Implementation of automatic feature selection methods for BCI realization

Andrzej Majkowski, Marcin Kolodziej, Remigiusz J. Rak
Institute of Theory of Electrical Engineering and Measurement and Information Systems
Warsaw University of Technology, Poland

***Abstract*—The main task of brain-computer interface is to translate signals generated by neurons of the brain into commands. For the effective operation of BCI, efficient methods of feature selection of EEG signal are needed. In this article authors propose the use of correlation and *t*-statistics to feature selection.**

I. Introduction

Brain-computer interface (BCI) is one of the new challenges, which is part of a broader issue known as man-machine communication, allowing the conversion of thoughts into action. Potentially, BCI provides a direct link between the brain and the outside world. The most important application of BCI is to improve living conditions of paralyzed people. The main task of BCI is to read the signals generated by neurons of brain, to decipher their meaning and generate appropriate signals to control another system. The EEG signal is most often used for that purpose.

The first step of EEG signal analysis is feature extraction. Then, for effective operation of brain-computer interface, usually an efficient method of feature selection is needed [2,3,9]. It is because the best features for a user can change from day to day and even from hour to hour. Therefore, BCI systems operate in two modes: calibration (adjustment to the user) and implementation. Calibration mode should be performed before every usage of the interface.

The number of extracted features depends on the used cerebral potential and can range from a few to several thousands. Too large number of features makes effective training of the classifier impossible, so there is a need for feature reduction. The aim of feature selection is to choose the best ones which will effectively differentiate classes of EEG signal.

II. The proposed feature selection methods

There are many feature selection algorithms. The fastest of them belong to ranking methods. In this article authors propose ranking feature selection algorithms based on correlation and *t*-statistics. In the first method features are ranked according to a measure that describes how a particular feature is correlated with others. The proposed method is based on the assumption that features less correlated will bring more information to the classifier. As a measure of correlation between two variables X_1 and X_2 (features) the Pearson linear correlation coefficient is used:

$$r_{X_1X_2} = \frac{cov(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} \tag{1}$$

where $cov(X_1, X_2)$ is the covariance of variables X_1 and X_2, σ_{X_1} and σ_{X_2} depict standard deviation of variable X_1 and variable X_2. Values of the correlation coefficient $r_{X_1X_2}$ belong to the range [-1,1]. The higher the absolute value of the correlation coefficient, the stronger linear relationship between variables. Equality $r_{X_1X_2} = 1$ depicts an exact positive linear relationship between features X_1 and X_2, $r_{X_1X_2} = 0$ denotes that there is no linear relationship between features. If $r_{X_1X_2} = -1$ there is an exact negative linear relationship between features - if one variable increases the other decreases. As it was already mentioned, if a feature is less correlated with others, it will bring more information to the classification process. The authors determined the correlation of features by implementing a matrix, which consists of correlation coefficients for all pairs of variables X_j and X_k. It is a square matrix, with ones on the diagonal. In the next step, elements of rows of the correlation matrix are added and a measure S is created:

$$S_i = r_{i1} + r_{i2} + ... + r_{iK} \tag{2}$$

In the formula (2) K - depicts the number of features and i - the feature index for which the measure S is determined. The smaller value of the sum S, the less correlated is a particular feature with others. In this way, it is possible to rank features from the least to the most correlated.

The second proposed method is based on ranking features in accordance to the Welch *t*-statistics. The best features are those which have the greatest value of Welch *t*-test (3):

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \tag{3}$$

In the formula (3) $\overline{X_1}$ - depicts the mean value of variable X_1 in the 1-st trial, σ_1^2 variance of the variable X_1 in the 1-st trial and N_1 number of variable representations in the 1-st trial. There are the same meanings for the variable X_2. In the experiment authors used the absolute value of t. The t coefficient can be interpreted as a value that describes the degree of differentiation between two considered classes. The greater the difference of mean values in the numerator of the formula (3) and smaller variances of the variables in the denominator, the greater the t ratio, and therefore greater the opportunity for diversification of classes using these variables. Features can be sorted according to the growth of the t factor, and then those with the highest values will be chosen for

978-1-4577-1772-7/12/$26.00

analysis. Another problem is a proper estimation of the overall number of features that gives the best results of classification.

III. Experiment description

The experiment included several steps. In the first step an appropriate set of features from the previously recorded EEG signal was extracted. For feature extraction a Fast Fourier Transform (FFT) was used. In the next step features were selected by using S and t coefficients. At the end, the quality of selected features was verified. As a measure of the quality the classification error was used. Additionally, the authors tried to find the influence of the number of selected features on the quality of classification.

