Marcin KOŁODZIEJ, Andrzej MAJKOWSKI, Remigiusz J. RAK

Warsaw University of Technology

Linear discriminant analysis as EEG features reduction technique for brain-computer interfaces

Abstract: *BCI systems analyze the EEG signal and translate patient intentions into simple commands. Signal processing methods are very important in such systems. Signal processing covers: preprocessing, feature extraction, feature selection and classification. In the article authors present the results of implementing linear discriminant analysis as a feature reduction technique for BCI systems.*

Streszczenie: *Systemy BCI analizują sygnał EEG i tłumaczą intencje użytkownika na proste polecenia. Ważnym elementem systemów BCI jest przetwarzanie sygnału. Obejmuje ono: przetwarzanie wstępne, ekstrakcję cech, selekcję cech i klasyfikację. W artykule autorzy prezentują wyniki badań z zastosowaniem liniowej analizy dyskryminacyjnej jako narzędzia do redukcji cech. (**Liniowa analiza dyskryminacyjna jako narzędzie redukcji cech sygnału EEG**)*

Keywords: Linear Discriminant Analysis, LDA, feature reduction, feature selection, brain-computer interface, BCI, EEG
Słowa kluczowe: Liniowa Analiza Dyskryminacyjna, LDA, redukcja cech, selekcja cech, interfejs mózg-komputer, EEG

Introduction

Constructing of an efficient brain-computer interface (BCI) is one of the most challenging scientific problems and focuses scientists attention from all over the world. Often BCI interfaces are based on EEG signals recorded from the surface of the scalp, because this method of brain activity monitoring is noninvasive, easy to use and quite inexpensive. Brain-computer interfaces make use of several brain potentials such as: P300, SSVEP or ERD/ERS [1,2,3]. The most difficult case for implementation is BCI based on brain potentials associated with movements (ERD/ERS). The ERD/ERS name origins from the phenomenon of EEG signal power rise or fall in the frequency bands 8-12 Hz and 18-26 Hz, when a subject spontaneously imagines a movement. In our experiment we tried to classify EEG signals for a single asynchronous trial of imagining movement.

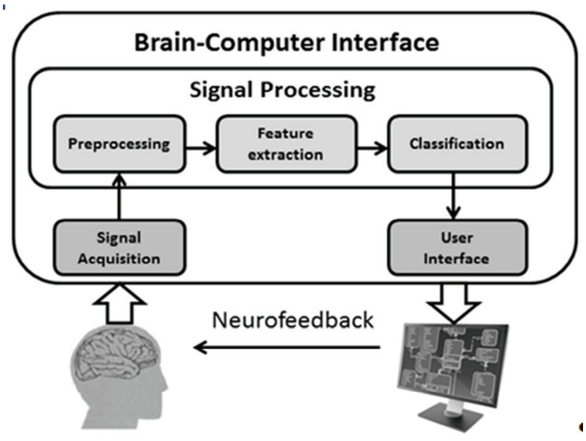

Fig. 1. The scheme of brain-computer interface

The most important part of BCI system is EEG signal processing, which include preprocessing, feature extraction and classification (fig. 1). In our experiment feature extraction based on linear discriminant analysis is taken simultaneously with feature reduction.

Data acquisition

In the experiment we used a dataset of EEG signals provided by IDIAP Research Institute (Silvia Chiappa, José del R. Millán). The set contains data from 3 normal subjects acquired during 3 non-feedback sessions. The subjects were relaxed, sat in normal chairs with arms resting on their legs. Each subject had three tasks to execute: imagination of repetitive self-paced left hand movements, imagination of repetitive self-paced right hand movements, generation of words which started with the same random letter. The subject performed a given task for about 15 seconds and then switched randomly to another task on the operator's request. EEG signals were recorded with a Biosemi system using a cap with 32 electrodes located at standard positions of the conventional 10-20 system. Authors used only 8 selected electrodes: F3, T7, C3, CP1, C4, T8, F4 and Cz. The sampling rate was 512Hz. No artifact rejection or correction was performed. Dataset contains raw EEG signals. Each training file contains additional 33^{rd} component indicating classes to which the particular parts of signal belong.

