Neural Networks 24 (2011) 1120–1127

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

On the use of interaction error potentials for adaptive brain computer interfaces

A. Llera [a,b,*], M.A.J. van Gerven [a,b,c], V. Gómez [a,b], O. Jensen [a,b], H.J. Kappen [a,b]

[a] Radboud University Nijmegen, The Netherlands

[b] Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

[c] Institute for Computing and Information Sciences (ICIS), Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 19 July 2010
Received in revised form 28 January 2011
Accepted 22 May 2011

Keywords:
Brain computer interfaces
Error potential
Adaptive classification

ABSTRACT

We propose an adaptive classification method for the Brain Computer Interfaces (BCI) which uses Interaction Error Potentials (IErrPs) as a reinforcement signal and adapts the classifier parameters when an error is detected. We analyze the quality of the proposed approach in relation to the misclassification of the IErrPs. In addition we compare static versus adaptive classification performance using artificial and MEG data. We show that the proposed adaptive framework significantly improves the static classification methods.

1. Introduction

The interest in Brain Computer Interfaces has quickly grown in the last few years. The possibility to provide disabled people with new communication channels, such as BCI spellers, new mobility channels such as BCI driven wheel chairs or BCI controlled mechanical prostheses makes this a very attractive research field. However, the applicability of current BCI systems is still limited because of a number of problems. One of these problems is the presence of non-stationarities in the data (Shenoy, Krauledat, Blankertz, Rao, & Müller, 2006). This causes patterns associated with each task during the training of the BCI to be different during testing, leading to a poor performance.

Several approaches have been proposed to overcome this problem by the introduction of adaptive classification methods (Sykacek, Roberts, & Stokes, 2004; Shenoy et al., 2006; Pfurtscheller, Neuper, Schölgl, & Lugger, 1998). In Shenoy et al. (2006), it is shown how the probability distributions associated with class features change between training and test sessions, and assuming that the labels of new incoming trials are known, it is shown that proper updates in the classifier parameters would improve the performance of the original static classifier. Note that in the BCI setting we normally do not know the user intention, so the labels of the trials are unknown. We propose the use of neural feedback to detect incorrect performance of the device, and to be able to recover the labels in the case of a binary classification task. The on-line detection of the wrong performance of a BCI has been addressed before by means of Interaction Error Potentials (IErrP) (Ferrez & Millán, 2008, 2005; Seno, Matteucci, & Mainardi, 2010).

Error-related potentials are potentials detected in the recorded electroencephalogram (EEG) of a subject just after an error occurs. The error is the difference between the expected and the actual result of an action. Error-related potentials have been studied in many different scenarios since the late 1980s (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1990; Miltner, Braun, & Coles, 1997; van Schie, Mars, Coles, & Bekkering, 2004; Ferrez, 2007). It is well known that the presence of an error is usually followed by what are called event-related negativity and positivity which are present in the alpha band in the fronto-central channels. More recently, a study using Magnetoencephalography (MEG) (Mazaheri, Nieuwenhuis, van Dijk, & Jensen, 2009) has shown that an erroneous reaction to stimuli is followed by an increase in the frontal theta and a decrease in the posterior alpha and central beta powers.

Based on the nature of the feedback, the error-related potentials can be categorized as response error potentials (Falkenstein et al., 1990; Mazaheri et al., 2009), feedback error potentials (Miltner et al., 1997), observation error potentials (van Schie et al., 2004) and the most interesting for us, interaction error potentials (IErrP) which are present when a device delivers an erroneous feedback (Ferrez, 2007).

Since the IErrP are present in the recorded EEG of a subject controlling a device just after the device returns an unexpected feedback (the BCI makes a classification error) (Ferrez & Millán, 2008), its detection can help to construct a more robust BCI, either by correcting the BCI output directly (Ferrez, 2007), or more interestingly, by adapting the BCI classifier to prevent similar mistakes in the future. This idea is illustrated in Fig. 1.

* Corresponding author at: Radboud University Nijmegen, The Netherlands.
E-mail address: a.llera@donders.ru.nl (A. Llera).

0893-6080/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2011.05.006

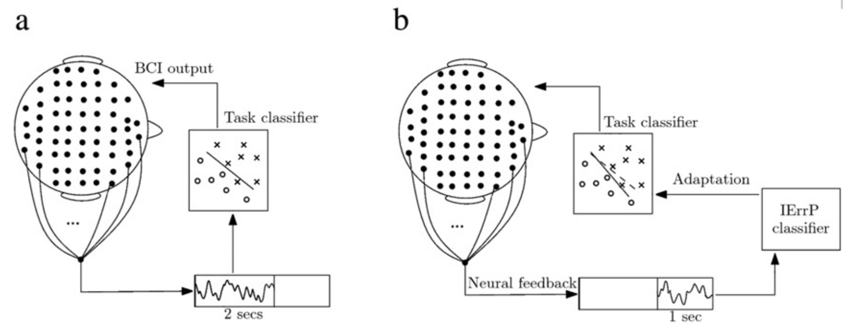

Fig. 1. (a) Classical scenario: brain activity is measured during a period of time, then the *task classifier* decides the class label given the measured activity and the device produces an output. (b) Proposed scenario: after task classification, an *IErrP classifier* uses feedback from the user (subsequent data from zero to one seconds after (a)) to compare previous user intention with the device output. If an IErrP is detected, the parameters of the task classifier are updated.