For the experiment, authors used a database of EEG signals, recorded for tests on brain-computer interface (BCI Competition) [7]. Registered EEG signals contain results for several sessions of three users. In this article there are presented the results of experiment for only one, selected user. However, very similar results were obtained for the other users.

Each user realized certain mental tasks. The tasks were following: imagining of the right hand movement, imagining of the left hand movement and generating words starting from a given letter. Sampling frequency of the recorded signal was 512 Hz. The EEG signal was recorded using 32 electrodes in the 10-20 system. All recordings were divided into one second windows (512 samples each). The windows were taken every 32 consecutive samples, so they overlap by 480 samples.

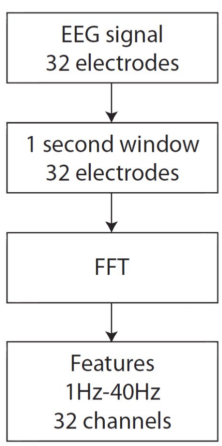

Fig. 1. The process of feature extraction from the EEG signal

For each window FFT was counted and in this way features were obtained [10]. Since only certain frequencies [6,9] carry out useful information about the expected potentials of the EEG signal, the number of features was limited to 40 spectral lines in the bandwidth from 1 Hz to 40 Hz. The block diagram of the feature extraction process is presented in fig.1. As the number of channels was 32 and the number of features for a single electrode was 40, the total number of features for each EEG window was 1280. Such a large number of features made classifier learning very difficult.

To perform feature selection the authors calculated feature correlation matrices for all classes (first method). Thus, the proposed method did not take into consideration belonging to a particular class. The only criterion for the quality of a feature was its correlation with other features. Features are ranked according to the value of S coefficients (2).

With *t*-statistics (second method) it is possible to rank features differentiating classes in pairs: (1,2), (1,3), (2,3). The authors suggested a common factor t_c (fig.2) describing statistics among all classes:

$$t_c = t_{12} + t_{13} + t_{23} \tag{4}$$

This ratio can be interpreted as the total sum of differences among classes. The higher the t_c value, the more features differentiate particular classes. Authors suggested this approach, because their tests proved that features are ranked very similarly like in a case when coefficients t_{12}, t_{13} and t_{23} are used separately. To determine the t_c coefficient for each of 1280 features, first coefficients t_{12}, t_{13} and t_{23} were calculated according to formula (3). The next step was to rank the features from the largest to the smallest t_c values. Then, during the classification step, a number of N best features (with the largest t_c) was selected.

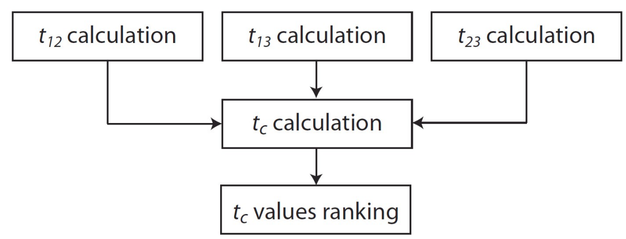

Fig. 2. The process of calculating t_c coefficients and selecting the best features

Classification error is an important factor for testing a feature selection methods. To determine the error a 10-cross validation test was used [3]. The entire dataset was divided into 10 subsets. The 1-9 subsets were used for classifier learning and the last one for testing. Next, the learning process was performed for the subsets 1-8 and 10. The testing was carried out on 9-th subset. This process was repeated for each of the 10 subsets. Tests were conducted using two classifiers - linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). The LDA classifier which linearly separated classes turned out more reliable. This classifier was characterized by better generalization, for the data from a new session. The results obtained by using QDA were better for a particular session. However, for a new data set (new session) classifier performed a lot worse. The classifier learning process is not only affected by the quality of features, but also by their number. Therefore, we changed the number of selected features (ranked from largest to smallest value of t_c).

