Classification of Motor Imagery Using Combination of Feature Extraction and Reduction Methods for Brain-Computer Interface

Classification of Motor Imagery Using Combination of Feature Extraction and Reduction Methods for Brain-Computer Interface

Vacius Jusas

Department of Software Engineering; Kaunas University of Technology; Kaunas, Lithuania;
e-mail: vacius.jusas@ktu.lt

Sam Gilvine Samuvel

Department of Software Engineering; Kaunas University of Technology; Kaunas, Lithuania;
e-mail: sam.samuvel@ktu.edu

Corresponding author: sam.samuvel@ktu.edu

The motor imagery (MI) based brain-computer interface systems (BCIs) can help with new communication ways. A typical electroencephalography (EEG)-based BCI system consists of several components including signal acquisition, signal pre-processing, feature extraction and feature classification. This paper focuses on the feature extraction step and proposes to use a combination of several feature extraction and feature reduction methods. The research presented in the paper explores the methods of band power, time domain parameters, fast Fourier transform and channel variance for feature extraction. These methods are investigated by combining them in pairs. The application of two feature extraction methods increases the number of selected features that can be redundant or irrelevant. The utilization of too many features can lead to wrong classification results. Therefore, the methods of feature reduction have to be applied. The following feature reduction methods are investigated: principal component analysis, sequential forward selection, sequential backward selection, locality preserving projections and local Fisher discriminant analysis. The combination of the methods of fast Fourier transform, channel variance and principal component analysis performed the best among the combinations of methods. The obtained classification accuracy of the above-mentioned combination of the methods is much higher than that of the individual feature extraction method. The novelty of the approach is based on consolidated sequence of methods for feature extraction and feature reduction.

KEYWORDS: Brain computer interface, feature extraction, feature reduction, combination of methods.

1. Introduction

The analysis based on the electroencephalograph (EEG) records provides valuable information on the brain activities. The mental activities guide the changes in the brain's signals. The BCI framework is responsible for acquiring, measuring, and converting these brain signals into control commands. As a non-invasive measurement method of brain activity, EEG has attracted increasing interest because of its low risk, low cost, and significant potential for practical applications. EEG data are generally composed of multichannel signals recorded from several electrodes placed on the scalp to record the activity of various cortexes. The decoding of human motor intentions mainly utilizes motor imagery (MI) EEG signals and could provide users with direct control of various devices without utilizing any

peripheral nerves or muscle movements [6, 13]. The typical EEG-based BCI framework consists of several components, such as signal acquisition, signal pre-processing, feature extraction and feature classification. The component of feature extraction has a direct influence on the success of feature classification. Therefore, the methods of feature extraction are very important and of huge interest to the researchers. EEG is a type of random signal that contains highly complex information. Thus, a single feature extraction method usually cannot describe the properties of EEG signals fully. So, it is necessary to take full advantage of different types of feature extraction methods and explore the optimal combination of

input features. While several methods are combined to extract features from original EEG signals, the problem of dimensionality appears. Moreover, many extracted features might be redundant or irrelevant; consequently, they can have a negative effect on classification task. In this case, a dimensionality reduction becomes an important preprocessing step before feature classification.

This paper presents the efficient feature extraction approach for EEG signals classification and suggests utilizing a combination of several methods for feature extraction and reduction. The combination of several methods of feature extraction and reduction can well express the different characteristics of EEG signals. The paper is organized as follows: Section 2 focuses on the related work, Section 3 presents the methods

used in this research, Section 4 introduces our proposed approach. The experimental results are delivered in Section 5 and Conclusions form Section 6.

2. Related Work

Rodriguez-Bermudez et al. [17] presented a wrapper-based methodology for feature selection. Features are computed in different time segments using feature extraction methods for power spectral density (PSD) features, adaptive autoregressive (AAR) coefficients and Hjorth parameters. The features then are averaged and concatenated into a single vector. Next, the proposed framework is used to select the appropriate features. The framework has two stages. The first stage involves feature ranking and the second stage consists in the selection of the most suitable features utilizing a leave-one-out estimation based on the Allen's PRESS statistics. Two different procedures have been considered for feature ranking, such as least angle regression (LARS) and the Wilcoxon rank sum test. It is confirmed that the LARS algorithm provides better results than the Wilcoxon rank sum test. The experiment showed that PSD features are the most selected in all the cases; next, the Hjorth parameters; and the less informative variables correspond to the AAR coefficients.

