Pattern Recognition Letters 37 (2014) 54–62

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

A case study of linear classifiers adapted using imperfect labels derived from human event-related potentials

Timothy J. Zeyl, Tom Chau *

Institute of Biomaterials and Biomedical Engineering, University of Toronto, 27 King's College Circle, Toronto, Ontario M5S 1A1, Canada
Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, 150 Kilgour Road, Toronto, Ontario M4G 1R8, Canada

ARTICLE INFO

Article history:
Available online 31 May 2013

Keywords:
Partially supervised learning
Imperfect labels
Adaptive classification
EEG
Binary classification
Event-related potentials

ABSTRACT

Electroencephalogram signals used to control brain–computer interfaces (BCIs) are nonstationary, a problem that makes classification of mental tasks difficult in real-time. Event-related potentials associated with BCI errors have the potential to be used as online labels for adaptation of BCI classifiers; however, detection of event-related potentials is imperfect, which makes this a partially supervised classification problem. In this study, two linear binary classifiers are adapted using uncertain labels on artificial data sets representing various scenarios of concept drift as well as on a real motor imagery BCI data set. Both perfectly and imperfectly simulated labels are incorporated into the classifiers which are adapted in the following two ways: (i) only after trials where BCI mistakes were detected and (ii) after every trial regardless of whether or not an error was detected. We find that all data sets benefit from adaptation using imperfect labels and that adapting after all trials results in better performance than adapting only after detected errors, especially when the labels are imperfect and the classes are inseparable.

1. Introduction

Brain computer interfaces (BCIs) analyze and translate physiological signals that arise from brain activity into operational control of a computer or communication device. The sensing modalities used to monitor brain activity in BCIs can include electroencephalography (EEG), magnetoencephalography, near-infrared spectroscopy and transcranial doppler ultrasound (Wolpaw et al., 2002; Power et al., 2011; Myrden et al., 2011). The signals recorded by these modalities exhibit nonstationarities that make it difficult to classify brain activity consistently across sessions (Power, 2012), between mental tasks, and throughout a task (Kipiński et al., 2011).

Classification of nonstationary data is difficult because the class labels corresponding to those data may be changing over time, a scenario referred to as concept drift (Kolter and Maloof, 2007). Drifting concepts (or classes) may arise from changes in the distribution from which the classes are drawn. More formally, concept drift can be viewed as a time-varying posterior probability of the output class given the data in a Bayesian sense; that is, $P_{t+1}(y|x) \neq P_t(y|x)$ where $P_t(y|x) = P_t(x|y)P_t(y)/P_t(x)$ and t represents time (Elwell and Polikar, 2011). Thus, concept drift may be due to time variations in any of the distributions that give rise to the posterior probability $P_t(y|x)$; this includes nonstationarities in the data features $P_t(x)$, the prior class probability $P_t(y)$, and the likelihood of the data given that it was generated from a particular class $P_t(x|y)$. It is difficult to make robust classifications if any of these distributions are changing in time, a problem that is exacerbated when the distribution that is time-variant cannot be identified. Thus, adaptive classification has been proposed as a solution for concept drift.

Adaptive classification allows a classifier to adjust its parameters to concept drift without the need for explicitly modeling the changes in data distribution. Adaptation can thus improve performance with respect to a static classifier (i.e., one that does not adjust its decision parameters) with small computational cost by considering data points obtained online. Adaptive classification enables continuous use of a BCI with fewer re-calibration sessions, reduced training times and improved performance (e.g. Vidaurre et al., 2011b).

Adaptive classifiers can be categorized according to the following three levels of supervision: they can be supervised, in which the class labels of the data become known after classification; they can be unsupervised, in which the labels of online classifications are unknown; or they can be partially supervised, in which the labels obtained online may be vague, uncertain, or incomplete. A number of groups have investigated supervised adaptation for

* Corresponding author at: Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, 150 Kilgour Road, Toronto, Ontario, Canada M4G 1R8. Tel.: +1 416 425 6220x3515.

E-mail addresses: tim.zeyl@utoronto.ca (T.J. Zeyl), tom.chau@utoronto.ca, tchau@hollandbloorview.ca (T. Chau).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.05.020

use in BCIs (e.g., Shenoy et al., 2006; Sugiyama et al., 2007), but these systems require class information to account for class dependent concept drift (Krauledat, 2008). Such class information is not readily available in the context of a BCI as it would require the BCI user to explicitly provide class labels for mental tasks. Such task verification may not be possible when the BCI is the only means of communication for that user. This limitation necessitates the use of either unsupervised adaptation or partially supervised adaptation in BCIs.

Unsupervised adaptation does not use explicitly known labels to adjust the classifier parameters. Instead, either the class independent parameters are the only ones adapted to the data (Vidaurre et al., 2011a) or class information is estimated based on previous classifications and training label information (Blumberg et al., 2007; Li and Guan, 2006; Lu et al., 2009). One way to estimate class information is to take previous classifications directly as the true label information, this method is referred to as 'naïve labeling' (Kuncheva et al., 2008; Plumpton et al., 2012). These unsupervised methods have had some success, but they have two potential shortcomings: (i) they overlook class dependent concept drift when only class independent parameters are adapted and (ii) they are prone to accumulating error ('runaway performance') when incorrect classifications are reinforced.