IV. Results

The authors tried to find the dependence between the number of features and the quality of classification for features selected by using S coefficients, *t*-statistics and randomly

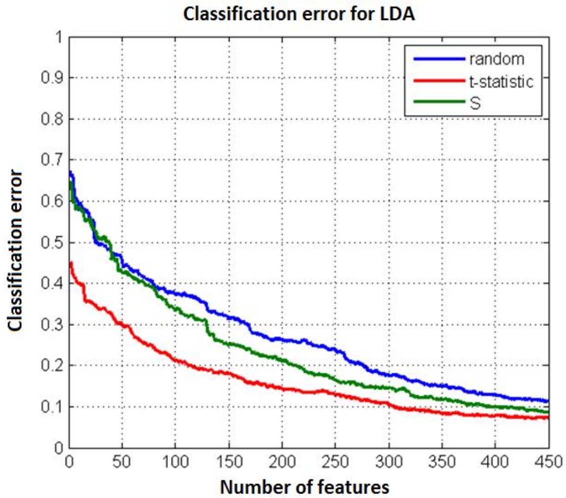

Fig. 3. Classification error for the LDA classifier counted for features chosen by t-statistics, S coefficients and for randomly chosen feature set

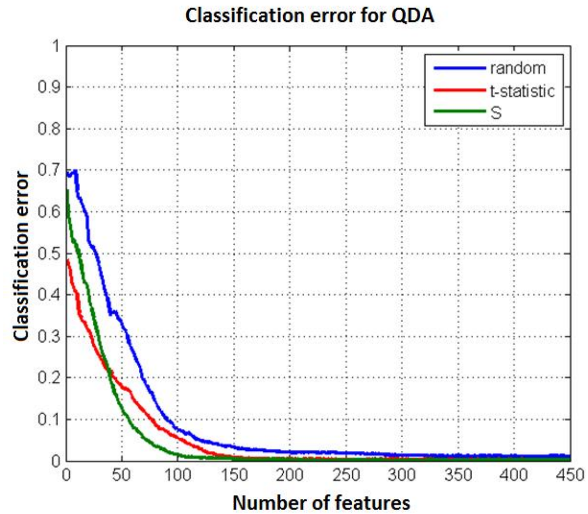

Fig. 4. Classification error for the QDA classifier counted for features chosen by t-statistics, S coefficients and for randomly chosen feature set

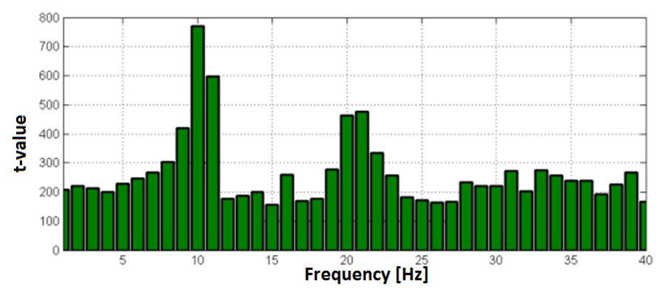

Fig. 5. The sum of the t_c coefficients for each frequency band. The higher the value of the sum, the more important is the feature

chosen. As it was mentioned above, tests were performed for both LDA and QDA classifiers. The classification results (dependence of the classification error on the numbers of features) are presented in figures 3 and 4. Higher number of features can better explain a model (phenomenon) and at the same time can reduce the classification error, especially for cross validation test, although it makes the calculation a lot more time consuming. But in a case of much greater number of features (about 800) we have observed increasing of classification error, which is not visible on figures 3 and 4, because of the limited range of horizontal axis.

For the LDA classifier selection method based on t-statistics helped to reduce the classification error. For example for 50 features the errors for t-statistics and randomly chosen features were respectively 0.31 and 0.45 (fig.3). The improvement is approximately about 14%. We do not observe so significant improvement for the feature selection method based on S coefficients. Feature selection based on t-statistics performs much better because it implements the differences between classes. For QDA classifier the selection method based on t-statistics made possible to reduce the classification error for about 15% (for 50 features corresponding errors were 0.18 and 0.33 - fig.4). For the classification method based on S coefficients the classification error (for 50 features) was 0.12. Thus, in that case, the method based on correlation gave better results.

An important element of the experiments was an attempt to indicate which features are the best (carry more information) for the classification process. For this purpose, the sum of the t_c coefficients were determined for each frequency band. The results clearly show (fig.5) that frequencies 9-11Hz and 20-23Hz are the best for differentiating classes which is consistent with existing theory [1]. In order to assess the suitability of the t-statistics for feature selection we draw the distribution of t_c values for 10Hz projected on user's scalp (fig.6). Higher values describe the areas from which the EEG potential should be registered. Whereas by drawing the distribution of t_{12}, t_{23} or t_{13} values on user's scalp, it is possible to assess which area is better to record EEG signals for differentiation of particular two classes. In fig.7 there are presented the t_{23} values that differentiate classes 2 and 3 for a frequency of 10 Hz. It is easy to designate areas where the electrodes should be stuck.