Garcia-Laencina et al. [8] investigated the dimensionality reduction techniques. Firstly, the following feature extraction methods were applied: the band power in two different frequency bands of the EEG data, the Hjorth parameters and the adaptive autoregressive coefficients. Then, the standard feature selection procedures were employed: sequential backward selection and sequential forward selection. At last, the authors explored the three dimensionality reduction techniques: principal component analysis, locality preserving projections, and the local Fisher discriminant analysis (LFDA). According to the experimental results, there was no an individual feature extraction method that provided the best performance for all the subjects. Moreover, making use of input features from all the methods did not ensure an improvement in the accuracy of classification results. However, making use of dimensionality reduction techniques increased

the classifier performance with respect to the original feature space. It has been noticed that the LFDA showed the best results for all subjects with reduced computational complexity.

Yu et al. [23] analyzed the performance of the feature extraction method using the following spatial filter techniques: common average reference, common spatial pattern, no-spatial filter and feature reduction method using principle component analysis. The results of the experiment showed that the common average reference filter had a better performance than other universal spatial filter techniques. The feature reduction method of principal component analysis did not improve accuracy but maintained the classification performance while decreasing feature number effectively.

Gupta et al. [9] considered empirical mode decomposition and wavelet transform for feature extraction. To decrease the size of the feature vector, six multivariate filter methods such as Euclidean distance, Bhattacharyya distance measure, Kullback-Leibler distance, ratio of scatter matrices, linear regression and minimum redundancy-maximum relevance were investigated. For all the multivariate filter methods, the top 25 features were incrementally included one by one to develop the decision model using sequential forward selection search method. Experimental results showed that the classification accuracy improved with the use of each of the six multivariate feature selection methods. Among all the investigated combinations of feature extraction and selection methods, the combination of wavelet transform and linear regression performed the best.

Ren et al. [16] proposed a feature extraction framework that combined hybrid feature extraction and feature selection method. The following feature extraction methods of different types were applied: autoregressive model, discrete wavelet transform, wavelet packet transform and sample entropy. The hybrid input feature vector was composed of 83 features. For feature selection, algorithms of minimal redundancy-maximal relevance and Fisher score based on global search strategy were used. For comparison purposes, the feature reduction method of principal component analysis was introduced. The experimental results demonstrated that class separability was not improved after transformation utilizing principal component analysis. The hybrid features selected on the basis of Fisher score yielded the highest classification accuracy.

Baig et al. [2] implemented a hybrid method that used common spatial patterns filter to extract feature space, then used the method of differential evolution with a classifier (wrapper) to discover the optimized feature subset. The Authors also implemented the following evolutionary algorithms: particle swarm optimization, simulated annealing, ant colony optimization, and artificial bee colony. The experimental results demonstrated that the proposed hybrid method performed well with classifier of support vector machine. Moreover, the comparison results of implemented evolutionary algorithms showed the superiority of the method of differential evolution. However, the proposed method is slow compared to the typical feature selection algorithms and the classifier of the wrapper technique makes it even slower.

Zhang et al. [25] combined autoregressive model and sample entropy for the feature extraction. Each feature vector acquired on the basis of the combination strategy contained two parts: autoregressive coefficients and sample entropy values. In the classification stage, the authors utilized support vector machine with radial basis function (RBF) kernel as the classifier. Experimental results showed that the combination strategy of the feature extraction obtained a better accuracy in comparison with autoregressive model-based method.

Han et al. [10] proposed an EEG classification framework based on EEG feature compression and convergent iterative channel positioning. The framework begins with an EEG signal pre-processing and single channel based on autoregressive coefficients or time domain parameters feature extraction. Next, after dividing all the trials into the training and testing sets, all the features from the different channels for the training trials are gathered and cluster signatures assigned to them through k-means. Then, EEG signals are mapped from the three-dimensional matrix (trial channel-feature) to a two-dimensional (trial-channel) matrix by compressing all feature vectors into their cluster signatures. For the two-dimensional matrix, RFS, RUFS, or SSLSR methods to rank and select the channels are employed. The results of the experiment showed that the execution of the proposed system is comparable with the state-of-the-art techniques. In the experiment, they performed only two motor imagery tasks. In this research, we used a four different motor imagery tasks.