As supervised and unsupervised adaptation may be impossible or inadequate for BCIs, an alternative is to use partial supervision, in which there is a label associated with each data point but the label is subject to a degree of uncertainty. In the context of a BCI, such a label could come in the form of an event-related potential that is elicited in response to feedback that contradicts a user's intention; that is, when the BCI makes a mistake. This type of event-related potential has been termed the interaction error-related potential (iErrP) (Ferrez and Millán, 2008) and has been measured using EEG most prominently at the Fz and Cz electrode locations according to the international 10–20 electrode naming system. The iErrP is a slowly varying potential that occurs within several hundred milliseconds of erroneous BCI feedback presentation and thus its detection could be useful for either correction or labeling of the most recent classification. However, its detection is imperfect as some BCI errors are missed and some correct BCI behaviors are labeled as errors. Nonetheless, the iErrP has been incorporated successfully into P300 speller BCIs (Combaz et al., 2011; Spüler et al., 2012) and has been proposed as a labeling device for adaptive motor imagery BCIs (Blumberg et al., 2007; Llera et al., 2011). In this way, partially supervised adaptation has the potential to account for class-dependent concept drift.

There are several approaches that can be taken when using the iErrP as a labeling device for adaptive classifiers; these approaches adhere to various methods of learning in the presence of concept drift. In drifting environments, one approach is to actively search for concept drift and make some changes to the classifier only when drift is detected (Gama et al., 2004; Baena-García et al., 2006). Within this category, concept drift detectors can be derived from the classification success rate (Ross et al., 2012) in which decreases in accuracy indicate drifting concepts. In this way, the presence of iErrPs can indicate a drifting concept which could trigger adaptation. A second approach for drifting environments is to continually adapt to streaming data based on the label information that is available under the assumption that concepts are continually drifting. Using this approach, the label information from an adaptive BCI classifier could be assumed correct in the absence of iErrPs and could adapt continuously.

In this study, we use the presence of iErrPs as a proxy for concept drift and we compare two methods of adaptation. The first we call *learnNeg*, in which adaptation only takes place when an iErrP is detected; if no iErrP is detected this method assumes the associated classification was correct and there is no concept drift. The second method we call *learnAll*, in which adaptation takes place after every data sample; the absence of an iErrP is taken to mean that the previous classification of the classifier was correct but the concept may still be drifting.

It is important that adaptive classifiers be able to adapt to changing concepts quickly, thus when studying adaptive classifier performance, it is important that the dynamics of the problem be taken into account. We examine the dynamics of two incremental adaptive classifiers in several simulated binary classification problems and ask whether the performance is dependent on the learning protocol chosen (i.e. *learnAll* vs. *learnNeg*). We examine how these learning types are affected by the condition of partial supervision (i.e. imperfect labels generated with iErrPs) and by class separability. We accompany this analysis with an evaluation of the performance of adaptive classifiers on a real nonstationary motor imagery BCI data set.

2. Materials and methods

In this case study, we examine two adaptive classifiers under both the *learnNeg* adaptation protocol and the *learnAll* protocol: the first adaptive classifier is derived from logistic regression and the second is derived from linear discriminant analysis (LDA). The logistic regression classifier invokes a simple linear decision boundary and has been proposed in the adaptation of a BCI task classifier using iErrPs (Llera et al., 2011). Likewise, the LDA classifier has performed well in two-class BCIs (Blankertz et al., 2011) and has been adapted for online BCIs (Blumberg et al., 2007; Vidaurre et al., 2011a). The adaptations made by the classifiers in this study are incremental, meaning that no data points are stored for successive adaptations and the previous parameters are forgotten after updates have been made.

2.1. Logistic regression

Logistic regression minimizes the negative log-likelihood of the classifier parameters (Bishop, 2006), or the cross-entropy,

$$G = -\sum_{n=1}^{N}\left[y^{(n)}\ln\sigma^{(n)} + (1-y^{(n)})\ln(1-\sigma^{(n)})\right] \quad (1)$$

where $$\sigma^{(n)} = \frac{1}{1+\exp(-\mathbf{w}^{\mathsf{T}}\mathbf{x}^{(n)})} \quad (2)$$

and superscript (n) indexes the current sample, $\mathbf{w} = [w_0, w_1, \ldots, w_d]^\top$ are the weights to be learned with bias w_0, $\mathbf{x}^{(n)} = [1, x_1, \ldots, x_d]^\top$ is the $(d+1)$-dimensional feature vector, and $y^{(n)} \in \{0,1\}$ is the true class of the data. The weights can be trained or updated to account for new data samples using stochastic gradient descent, where the gradient at new sample n can be written as

$$\frac{\partial G^{(n)}}{\partial \mathbf{w}} = (y^{(n)} - \sigma^{(n)})\mathbf{x}^{(n)} \quad (3)$$

and the weights are updated as $\mathbf{w}^{(n+1)} = \mathbf{w}^{(n)} + \eta\frac{\partial G^{(n)}}{\partial \mathbf{w}}$, with learning rate $\eta \in \mathbb{R}, \eta > 0$.

Llera et al. (2011) have modified Eq. (3) to be updated only when an iErrP is observed, which corresponds to our definition of a *learnNeg* strategy. To simulate iErrPs, they classify each sequential test sample $\mathbf{x}^{(n)}$ according to

$$\tilde{y}^{(n)} = I[\sigma^{(n)} \geqslant 0.5] \quad (4)$$

where $\tilde{y}^{(n)} \in \{0,1\}$ is the output of the classifier and I is the indicator function.[1] In the simulations that appear in Section 2.3, if the

[1] The indicator function has the following behavior: $I[TRUE] = 1; I[FALSE] = 0$.