V. Conclusion

The proposed feature selection methods are known as ranking methods [9]. These methods allow to determine the best features by ranking them. The advantage of applying such a method to brain-computer interface is the high speed of feature selection in comparison to other methods, such as genetic algorithms [4]. In the first approach (S coefficients) the variables are ranked according to the degree of correlation with each other. In the second method (t-statistics) the t coefficients, which differentiated each pair of classes, were summed up. Also the number of features affects the quality of classification. It can be seen that, with the increasing number of features, the importance of the feature selection method decreases. However, to large number of features makes the calculation (in real time) increasingly time-consuming. The

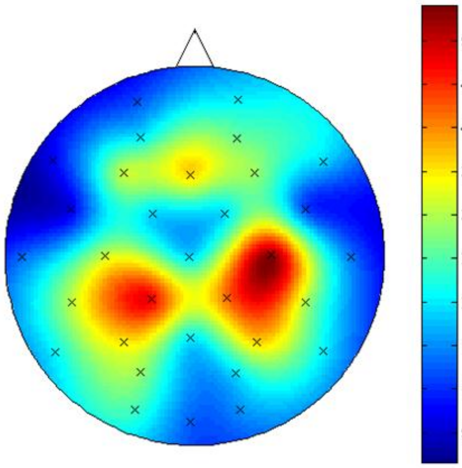

Fig. 6. Values of t_c for 10 Hz mapped on user's scalp

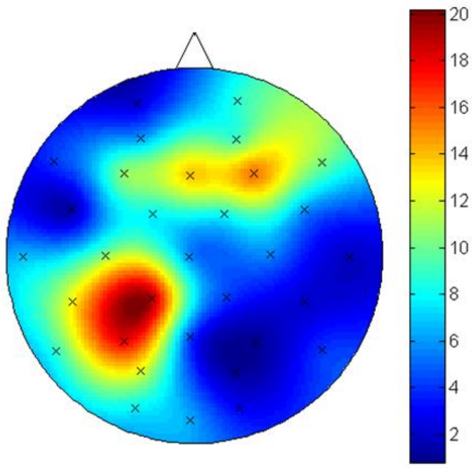

Fig. 7. Diagram showing the distribution of t_{23} values that differentiate classes 2 and 3 for a frequency of 10 Hz

basic problem is therefore to estimate the proper number of features. It can be done by using the proposed feature selection methods and cross validation tests.

This work was partly supported by MNiSZW grant N N510 526239.

REFERENCES

[1] *G. Pfurtscheller and F. H. Lopes da Silva,* Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clinical Neurophysiology, 110:1842–1857, 1999.*

[2] *J. R. Wolpaw, G. E. Loeb, B. Z. Allison, E. Donchin, O. F. do Nascimento, W. J. Heetderks, F. Nijboer, W. G. Shain, J. N. Turner,* BCI Meeting 2005 Wworkshop on signals and recording methods, *IEEE Trans Neural Syst Rehabil Eng: A Pub IEEE Eng Med Biol Soc, 14:138:141, 2006.*

[3] *J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan,* Braincomputer interfaces for communication and control, *Clin Neurophysiol, 113:767:791, 2002.*

[4] *M. Kolodziej, A. Majkowski ,R. Rak,* A new method of EEG classification for BCI with feature extraction based on higher order statistics of wavelet components and selection with genetic algorithms, *International Conference on Adaptive and Natural Computing Algorithms, ICANNGA'11, April 14-16: 2011.*

[5] *J. J. Vidal,* Real-time detection of brain events in EEG, *IEEE Proc: Special Issue on BiolSignal Processing and Analysis, 65 ,1977.*

[6] *D. J. McFarland , J. R. Wolpaw,* Sensorimotor rhythm-based brain-computer interface (BCI): feature selection by regression improves performance, *IEEE Trans Neural Syst Rehabil Eng, 2005 13:372-9.*

[7] *J. R. Milln,* On the need for on-line learning in brain-computer interfaces, *Proc. Int. Joint Conf. on NeuralNetworks, 2004.*

[8] *F. Galn, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, J. R. Milln,* A brain-actuated wheelchair: Asynchronous and non-invasive Brain-computer interfaces for continuous control of robots Original Research Article, *Clinical Neurophysiology, 119(9) 2159:2169,2008.*

[9] *S. Sanei S, J. A Chambers,* EEG Signal Processing, *ISBN: 978-0-470-02581-9, 2007.*

[10] *M. Kolodziej , A. Majkowski, R. Rak,* Matlab FEToolbox - an universal utility for feature extraction of EEG signals for Brain-Computer Interface realization, *Przeglad Elektrotechniczny, 1-2010*