3. Methods

In this section, we are presenting all methods which will be applied in the experiments. First of these methods is a common spatial pattern which will be used for pre-processing. The inclusion of the pre-processing stage is an important procedure due to the high variability of the EEG measurements [1].

3.1. Common Spatial Patterns

The common spatial pattern technique was initially proposed for classification of multi-channel EEG during imagined hand movements by Ramoser et al. [15]. The primary idea is to utilize a linear transform to project. The multi-channel EEG data into a low dimensional spatial subspace with a projection matrix, whose every row comprises of weights for channels. This change can augment the variance of two class signal matrices. CSP method depends on the concurrent diagonalization of the covariance matrices of two classes.

The points of the algorithm are described as follows with the case of characterizing single preliminary EEG during hand and foot movements X_H and X_F namely the preprocessed EEG matrices under two conditions (hand and foot) with dimensions N*T, where N is the number of channels and T is the number of samples per channel [22]. The standardized spatial covariance of the EEG can be represented as follows:

$$R_H = (X_H X_H^T)/(trace(X_H X_H^T)),$$
$$R_F = (X_F X_F^T)/(trace(X_F X_F^T)). \tag{1}$$

3.2. Feature Extraction

3.2.1. Band Power

There exist various methods for band power (BP) feature extraction from EEG signals. In this investigation, we used a method implemented in the Biosig biomedical signal processing library that enumerates the band power by band-pass filtering the signal [18]. According to this method, initially, the signal is filtered through a band-pass filter that is designed for a given frequency band. In case of a Biosig library, a 4th order Butterworth infinite impulse response (IIR) filter is used. The next step consists in squaring each sample of the resulting signal $x(t)$[1], which contains only the required frequency components, to achieve the time course of power

$$p = x^2(t). \tag{2}$$

Given the smoothing window size w, the following smoothing operation is applied to the signal obtained from the previous step:

$$\overline{p}[n] = 1/w \sum\nolimits_{k=0}^{w} p[n-k]. \tag{3}$$

This means that the band power for sample n is equivalent to the average power of w preceding samplings. The final feature values are equal to $\ln(\overline{p}[n])$. The logarithm is used because it can enhance the performance of linear classification.

3.2.2. Time Domain Parameters

Another feature extraction technique used in this research is related to time domain parameters (TDP) [21]. The TDP implemented in the Biosig library can calculate the time-varying power of the first k derivatives by using the following equation:

$$p_i(t) = (d^i x(t))/dt^i, i = 0,1,...,k. \tag{4}$$

The obtained values might be smoothed by using an exponential moving average window filter. It can be implemented using the following infinite impulse response (IIR) filter [18]:

$$y[n] = u*p_i[n] - (1-u)*y[n-1]. \tag{5}$$

The p_i indicates the input signal (i th order derivative), y indicates filtering result. The u value is used as a parameter for computing the time domain parameters. The final feature values are equal to ln (y[n]).

Ofner et al. [14] compared feature extraction methods based on TDP, BP, Hjorth, adaptive autoregressive (AAR) parameters, bilinear AAR parameters, multivariate AAR parameters and then concluded that TDP is the most efficient of all compared feature extraction methods.

3.2.3. Fast Fourier Transform

A fast Fourier transform (FFT) is an algorithm that samples a signal over a period of time or a certain space and partitions it into its frequency components. FFT is an approach for efficiently processing the discrete Fourier transform of a series of data samples [7]. In the discrete time case, the data to be transformed could be partitioned into frames

(which commonly overlap, to reduce artifacts at the boundary). Each frame is Fourier transformed, and the complex result is added to a matrix, which records a magnitude and phase for each point in time and frequency. This can be expressed as follows:

$$STFT\{x[n]\}(m,w) \equiv X(m,w)$$
$$= \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{(-jwn)}. \qquad (6)$$

In such case, m is discrete, and w is continuous, but in most common applications of the short-time Fourier transform (STFT) is performed utilizing the fast Fourier transform, so both variables are discrete and quantized. The FFT demonstrates the overall robustness and consistency in generating the most distinguishable sets of features from MI induced EEG trials [12, 19].

