Indian Journal of Science and Technology, *Vol 9(32), DOI: 10.17485/ijst/2016/v9i32/100742, August 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Classification of EEG Signal using Correlation Coefficient among Channels as Features Extraction Method

Saurabh Diwaker[1*], S. K. Gupta[2] and Neeraj Gupta[3]

[1]Department of Information Technology, Krishna Institute of Engineering and Technology, Ghaziabad - 201206, Uttar Pradesh, India; Saurabh.diwakar@kiet.edu
[2]Department of Computer Science, Bundelkhand Institute of Engineering and Technology, Jhansi - 284128, Uttar Pradesh, India; guptask_biet@rediffmail.com
[3]Department of Electrical and Electronics, Krishna Institute of engineering and Technology, Ghaziabad - 201206, Uttar Pradesh, India; neeraj.gupta@kiet.edu

Abstract

Objectives: In this paper, we have evaluated the effectiveness of classification of Electroencephalogram (EEG) signals using the correlation between channels as a method of features selection. **Methods/Statistical Analysis**: First data is broken sample wise, then correlation coefficient between channel pair for each sample is calculated. After that mean of the correlation coefficient of all channel pair for each class over all samples is calculated and in a similar manner, standard deviation from the mean is also calculated. For feature selection we have plotted a pair of the Gaussian curves between channels of two separate classes and choose those channels which give us lower misclassified area as features. Then these features are used for training purpose of Support Vector Machine (SVM). **Findings:** Most of the previous researches follow either signal processing approach or machine learning approach while we emphasized upon the nature of the signal propagation amongst the neurons. The basic idea behind the feature selection is taken from the way the signals propagate from one neuron to the other. In our work we assume that EEG signals follow the normal distribution and verify the fact using chi-square test. On applying SVM the accuracy of classification on testing data confirms that correlation among channels can be used for feature selection. **Application/Improvements:** The results can be improved by improving the pre-processing of EEG signals. It can be used to develop a Brain Computer Interaction (BCI) system.

Keywords: Correlation Coefficient, Electroencephalogram (EEG), Support Vector Machine (SVM)

1. Introduction

A BCI System allows direct communication between a computer and human brain through neural signals that are recorded in the form of electrical activity along scalp most commonly known as Electroencephalographs or EEG signals. The main task establishing a communication link between a computer and human brain is to predict what the human intends to do. For this we need classification of EEG signals into the movements about which human is thinking. EEG signals are extremely complex signals and vary with subjects, action and channels[1,2]. As a classification problem EEG signal classification has always been a challenging task as these are very complex in nature[2,3]. In[4] observed that EEG signals have low SNR and poor spatial resolution in their preliminary form, and hence are hard to classify using any mechanism so preprocessed the signal using the energy entropy of the signal and then used Fisher classifier to classify the samples[4]. A multi-scale filter with different size of filter window was used in, order to find the major frequency band components from EEG signals. Analysis of different energy bands led to an increase in adaptability of the overall system. Once the major frequency bands were