Feature extraction

There exists many methods of feature extraction. The most widely used for EEG signal are methods based on frequency analysis, for example discrete Fourier transform (DFT) or power spectral density (PSD). We used a method based on DFT. At first the EEG signal is divided into one-second windows overlapping by a half of second. The half-second overlap enables to generate large enough set of data for efficient classifier learning.

In this way we obtained 1280 features from each one-second window [3]. Such a large number of features makes the classification process very difficult. For further analysis frequency band was restricted to 8Hz-30Hz. In this way only 184 features from one-second window were taken into consideration. The scheme of feature extraction procedure is presented in fig. 2.

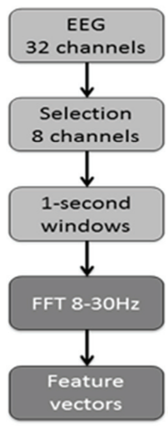

Fig. 2. The scheme of feature extraction

Only some features contain useful information for distinguishing classes. It was found that a linear combination of features can create a new, better feature set of reduced dimension. In the next step authors used linear discriminant analysis to feature reduction.

Feature reduction by linear discriminant analysis

Linear discriminant analysis (LDA) is a well known feature reduction technique [4]. LDA is used to find a linear combination of features that can better separate two or more classes. The LDA finds such direction ***a*** that provide maximum linear separation of classes. An example of a data projection on directions ***a*** and ***b*** is given in fig. 3. There are many possibilities for finding directions but only some are optimal for data discrimination.

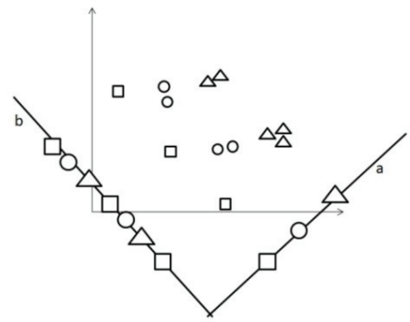

Fig. 3. LDA finds a direction (**a**) that maximize the separation of data

A measure of data separation can be expressed as the maximum of separation coefficient *F* (1):

$$F = \frac{tr(\boldsymbol{S_m})}{tr(\boldsymbol{S_w})} \quad (1)$$

where $\boldsymbol{S_m}$ depicts the between-class scatter, $\boldsymbol{S_w}$ within-class scatter. The bigger the value of *F* (1) the grater probability of classes separation.

Let us assume that we have *C* classes, each containing *N* observations $\boldsymbol{x}_i$. The measure of within-class scatter $\boldsymbol{S^c}$ for the class *c* can be estimated as (2):

$$\boldsymbol{S}^c = \sum_{i=1}^{N}(\boldsymbol{x}_i^c - \boldsymbol{\mu}^c)\,(\boldsymbol{x}_i^c - \boldsymbol{\mu}^c)^T \quad (2)$$

where $\boldsymbol{\mu}^c$ indicates mean of the all observations $\boldsymbol{x}_i$ for *c-th* class. Generalization $\boldsymbol{S_w}$ of the within-class scatter for all *C* classes can be calculated as:

$$\boldsymbol{S}_w = \sum_{i=1}^{C}\frac{n_i}{N}\boldsymbol{S}^i \quad (3)$$

where n_i is the number of $\boldsymbol{x}_i$ observations in each class and *N* is a total number of all observations. The value of between class scatter $\boldsymbol{S}_B^C$ for class c can be calculated as:

$$\boldsymbol{S}_B^C = \sum_{i=1}^{C}(\boldsymbol{\mu^i} - \boldsymbol{\mu})(\boldsymbol{\mu^i} - \boldsymbol{\mu})^T \quad (4)$$

where $\boldsymbol{\mu}^i$ indicates the mean of the all observations $\boldsymbol{x}_i$ for i-th class and $\boldsymbol{\mu}$ indicates the mean of the all observations $\boldsymbol{x}_i$ for all classes. Generalization of between-class scatter $\boldsymbol{S_m}$ *for* all *C* classes can be calculated as:

$$\boldsymbol{S}_m = \sum_{i=1}^{C}\frac{n_i}{N}\boldsymbol{S}_B^i \quad (5)$$

where n_i means the number of $\boldsymbol{x}_i$ observations in each class and N is a total number of all observations.