3.2.4. Channel Variance

Channel variance (CV) for every i-th EEG channel is the second moment of the signal computed about its mean $\bar{x}$ [19]. The result is normalized using Box-Cox transformation [4] for the final feature vector:

$$y_i = log(1 / N \sum_{k=1}^{N} (x_i[k] - (\overline{x_i})^2,$$
$$i = (\overline{1, n}). \qquad (7)$$

Uktveris and Jusas [19] considered many feature extraction methods, among them BP and TDP. The obtained results show that the performance of CV method is second only to that of the FFT method.

3.3. Feature Reduction

3.3.1. Locality Preserving Projections

Locality preserving projection (LPP) is a nonlinear dimensionality reduction technique. The optimality criterion based on the LPPs for extending the local mutual relationship can exist among the input data vectors to the vectors of the projected subspace [11]:

$$D_N = min_{i,j}^{\Sigma} (\overrightarrow{y_i} - \overrightarrow{y_j})'(\overrightarrow{y_i} - \overrightarrow{y_j})s_{i,j}. \qquad (8)$$

The equation for the local relationships between the input data vectors may be described as the similarity matrix, S = {s(i,j)} N×N, where the similarity relationship is defined as follows:

$$S_{i,j} = \begin{cases} exp(-\|\overrightarrow{x_i} - \overrightarrow{x_j}\|^2 / \rho), e(\overrightarrow{x_i} - \overrightarrow{x_j})=1 \\ 0, e(\overrightarrow{x_i}, \overrightarrow{x_j})=0 \end{cases}. \qquad (9)$$

The function of e (x_i, x_j) is defined as an indicator function. The x_i, x_j are neighbours and ρ is the heat kernel factor.

3.3.2. Local Fisher Discriminant Analysis (LFDA)

Local Fisher discriminant analysis is a combination of the concepts of both Fisher discriminant analysis (FDA) and LPP to define the local FDA. Typically, LFDA computes both between-class covariance matrix and the within-class covariance matrix in a local manner [8].

The LFDA transformation matrix T is defined as follows:

$$T_{LFDA} = \arg \max_{T \in \mathbb{R}^{d \times r}} [tr((T^T s^{(\omega)}t)^{-1} T^T s^{(b)}T)]. \qquad (10)$$

LFDA tries to map the nearby data pairs of the same class as close together as possible in the reduced space; meanwhile, the data pairs in different classes are separated from one another.

3.3.3. Principal Component Analysis (PCA)

PCA is a filtering technique, which utilizes an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called main components. In general, PCA has been successfully applied in many scientific fields and PCA also receives much attention in BCI experiments [8]. PCA is an unsupervised method that calculates a linear mapping F in order to accomplish a low-dimensional representation of the original data in which the amount of variance is maximal.

PCA finds F that maximizes the cost function:

$$F_{PCA} = \frac{\arg \max}{F \in \mathbb{R}^{D \times r}} [trace(F^T SF)], \qquad (11)$$

where S is a covariance matrix of F. This linear transformation is formed by an orthogonal basis from the top eigenvectors of the covariance matrix [8]. During transformation, the eigenvectors corresponding to the r largest eigenvalues are retained.

4. The Proposed Approach

CSP is generally perceived as a signal pre-processing method that decomposes the raw EEG into subcomponents (spatial patterns) having maximum differences in variance. This technique allows better feature sep-

aration in feature space and thus a more exact signal classification. Likewise, the property of CSP to decrease feature dimensionality is exceptionally reasonable for EEG data complexity reduction. It has been shown by Uktveris & Jusas [19] that this technique enables obtaining a substantial increase in EEG signal classification performance. Therefore, this method is exceptionally suitable for signal filtering.

The single feature extraction method typically cannot define the properties of EEG signals completely. Therefore, this paper proposes a combination of two methods for EEG feature extraction, which include different categories of characteristics. However, the application of two feature extraction methods produces a larger set of features. Some of the extracted features might be repetitive or insignificant; subsequently, they can have a negative impact on classification task. In this situation, a dimensionality reduction turns into an important preprocessing step before feature classification.

CSP is a method to analyse multi-channel data based on recordings from two classes (tasks). The CSP method is used for reducing the number of the channels. A diagram of our proposed approach is shown in Figure 1. The first stage of Figure 1 shows the input of EEG brain signals. The second stage shows the common spatial patterns used in signal pre-processing for separating a multivariate signal.