*Author for correspondence

retrieved, Principal Component Analysis (PCA) was used for feature extraction as well as dimensionality reduction[5]. In[6] transformed the randomly observed data into its constituents that are statistically independent using Independent Component Analysis but the original combination is not guaranteed to be a linear combination. Linear ICA separates the artifacts from EEG signals that arouse mainly due to eye blinks and muscle movements. In used mutual information to check the amount of independence between two datasets. Since the combination found in the EEG signals cannot be guaranteed to have a linear combination of the artifacts and the original signal uses the nonlinear ICA. In[6] used separate classifiers for each component and then used voting to resolve the disputes, if any. Autoregressive model and Fast Fourier Transform were used to fetch features and the latest version of learning classifier systems XCS for processing, some parts of beginning of data as well as the ending part of data were removed[7] and then the remaining data was segmented into 128, 256, 512 samples per segment by dividing it into 0.5, 0.25, 0.125 seconds respectively. In[8] Extracted features from the EEG data set using system identification methods that are using the model of the processes that produce the EEG signals. They used scalar Autoregressive model to represent the signal. The model that was used comprised of white noise and previous signals that contribute to make the next signal. Choosing order of the filter depends upon the Autocorrelation function used to represent the EEG signal[8]. The autocorrelation coefficients are the features of the classifier. In[9] dimensions are the same as the order of the system and feed forward type neural network was used for classification purpose. These coefficients were the inputs to the ANN based classifier. Adaptive Neuro-Fuzzy Inference Systems for EEG signal classification along with the fractal dimensions, in the ANFIS were introduced in[9]. In[10] used Higuchi Algorithm for feature extraction to extract the fractal dimensions, and PCA for dimensionality reduction. In[10] used PCA with SVD was used to extract features and then the classifications methods like, Bayesian, k-nearest neighbor, were used. The features were extracted with the fact that fishers criterion for maximum separation is satisfied. A new method for automatic detection and classification of sleep stages by multichannel EEG signal monitoring was given by in[11]. For training vector quantization and sleep stage definition by mass function per every sleep using generalize log likelihood distortion[11] Classification was done using Kullback-Leiber (KL) divergence. We observed that most of the work in the past concentrated on EEG Signals using the aspects of Signal processing and tried to get some features using their signal processing methods which lacked the physiological basis of the EEG signals. Physiological basis is the way neural signals propagate in brain. In the next portion, there will be information about the material used and methodology of work is given. In the same section reason behind considering correlation coefficient is explained. In methodology pre-processing, feature selection and classification are covered. Later Results were discussed. After that conclusion and future scope of the paper is written.

2. Material and Methodology

Classification of EEG signal is a four steps process; namely data acquisition, preprocessing, feature extraction and selection of classifier. In the paper we emphasized upon the nature of the signal propagation amongst the neurons. The basic idea behind the feature selection is taken from the way the signals propagate from one neuron to the other. For a given thought or motion, there must be a neuron from where the whole process begins and then the electric signal propagates through the synaptic junction to reach the neighboring neuron(s), from there to some other neuron(s) in the neighborhood and so on. Thus, the excitation of the source neuron will excite some neighboring neuron which, in turn, will excite some other neuron. Now, if an EEG electrode is placed in the vicinity of an excited neuron, then it will record the activity which should show up as relatively higher values of the voltages. Therefore, we can expect that the signals from the neurons that are getting excited will be strongly correlated. On the other hand, the signals from those neurons that are not getting excited simultaneously will not be correlated. Moreover, since there would be a specific set of neurons getting excited, corresponding to a specific thought or motion, so the high values of correlations would be observed only within the signals from those specific neurons and not from other pairs of neurons. The above observation leads us to conclude that correlations between the signals would have good discriminating features and can be used to differentiate one thought from another. It may be argued that since there is a transfer of signal taking place, the strength of signal may decrease as it moves along a particular path. While this may be true, it will not affect the correlations between the signals of different nodes and therefore, correlations will continue to serve its purpose as a good feature. Based on the above conjecture, we have conducted some tests on EEG signals. We calculated the correlations between the signals of different channels for the same thought. A visual examination of these plots seems to strongly justify our conjecture. In Figure 1, we present the actual plots for the signals from different pairs of channels. X-axis denotes the sample number and Y-axis denotes the amplitude value in microvolt. The pairs were deliberately chosen such that some had high correlation and others had low correlation. Whenever a person thinks of something, some area of brain gets activated. EEG signal are collected using many channels over cerebral cortex. So EEG signal are collected from all parts of brain, while for a particular motion only some parts of the brain contribute. While collecting EEG signal many electrodes are placed over one area of brain. In[12] soon thinking of some motion all the electrodes of that area contributed. That is why we expect that there should be correlation between specific channels of EEG signal corresponding to a particular thought or motion. While the correlation plots above show that there is some merit in our conjecture, we also realize that there are only a few channels that will be strongly correlated corresponding to a particular motion or thought. However, given N channels, we can have N(N-1)/2 possible pairs and therefore the same number of correlations. As mentioned earlier, N can be as large as 128 in modern EEG machines. Most of the correlations will not have the discriminating power that we seek while only a few of them will carry the information that we seek. This implies that if we want to use correlations as features, we will have to apply some tests to choose only those correlations that are useful for our purpose. The procedure adopted for making this choice is explained in subsequent sections.

