Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 123 (2018) 488–493

Procedia Computer Science

www.elsevier.com/locate/procedia

8th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2017

A Greedy Feature Selection Algorithm for Brain-Computer Interface Classification Committees

Alexander G. Trofimov[1], Sergei L. Shishkin[2], Bogdan L. Kozyrskiy[1,2] and Boris M. Velichkovsky[2]

[1]National Research Nuclear University MEPhI, Moscow, Russia

[2]National Research Centre "Kurchatov Institute", Moscow, Russia

atrofimov@list.ru

Abstract

We propose an approach to electroencephalogram feature selection and classification problems in brain-computer interfaces based on a committee of weak classifiers. The design of a classification committee is formulated as an optimization problem and the greedy algorithm for its solving is considered. The proposed approach is applicable when the objects to be classified are characterized by a large number of features while a few train samples are available. Classification performance of the committee was evaluated on real data and improvement over traditional classification methods was observed.

Peer-review under responsibility of the scientific committee of the 8th Annual International Conference on Biologically Inspired Cognitive Architectures

Keywords: electroencephalogram; brain-computer interface; feature selection; classification; committee; greedy algorithm

1 Introduction

Brain-computer interfaces (*BCIs*) are communication systems in which messages sent by the user to the outside world are detected from his or her brain signals and do not pass through the usual output channels such as peripheral nerves and muscles [1–3].

Operation of an electroencephalogram (EEG) based BCI systems in general involves the following steps [1]:

1) Correction of the raw EEG data. This step includes filtering from noise, baseline correction, removal from artifacts and some other preprocessing procedures.

2) Extraction of EEG features. The purpose of this step is to transform the initial space of the EEG data to the multidimensional feature space. Once such a transformation is performed the problem of EEG signal recognition is reduced to well-known problem of multidimensional data classification.

Peer-review under responsibility of the scientific committee of the 8th Annual International Conference on Biologically Inspired Cognitive Architectures

10.1016/j.procs.2018.01.074

3) EEG classification. At this step one of a variety of existing algorithms for data classification, such as discriminant analysis, Bayesian approach, artificial neural networks etc., is applied to the EEG features.

4) Generation of a message or a command to a device based on post-processing of the classification results.

The problem of extraction of the EEG features for classification, i.e. step 2, can be probably considered as the most difficult one. There is no unique and unambiguous way to solve it. In most cases, it is not known a priori which way would be better and whether the data from different classes will be separable well in the feature space proposed. One approach to solve this problem is to try various feature spaces and to choose one where the data are the best separable. The disadvantage of this approach is the need to define a class of functional transformations to be applied to the EEG signals in order to construct the feature vectors. This class of functional transformations must be set by the researcher based on his or her intuition; in the case of poor choice, the classification problem cannot be solved with acceptable accuracy.

The simplest way to construct the feature vector is vectorization of all the observed EEG values (from all channels and all time instances). It can result in high dimensionality of the feature space; in addition, this approach is associated with crucial dependence of the number of EEG channels and sampling frequency [4]. Another popular approach is to use the spectral characteristics of EEG. This approach is justified only when it is known a priori that EEG signals corresponding to different mental states may have significantly different spectra [4,5].

Importantly, prior knowledge about useful EEG features is very inexact. Most of knowledge about EEG components that are sensitive to brain states is obtained in psychophysiological experiments where many trials can be collected and averaged, and statistical inferences are normally made on group level, while BCI classifiers must be adapted to particular individuals and make decisions using a single or at least a few trials. Thus, for constructing a BCI classifier it is important to be able to deal with as many input features as possible. At the same time, BCI classifier must be trained on individual training data, which are typically available in very limited amount, so special attention should be paid to the ability of the BCI algorithms to prevent overfitting.

We propose an approach to solve the EEG classification problem based on a committee of weak classifiers working in the simplest feature spaces. Thus, instead of searching for a suitable functional transformation from the EEG signal space to the feature space it is proposed to construct a committee of simplest classifiers each of which individually may not have the required classification accuracy but in ensemble can amplify each other. This approach is based on the well-known idea of boosting [6–8].

A variety of boosting algorithms was developed to date, in particular AdaBoost, LogitBoost, BrownBoost [9–11] etc. Most of them use weighting of training data and cascading of weak classifiers. At the same time all weak classifiers deal with the same training subsamples, i.e. work in the same feature space.

Unlike the classical boosting algorithms, in the proposed approach each weak classifier performs in its own feature space. Thus, the method of constructing the classification committee can also be seen as a method of feature selection for classification.

The proposed method can be considered as a variation of random subspace ensemble (RSE) method developed by Ho [12], in which the individual classifiers are created only based on randomly selected feature subspaces and final prediction is combined by classifiers' voting. On the other hand, the proposed method is close to the correlation-based feature selection (CFS) algorithm developed by Hall [13] where the best subset of features is selected by taking into consideration two criteria: 1) how good the individual features are at predicting the class and 2) how much they correlate with the other features.