The third block shows two feature extraction methods. In this case, we will consider BP, TDP, FFT and CV approaches and use them in pairs. According to obtained results of the experiment, we will decide which combination of feature extraction methods is the best one. For the dimensionality reduction, we consider PCA, LPP, sequential forward selection (SFS), sequential backward selection (SBS) and

LFDA methods.

These feature extraction methods were chosen for the following reasons:

1 BP method is one of the most frequently used feature extraction methods;

2 TDP method is the most efficient compared to feature extraction methods BP, Hjorth, and adaptive autoregressive parameters (Ofner et al.).

3 The results obtained in [19] show that the performance of FFT method and CV method are the best

Figure 1

Block diagram of the proposed approach
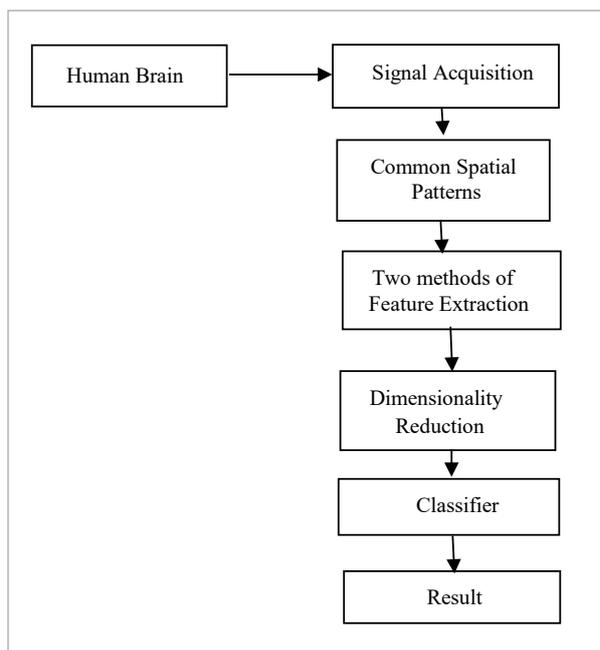
among the other feature extraction methods for the deep learning approach. The methods BP and TDP were included among the considered methods.

4 Furthermore, the considered methods BP, TDP, FFT and CV represent the signal characteristics from different views: BP – energy in frequency domain; TDP – time domain parameters, FFT – amplitude and phase in frequency domain, and CV – second moment of the signal.

According to the results of the experiment, we will decide which feature reduction method is the best suited to the previous combination of feature extraction methods.

5. Experimental Studies

5.1. Data

The 2a data set is utilized from the BCI Competition IV for the experiments [3]. This data set includes data from nine users over two sessions and of each recorded on the individual days. During the trial, participants performed one out of four different motor

imagery tasks: such as movement, imagination of left hand, right hand, both feet, and tongue. In total, each of the two sessions consisted of 288 trials (72 trials per class) in random order. To mark the starting of a trial, a cross appeared on the black screen.

During the process, the subjects heard a tone indicates the trial on the set. The subjects could see an indicator that pointed either to the left, right, top or bottom of the screen. The performance is based on the corresponding motor imagery task until the cross disappeared after 6s. There can be a short break between 1.5 and 2.5s before the next trial. The data consist of 22 EEG signals recorded in a monopolar manner (referenced to the left mastoid and grounded to the right mastoid). Signals were sampled at 250 Hz and band-pass-filtered in the range of 0.5 and 100 Hz.

5.2. Results

The classification tests for MI data are carried out using ten-fold cross validation in this study. The combination of the methods is evaluated by means of kappa coefficient. The classification results are also evaluated and compared by using the kappa coefficient, which takes the value 0 for a random classifier and 1 for a perfect classifier that always classifies correctly. The estimation of kappa coefficient is computed using the equation below:

$$K = \frac{(P_0 - P_e)}{1 - P_e},\qquad(12)$$

where P_0 defines classification accuracy and P_e defines the hypothetical accuracy of a random classifier on the same data.

$$K = \frac{(P_0 - 0.25)}{1 - 0.25}.\qquad(13)$$

We consider the value for P_e =0.25. The final proportion of execution of a given algorithm is the maximum value of the kappa coefficient from the computed time-course. The final results for the individual feature extraction methods are presented in Table 1. Two classifiers, namely, the Linear Discriminant Analysis (LDA) [20] and support vector machine (SVM) [5] were utilized to evaluate the individual feature extraction methods.