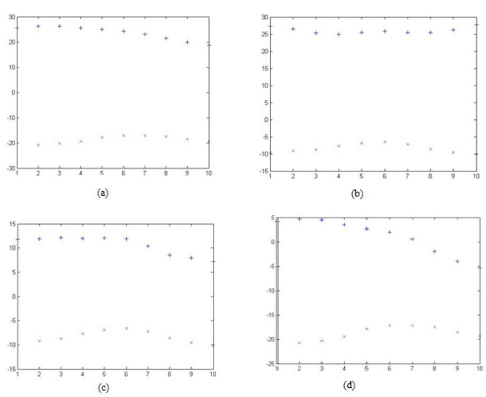


Figure 1. Plots of strongly correlated and uncorrelated channel pairs. **(a) (b)** Plots of channel 11('+') and channel 30 ('*') for class +1 and -1 respectively (Highly correlated). **(c) (d)** Plots of channel of channels 11 ('+') and channel 22 ('*') for class +1 and -1 respectively (Highly correlated).

2.1 Material

We used publically available data provided by Eberhard-Karls-Universität Tubingen, Germany in BCI completion III. IT is a two class dataset recorded on a single subject for two tasks. There are total 278 trials in training set and 100 trials in testing set. Each trail is of 3 second duration and sampling rate is 100Hz[12].

2.2 Methodology

As described above that EEG signal classification is a four step process. So here we are discussing each step in detail.

2.2.1 *Preprocessing*

First EEG data is Split trial wise. After breaking the data, correlation coefficient of all pairs of channels is calculated. Correlation coefficient can be found using

$$\mathrm{corr}(x, y) = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} \tag{1}$$

Where cov(x,y) is covariance between x and y and can be calculate using

$$\mathrm{cov}(x, y) = E[(x - \mu_x)(y - \mu_y)] \tag{2}$$

μ_x and μ_y are expected values of random variables x and y.

σ_x and σ_y are standard deviation of x and y.

After getting correlations between channels of each trial, mean of correlation coefficient over all trials for each channel pair is calculated using:

$$\mu^{c}_{ch,ch'} = \frac{1}{N_c} \sum_{t=1}^{N_c} M^{t,c}_{ch,ch'} \tag{3}$$

$\mu^{c}_{ch,ch'}$: Mean of correlation coefficient between channels ch and ch' for class c taken over all the trial.

N_c: Number of trials of a class c.

$M^{t,c}_{ch,ch'}$: Correlation coefficients between channel ch and ch' for class c taken over all the trials.

Then standard deviation of correlation coefficients of all channel pairs of a class taken over all the trial from the mean is found out using formulae.

$${\sigma^{c}_{ch,ch'}}^{2} = \left\{\frac{1}{(N_c - 1)}\right\} \sum_{t=1}^{N_c} \left(M^{t,c}_{ch,ch'} - \mu^{c}_{ch,ch'}\right)^2 \tag{4}$$

${\sigma^{c}_{ch,ch'}}^{2}$: Standard deviation of correlation coefficient between channel ch and ch' of each trial from the mean.

Now we model the correlation coefficients of each channel pair as a normal distribution. The choice of this distribution is due to the fact that general things around us in real life follow the normal distribution. This is clearly an assumption and we performed a goodness of fit test to verify this assumption. Specifically, we performed a CHI-SQUARE test which gave positive results for confidence levels up to 84.32% with this test. We also used the curve fitting tool from the statistical toolbox of MATLAB, which gave results as shown in Figure 2. The result is in support of our assumption to take the distribution of correlation coefficients as a normal one.

2.2.2 Feature Selection

The basic criterion for choosing a feature is that its value should help us to discriminate between the classes. However, it is known that the optimal choice of features is an NP hard problem and thus, some heuristics is applied to perform the selection. In this paper we check for the discriminating power of the feature. Qualitatively, we can look upon the discriminating power of a feature as follows: If we use only one feature for building a classifier, then the efficiency of the resulting classifier will give the discriminating power of that feature. Thus, the quantity that we are interested in the misclassification area for each feature i.e. the correlation value of a particular channel pair. This can be obtained manually or by calculation. After plotting the normal curve of the correlation values obtained for a particular channel pair of different classes, feature can be selected by either looking for well separated curves or by calculating the misclassified area. This can be achieved as follows: For a given channel pair (i.e., for a given possible feature) we get two normal distributions corresponding to the two classes that we are trying to classify. The misclassification will be the area of overlap between these two normal distributions. To find this area first we have to find the point at which both curves cross each other. This is called the decision boundary of two curves. While plotting the curve all plots did not give clear decision boundary. Some channel pairs gave plots that had more than one decision boundary Figure 3(b). These are obviously cases of high misclassification and were rejected outright as good features. There are some of the samples that get misclassified because they lie in the common area. So these are responsible for error in classification.