It can be proved that directions providing the best class separation are eigenvectors with the highest eigenvalues of matrix $\boldsymbol{S}$ [4]:

$$\boldsymbol{S} = \boldsymbol{S}_w^{-1}\boldsymbol{S}_m \quad (6)$$

Generally the matrix $\boldsymbol{S}$ is not a symmetric matrix and calculation of eigenvectors can be difficult. This problem can be solved by using generalized eigenvalue problem [5]. A transformed data set can be obtained by:

$$\boldsymbol{y} = \boldsymbol{x}^T\boldsymbol{W} \quad (7)$$

where $\boldsymbol{W}$=[$\boldsymbol{w_1}$,$\boldsymbol{w_2}$,..,$\boldsymbol{w_M}$] is a matrix build with the *M* eigenvectors of matrix $\boldsymbol{S}$ connected with the highest eigenvalues. LDA reduces the original feature space dimension to *M*. A new data set $\boldsymbol{y}$ is created as a linear combination of all input features $\boldsymbol{x}$ with weights $\boldsymbol{W}$. In authors experiments the total number of features is 184. For further analysis only two LDA components were taken. The data after LDA transformation can be seen in the fig. 4. The data are easily separable. In next step of the experiment data classification was performed.

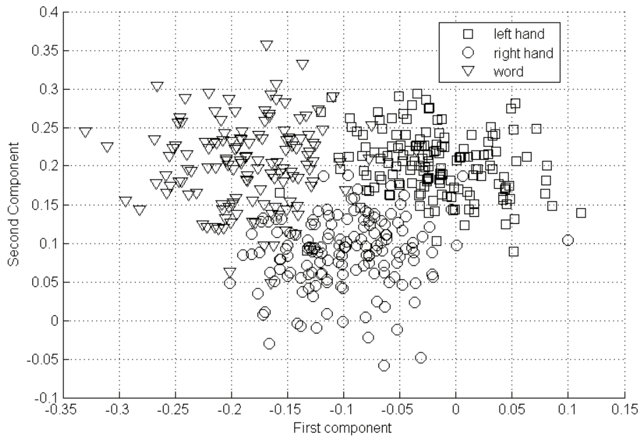

Fig. 4. Data after LDA transform (two components) for the first session of first subject

Classification accuracy

At first for classification authors used K-nearest-neighbor classifier (10-KNN). Because the number of observations was not too large, authors used 10-cross-validation test [4]. Data were divided into 10 parts, nine parts were used to learn classifier and one to test it. Such an operation was repeated 10 times, so each of the ten sets was tested. The total error was calculated as the average of errors for all ten sets. The 3 subjects and 3 session were tested. The results for for 3 different subjects are presented in tables 1, 2, 3.

Table 1. Classification error for the first subject

First subject			
Session 1	Session 2	Session 3	Mean
5.7%	3.3%	5.3%	4.8%

Table 2. Classification error for the second subject

Second subject			
Session 1	Session 2	Session 3	Mean
8.8%	9.1%	9.1%	9%

Table 3. Classification error for the third subject

Third subject			
Session 1	Session 2	Session 3	Mean
12.8%	7.1%	7.3%	9.1%

The results showed that LDA performs well as a feature reduction method. Classification error was very small. In our case the average classification error for separation of three classes was only 4.8% for the first subject. The average classification error for separation of three classes was 9% for the second and third subject.

In the next step authors tried to find the dependence of LDA component number on classification error. The results of this experiments for third subject and third session is presented in the fig. 5. As it is shown in fig. 5 the smallest error was obtained for the first two LDA components.