The best results were obtained when BP and CV features were used. When utilizing band power features with LDA classifier, the kappa value on the testing set was equivalent to 0.45, on this term of SVM it is 0.44. When using CV features with SVM classifier, the testing set was equivalent to 0.45 as well, on this term of LDA it is 0.43.

The average accuracies of the combination of methods are presented in Table 2. The two methods of feature extraction and feature reduction of PCA are combined. The best results for most subjects were achieved when using the FFT, CV and PCA in combination with SVM classifier. The kappa coefficient for this combination is equivalent to 0.54 when using SVM classifier and 0.48 when using the LDA classifier.

Table 1
Average accuracy of single methods

Methods	class	1	2	3	4	5	6	7	8	9	No of Features	Average
BP	LDA	0.74	0.11	0.54	0.35	0.17	0.31	0.63	0.61	0.62	24	0.45
BP	SVM	0.72	0.28	0.55	0.36	0.16	0.28	0.50	0.55	0.56	24	0.44
TDP	LDA	0.75	0.19	0.50	0.31	0.19	0.29	0.52	0.48	0.54	48	0.42
TDP	SVM	0.60	0.12	0.47	0.25	0.17	0.21	0.45	0.41	0.45	48	0.35
FFT	LDA	0.63	0.12	0.51	0.29	0.18	0.37	0.63	0.65	0.52	24	0.43
FFT	SVM	0.66	0.13	0.55	0.32	0.15	0.34	0.65	0.62	0.56	24	0.44
CV	LDA	0.60	0.17	0.53	0.32	0.14	0.32	0.60	0.56	0.59	24	0.43
CV	SVM	0.70	0.25	0.55	0.37	0.18	0.30	0.57	0.58	0.57	24	0.45

Table 2
Average accuracy of the combination of methods

Methods	class	1	2	3	4	5	6	7	8	9	No of Features	Average
BP+TDP+PCA	SVM	0.81	0.25	0.58	0.43	0.18	0.42	0.69	0.61	0.67	52	0.52
BP+TDP+PCA	LDA	0.74	0.11	0.55	0.35	0.17	0.31	0.63	0.60	0.62	52	0.45
FFT+CV+PCA	LDA	0.76	0.14	0.62	0.36	0.26	0.36	0.65	0.58	0.60	46	0.48
FFT+CV+PCA	SVM	0.82	0.26	0.56	0.44	0.24	0.46	0.71	0.65	0.75	46	0.54
FFT+TDP+PCA	LDA	0.65	0.17	0.63	0.39	0.11	0.25	0.69	0.65	0.63	72	0.46
FFT+TDP+PCA	SVM	0.67	0.18	0.64	0.36	0.13	0.23	0.69	0.68	0.66	72	0.47
CV+TDP+PCA	LDA	0.71	0.21	0.51	0.29	0.23	0.26	0.48	0.50	0.53	76	0.41
CV+TDP+PCA	SVM	0.68	0.20	0.49	0.25	0.20	0.29	0.47	0.54	0.50	76	0.40

Table 3
Average Accuracy of the Combination of Methods for LS-SVM

Methods	1	2	3	4	5	6	7	8	9	No of Features	Average
FFT+CV+PCA	0.80	0.28	0.55	0.49	0.30	0.41	0.79	0.68	0.77	46	0.56
FFT+CV+SBS	0.71	0.23	0.67	0.41	0.11	0.23	0.69	0.66	0.72	56	0.49
FFT+CV+SFS	0.70	0.20	0.64	0.37	0.15	0.28	0.68	0.63	0.69	58	0.48
FFT+CV+LPP	0.70	0.15	0.58	0.37	0.13	0.28	0.67	0.66	0.60	52	0.46
FFT+CV+LFDA	0.78	0.19	0.61	0.40	0.16	0.38	0.70	0.60	0.59	54	0.49

The best results were obtained when combining feature extraction methods of FFT and of CV. So, we tested this combination with various feature reduction methods such as SBS, SFS, LPP, LFDA in Table 3. For the classification, a classifier of Least squares support vector machine (LS-SVM) [24] was used. Since this classifier uses a set of linear equations for training, such a classifier is less computationally demanding and, moreover, it has shown a higher generalization capability to that of the SVMs with Gaussian kernels [12]. The best obtained result is 0.56 which is much higher than for the combinations of the other methods. Thus, the combination of methods such as FFT, CV and PCA yielded the most efficient results.

