. 3 shows an illustration of the evolution of base classifiers' weights between the beginning and the end of test session for two different cases in data set 1. Figure 3(a) and (b) show the normalized weights of the base classifiers for subject 3 at the beginning and the end of test session, respectively. For this subject, the base classifier learned using EEG

Table 1. Classification accuracy of the static weighted average ensemble and different scenarios of the adaptive weighted average ensemble when the size of calibration set is equal to 10 trials in the first data set and the update coefficient *UC* is equal to 0.5

	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean	Std.
Static	81.2	51.4	95.8	65.0	48.6	58.3	64.6	89.6	56.3	67.9	17.0
Guided	75.7	50.0	95.8	49.7	50.7	50.7	50.0	90.3	68.1	64.5	18.7
Realistic iErrP	81.9	60.2	94.0	55.5	51.7	54.7	73.7	90.2	71.2	70.3	15.9
Perfect iErrP	82.6	59.7	95.8	60.8	52.7	57.6	83.3	90.3	75.7	73.2	15.8

signals recorded from subject 4 and filtered in the 8–30 Hz frequency band maintained the highest weight during all the test set ("robust" classifier) which is reflected in the classification accuracy of the static weighted-average ensemble that is equal to the accuracy of the adaptive ensemble using a perfect iErrPs classifier. Oppositely, both adaptive ensemble method using realistic iErrPs classifier and adaptive ensemble method using perfect iErrPs classifier significantly increased classification accuracy for subject 7 in comparison to the static ensemble which is related to the huge change of base classifiers' weights between the beginning of the test session (Fig. 3(c)) and the end of it (Fig. 3(d)).

4 Conclusion

In this paper we presented an online inter-subjects classification framework for endogenous brain-computer interfaces. A straightforward way to learn from heterogeneous data recorded from different subjects is to use a weighted average ensemble in which each base classifier is trained using a single data set and weighted according to its accuracy in classifying brain signals of current BCI user. Static weighting of base classifiers using a small calibration set may increase classification accuracy in comparison to standard methods but this approach is limited by the non-stationary nature of brain signals. In the absence of true class labels during feedback phase, we proposed a new online adaptation approach of base classifiers' weights based on ensemble predictions reinforced by interaction error-related potentials (iErrPs). Results on two EEG data sets showed that our adaptive ensemble method based on a realistic iErrPs classifier allows increasing classification accuracy in comparison to the static method and preventing error accumulation compared to the adaptive method based only on ensemble predictions.

The proposed online adaptation method was limited to binary classification tasks. In future work we will extend it to multi-class classification and evaluate it in online experimental settings. Beyond the scope of BCI applications, our approach can be extended to other applications in which online transfer learning is needful and information about user's assessment of the system is accessible such as spam filtering application.

References

1. McFarland, D.J., Wolpaw, J.R.: Brain-computer interfaces for communication and control. Commun. ACM **54**(5), 60–66 (2011)
2. Nicolas-Alonso, L.F., Gomez-Gil, J.: Brain computer interfaces, a review. Sensors **12**, 1211–1279 (2012)
3. Lotte, F., Guan, C.: Learning from other subjects helps reducing brain-computer interface calibration time. In: International Conference on Audio Speech and Signal Processing (ICASSP), pp. 614–617 (2010)
4. Tu, W., Sun, S.: A subject transfer framework for EEG classification. Neurocomputing **82**, 109–116 (2011)
5. Liyanage, S.R., Guan, C., Zhan, H., Ang, K.K., Xu, J., Lee, T.H.: Dynamically weighted ensemble classification for non-stationary EEG processing. J. Neural Eng. **10**(3), 036007 (2013)
6. Dalhoumi, S., Dray, G., Montmain, J., Derosière, G., Perrey, S.: An adaptive accuracy-weighted ensemble for inter-subjects classification in brain-computer interfacing. In: 7th International IEEE EMBS Neural Engineering Conference (2015)
7. Ferrez, P.W., Del R Millan, J.: Error-related EEG potentials generated during simulated brain–computer interaction. IEEE Trans. Biomed. Eng. **55**(3), 923–929 (2008)
8. Llera, A., Van Gerven, M.A.J., Gomez, V., Jensen, O., Kappen, H.J.: On the use of interaction error potentials for adaptive brain computer interfaces. Neural Netw. **24**(10), 1120–1127 (2011)
9. Zeyl, T.J., Chau, T.: A case study of linear classifiers adapted using imperfect labels derived from human event-related potentials. Pattern Recogn. Lett. **37**, 54–62 (2014)
10. Blankertz, B.: BCI Competition IV. http://www.bbci.de/competition/iv (2008)
11. Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.R.: Optimizing spatial filters for robust EEG single-trial analysis. IEEE Sig. Process. Mag. **25**(1), 41–56 (2008)
12. Steyrl, D., Scherer, R., Forstner, O., Muller-Putz, G.R.: Motor imagery brain-computer interfaces: random forests vs regularized LDA – non-linear beats linear. In: 6th International Brain-Computer Interface Conference (2014)