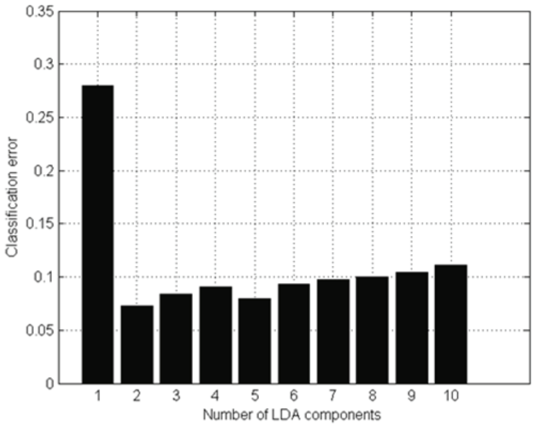

Fig. 5. Classification error dependence on LDA component number

Next authors tested classifier using quadratic discriminant analysis (QDA). The results are shown in tables 4, 5, 6.

Table 4. Classification error for the first subject

First subject			
Session 1	Session 2	Session 3	Mean
6.8%	3.5%	4.2%	4.8%

Table 5. Classification error for the second subject

Second subject			
Session 1	Session 2	Session 3	Mean
9.8%	7.5%	8.9%	8.7%

Table 6. Classification error for the third subject

Third subject			
Session 1	Session 2	Session 3	Mean
11.3%	6.9%	7.3%	8.5%

Classification errors for QDA were also very small. In that case the average classification error for separation of three classes was only 4.8% for first subject. The average classification error for separation of three classes was 8.7% for the second and 8.5% for the third subject.

Conclusions

In experiments authors used only 8 EEG channels. Discrete Fourier transform coefficients were used as features. For feature reduction a linear discriminant analysis was implemented. The experiments proved that it is possible to classify successfully the "mental tasks" with the use of only 8 electrodes. Linear discriminant analysis is a good tool for feature reduction. Only two components of LDA were used. Authors tested 10-NN, LDA and QDA classifiers. All classification methods gave small classification errors. For all this methods and for the first user, classification error was only 4%. For the second and third user classification error was about 9%. Such a small error is a very good result. It is worth noting that EEG signal contained artifacts such as EOG and ECG. The main disadvantage of implementing LDA for feature reduction is the need of using all FFT coefficients (features) as the input signals. The new set of features is calculated based on all observations. Such a situation will not occur in real BCI implementation. So the results for real BCI system can be worse.

REFERENCES

[1] Vidal, J.J., *Direct brain-computer communication*, Ann. Rev. Biophys Bioeng, 2, 1973.

[2] Wolpaw J.R., Birbaumer N., Heetderks W. J., Mcfarland D.J,Hunter Peckham P., Schalk G, Donchin E., Quatrano L.A., Robinson C.J, Vaughan T.M, *Brain–Computer Interface Technology: A Review ofthe First International Meeting*, IEEE Transactions on Rehabilitation Engineering, vol. 8, No. 2, June 2000.

[3] Kołodziej M., Majkowski A., Rak R. *Matlab FE_Toolbox - an universal utility for feature extraction of EEG signals for BCI realization*, Przegląd Elektrotechniczny 2010-1.

[4] Kantardzic M., “ *Data Mining: Concepts, Models, Methods, and Algorithms* ”, IEEE Press & John Wiley, November 2002.

[5] Gareis, Ivan E.; Acevedo, Ruben C.; Atum, Yanina V.; Gentiletti, Gerardo G.; Banuelos, Veronica Medina; Rufiner, Hugo L., *Determination of an optimal training strategy for a BCI classification task with LDA*, Neural Engineering (NER), 2011 , Page(s): 286 - 289

[6] Bhattacharyya, S.; Khasnobish, A.; Chatterjee, S.; Konar, A.; Tibarewala, D.N., *Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data*. Systems in Medicine and Biology (ICSMB), 2010 , Page(s): 126 - 131

[7] Sadeghian, E.B.; Moradi, M.H., *Continuous Detection of Motor Imagery in a Four-Class Asynchronous BCI*, Engineering in Medicine and Biology Society, 2007. EMBS 2007. 2007, Page(s): 3241 - 3244

Authors: *prof. dr hab. inż. Remigiusz J. Rak, e-mail: remigiusz.rak@ee.pw.edu.pl; dr inż. Andrzej Majkowski, e-mail: amajk@iem.pw.edu.pl; mgr inż. Marcin Kołodziej, e-mail: kolodzim@iem.pw.edu.pl Politechnika Warszawska, Instytut Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych, ul. Koszykowa 75, 00-661 Warszawa.*