Although the possibility of BCI adaptation using error feedback from the user has been previously proposed (Chavarriaga, Ferrez, & Millán, 2007), the impact of using an IErrP classifier to improve the original task classifier has not been studied before in a realistic BCI setting. In this article, we explore this idea in detail. In Section 2 we introduce a framework based on reinforcement learning for adaptive BCI using an IErrP classifier as a control signal. We analyze the effect of IErrP misclassification in terms of false positives and false negatives, and we measure how the performance of the task classifier is affected. In Section 3, we first perform the single trial classification of IErrP and then we apply the proposed adaptive method to MEG data collected during a binary forced-choice covert attention task.

2. Adaptive BCI classifier

In this section we introduce the proposed method to design a binary adaptive BCI. We consider a binary task with an adaptive task classifier that learns from the output of a static IErrP classifier in order to minimize erroneous feedback.

2.1. Adaptive learning rule

Consider the (unobserved) subject's intention, left or right, that we denote as target class $t \in \{0, 1\}$ respectively. The generated brain activity is measured and a vector of feature values $\mathbf{x} := (x_1, \ldots, x_n)$ is extracted which is relevant to discriminate between both classes. We use the logistic regression model (Bishop, 2007) which takes the form:

$$p(t = 1|\mathbf{x}, \mathbf{w}) = \sigma(\mathbf{x}, \mathbf{w}) = \frac{1}{1 + e^{-\sum_{i=0}^{n} w_i x_i}}, \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^{n+1}$ is the vector of weights, and $x_0 = 1$ accounts for the bias term.
The error in the prediction is quantified as the log-likelihood of the data:

$$G(\mathbf{x}, \mathbf{w}, t) = -(t \ln \sigma(\mathbf{x}, \mathbf{w}) + (1 - t) \ln(1 - \sigma(\mathbf{x}, \mathbf{w}))). \tag{2}$$

The output of the task classifier is defined as

$$\tilde{t} = \chi\left(p(t = 1|\mathbf{x}, \mathbf{w}) > \frac{1}{2}\right), \tag{3}$$

where χ returns 1 if its argument is true and 0 otherwise. An adaptive learning rule for the parameters $\mathbf{w}$ updates $\mathbf{w}$ in the direction of the gradient of (2):

$$\Delta w_i = \eta \frac{\partial G(\mathbf{x}, \mathbf{w}, t)}{\partial w_i} = \eta(t - \sigma(\mathbf{x}, \mathbf{w}))x_i, \tag{4}$$

where η denotes the learning rate.

In a realistic BCI setting however, the intention of the user t is unknown. We define $E \in \{0, 1\}$ as the user's true absence or presence of surprise following the output of the device. Thus $E = 0$ corresponds to $\tilde{t} = t$ and $E = 1$ to $\tilde{t} \neq t$. After the output of the task classifier ($\tilde{t}$) is delivered, subsequent brain activity (neural feedback) is measured and a feature vector $\mathbf{y} := (y_1, \ldots, y_m)$ is extracted and used by the IErrP classifier to provide an estimation of E, which we denote by $\tilde{E} \in \{0, 1\}$. Updates occur only when a surprise is detected ($\tilde{E} = 1$), in which case the observed output $\tilde{t}$ is presumably incorrect, so $t = 1 - \tilde{t}$ and the learning rule (4) for the task classifier becomes

$$\Delta w_i = \eta \tilde{E}(1 - \tilde{t} - \sigma(\mathbf{x}, \mathbf{w}))x_i, \tag{5}$$

where $1 - \tilde{t}$ is the opposite label from the output of the task classifier.

The performance of this model clearly depends on the flexibility of the model to adapt to changes at the correct time scale (Heskes & Kappen, 1991, 1992), but also on the asymptotic behavior of the task classifier in relation to the misclassification of IErrP. In Sections 2.2 and 2.3 we study this relation.

2.2. Effect of IErrP misclassification

The performance of a BCI system based on the previous framework clearly depends on the accuracy of the IErrP classifier. Previous researchers have reported classification rates of IErrP of around 80%, as well as the stability on IErrP detection across sessions (Ferrez & Millán, 2008). The misclassification of IErrPs can occur in two ways (see Fig. 2):

False positives. Correctly classified trials ($\tilde{t} = t$) are considered to be erroneous, causing an update of the task classifier parameters with the wrong class label. We characterize the rate of false positives with α_1.

False negatives. Erroneously classified trials ($\tilde{t} \neq t$) are considered as correct. As a consequence, the task classifier parameters will not be updated when it is desirable. We characterize the rate of false negatives with α_2.

Note that the effect of false positives results in learning from incorrectly labeled data, whereas false negatives result in discarding potentially useful learning samples.

2.3. Simulations

In order to better understand the asymptotic behavior of the task classifier in relation to the accuracy of the IErrP classifier (α_1 and α_2), we consider an artificial binary class classification problem in a one-dimensional feature space. For each class $t \in \{0, 1\}$, the feature is distributed according to a Normal distribution