6. Conclusions

The typical EEG-based BCI system consists of several components including signal acquisition, signal pre-processing, feature extraction and feature classification. Every step of the whole process may be improved. For the signal pre-processing, the method of common spatial patterns was used. The method decreased the number of channels from 22 to 8 for the data set 2a from the BCI Competition IV. Assume that the combination of two different feature extraction methods would increase the performance of the classifier. The following feature extraction methods, namely band power, time domain parameters, fast

Fourier transform, and channel variance have been combined in several ways. The combination of methods pertaining to the fast Fourier transform and to channel variance yielded the best result.

The use of two feature extraction methods increases the number of features that can be irrelevant or redundant.

Therefore, a feature reduction method should be applied. The method based on principal component analysis has been chosen from among several methods of its category (which includes sequential backward selection, sequential forward selection, locality preserving projections and local Fisher discriminant analysis) and used. The above-mentioned method in

combination with two feature extraction methods showed the best result. For the classification, three methods are used: linear discriminant analysis, support vector machine and least square support vector machine. The classifier of least square support vector machine showed the best performance. In essence, this paper presents the approach that includes a novel combination of the methods for feature extraction and reduction. The approach includes the following methods: common spatial patterns for signal pre-processing, fast Fourier transform and channel variance for feature selection and principal component analysis for the feature reduction. The experiment showed that this combination of the methods performs the most efficiently.

References

1. Alvarez-Meza, A. M., Velasquez-Martinez, L. F., Castellanos-Dominguez, G. Time-Series Discrimination Using Feature Relevance Analysis in Motor Imagery Classification. Neurocomputing, 2015, 151, 122-129.https://doi.org/10.1016/j.neucom.2014.07.077

2. Baig, M. Z., Aslam, N., Shum, H. P. H., Zhang, L. Differential Evolution Algorithm as a Tool for Optimal Feature Subset Selection in Motor Imagery EEG. Expert Systems with Applications, 2017, 90, 184-195. https://doi.org/10.1016/j.eswa.2017.07.033

3. BCI Competition IV Results. http://www.bbci.de/competition/iv/results/, March 2013.

4. Box, G. E. P., Cox, D. R. An Analysis of Transformations. Journal of the Royal Statistical Society, Series B, Wiley for the Royal Statistical Society, 1964, 26(2), 211-252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

5. Carvalho, S. N., Costa, T. B. S., Uribe, L. F. S., Soriano, D. C., Yared, G. F. G., Coradine, L. C., Attux, R. Comparative Analysis of Strategies for Feature Extraction and Classification in SSVEP BCIs. Biomedical Signal Processing and Control, 2015, 21, 34-42. https://doi.org/10.1016/j.bspc.2015.05.008

6. Choi, B., Jo, S. A Low-Cost EEG System-Based Hybrid Brain-Computer Interface for Humanoid Robot Navigation and Recognition. PLOS ONE, 2013, 8(9), Article Number: e74583. https://doi.org/10.1371/journal.pone.0074583

7. Cochran, W. T., Cooley, J. W., Favin, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., Maling, G. C., Nelson, D.

E., Rader, C. M., Welch, P. D. What is the Fast Fourier Transform? Proceedings of the IEEE, 1967, 55(10), 1664-1674. https://doi.org/10.1109/PROC.1967.5957

8. Garcia-Laencina, P. J., Rodriguez-Bermudez, G., Roca-Dorda, J. Exploring Dimensionality Reduction of EEG Features in Motor Imagery Task Classification. Expert System with Applications, 2014, 41(11), 5285-5295. https://doi.org/10.1016/j.eswa.2014.02.043

9. Gupta, A., Agrawal, R. K., Kaur, B. Performance Enhancement of Mental Task Classification Using EEG signal: A Study of Multivariate Feature Selection Methods. Soft Computing, 2015, 19(10), 2799-2812. https://doi.org/10.1007/s00500-014-1443-1

10. Han, J., Zhao, Y., Sun, H., Chen, J., Ke, A., Xu, G., Zhang, H., Zhou, J., Wang, C. A Fast, Open EEG Classification Framework Based on Feature Compression and Channel Ranking, Front Neuroscience, 2018, 12, Article Number: 217. https://doi.org/10.3389/fnins.2018.00217

11. He, X., Niyogi, P. Locality Preserving Projections. Advances in Neural Information Processing Systems. NIPS'03 Proceedings of the 16th International Conference on Neural Information Processing Systems, 2003, 153-160.

12. Herman, P., Prasad, G., McGinnity, T. M., Coyle, D. Comparative Analysis of Spectral Approaches to Feature Extraction for EEG-Based Motor Imagery Classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2008, 16(4), 317-326. https://doi.org/10.1109/TNSRE.2008.926694

13. Martinez-Leon, J.-A., Cano-Izquierdo, J.-M., Ibarrola, J. Are Low Cost Brain Computer Interface Headsets

Ready for Motor Imagery Applications? Expert Systems with Applications, 2016, 49, 136-144. https://doi.org/10.1016/j.eswa.2015.11.015

14. Ofner, P., Muller-Putz, G. R., Neuper, C., Brunner, C. Comparison of Feature Extraction Methods for Brain-Computer Interfaces. In Proceedings of 5th International BCI Conference, Graz, Austria, 2011, 1-4.

15. Ramoser, H., Muller-Gerking, J., Pfurtscheller, G. Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement. IEEE Transactions on Rehabilitation Engineering, 2000, 8(4), 441-446. https://doi.org/10.1109/86.895946

16. Ren, W., Han, M., Wang, J., Wang, D., Li, T. Efficient Feature Extraction Framework for EEG Signals Classification. IEEE Seventh International Conference on Intelligent Control and Information Processing (ICICP), Siem Reap, Cambodia, 2016, 167-172. https://doi.org/10.1109/ICICIP.2016.7885895

17. Rodriguez-Bermudez, G., Garcia-Laencina, P. J., Roca-Dorda, J. Efficient Automatic Selection and Combination of EEG Features in Least Squares Classifiers for Motor Imagery Brain-Computer Interfaces. International Journal of Neural Systems, World Scientific Publishing Company, 2013, 23(4), Article Number: 1350015. https://doi.org/10.1142/S0129065713500159

18. Szachewicz, P. Classification of Motor Imagery for Brain-Computer Interfaces. Master's thesis, Poznan University of Technology, Faculty of Computing and Information Science, Institute of Computing Science, Poznan 2013.

19. Uktveris, T., Jusas, V. Application of Convolutional Neural Networks to Four-Class Motor Imagery Classification Problem. Information Technology and Control, 2017, 46(2), 260-273. https://doi.org/10.5755/j01.itc.46.2.17528

20. Vega, R., Sajed, T., Mathewson, K. W., Khare, K., Pilarski, P. M., Greiner, R., Sanchez-Ante, G., Antelis, J. M. Assessment of Feature Selection and Classification Methods for Recognizing Motor Imagery Tasks from Electroencephalographic Signals. Artificial Intelligence Research, 2017, 6(1), 37-51. https://doi.org/10.5430/air.v6n1p37

21. Vidaurre, C., Kramer, N., Blankertz, B., Schlogl, A. Time Domain Parameters as a Feature for EEG-based Brain-Computer Interfaces. Neural Networks, 2009, 22(9), 1313-1319. https://doi.org/10.1016/j.neunet.2009.07.020

22. Wang, Y., Gao, S., Gao, X. Common Spatial Patterns for Channel Selection in Motor Imagery Based Brain Computer Interface. In Proceedings of the IEEE, Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 2005, 5392-5395.

23. Yu, X., Chum, P., Sim, K.-B. Analysis the Effect of PCA for Feature Reduction in Non-stationary EEG Based Motor Imagery of BCI system. Optik, 2014, 125(3), 1498-1502. https://doi.org/10.1016/j.ijleo.2013.09.013

24. Zarei, R., He, J., Siuly, S., Zhang, Y. A PCA Aided Cross-covariance Scheme for Discriminative Feature Extraction from EEG Signals. Computer Methods and Programs in Biomedicine, 2017, 146, 47-57. https://doi.org/10.1016/j.cmpb.2017.05.009

25. Zhang, Y., Ji, X., Liu, B., Huang, D., Xie, F., Zhang, Y. Combined Feature Extraction Method for Classification of EEG Signals. Neural Computing and Applications, 2017, 28(11), 3153-3161. https://doi.org/10.1007/s00521-016-2230-y