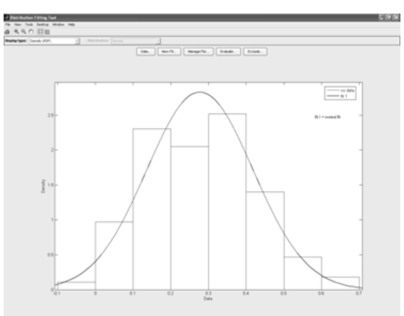


Figure 2. Plot of normal distribution of correlation coefficients between electrode 11 and 30 (highly correlated electrodes) as per the curve fitting tool in statistical toolbox of MATLB.

Here misclassified area of one class can be found out using formula

$$\text{error } 1 = \int_{-\infty}^{x_{i,j}} \frac{1}{\sqrt{2\pi{\sigma_j}^2}} e^{\frac{-(x-\mu_j)^2}{2{\sigma_j}^2}} dx \tag{5}$$

Where

$x_{i,j}$ is intersection of the normal distributions of two different classes.

${\sigma_j}^2$ is standard deviation of normal distribution of class j from mean over all trials.

μ_j is mean of normal distribution of class j over all trials.

Similarly, misclassified area of another class is

$$\text{error } 2 = \int_{x_{i,j}}^{\infty} \frac{1}{\sqrt{2\pi{\sigma_i}^2}} e^{\frac{-(x-\mu_i)^2}{2{\sigma_i}^2}} dx \tag{6}$$

is standard deviation of class i from mean over all trial.

μ_i is mean of class i over all trial.

Therefore, the total error (colored area in figure 5) is

Error = error 1+error 2

$$\text{error} = \int_{-\infty}^{x_{i,j}} \frac{1}{\sqrt{2\pi{\sigma_j}^2}} e^{\frac{-(x-\mu_j)^2}{2{\sigma_j}^2}} dx + \int_{x_{i,j}}^{\infty} \frac{1}{\sqrt{2\pi{\sigma_i}^2}} e^{\frac{-(x-\mu_i)^2}{2{\sigma_i}^2}} dx \tag{7}$$

After finding the error for all channel pairs, results are sorted in increasing order of error. The channel pairs that have the least error are selected as a feature. It is observed that if we select the correlation values of 7 channel pair as features then we can build classifiers with low error rates.

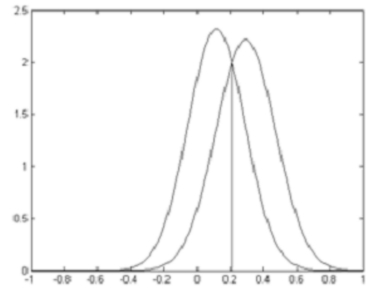
2.2.3 Classification

Once the features have been selected, we are in a position to build the classifier. In this paper we have used

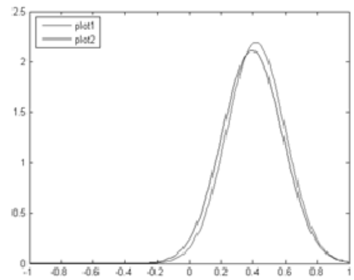
Support Vector Machine (SVM) as a classifier. We used the standard tool, SVM light was used in the present work for implementing the SVM.

3. Results and Analysis

As stated earlier, classification of EEG signal multistep process and result of each step is depends upon previous step output. If there is an error in one step then that will propagate in upcoming steps and affect the overall accuracy of classification. Therefore, result of each step is crucial and need to be analyzed after finishing each step. To check the result of all steps some basic concept were applied, for example correlation coefficients should be between -1 and +1, there should be some channels which should not have very high misclassified area (low correlated channels) and some channels should have low misclassified area (high correlated channels), there should be N*(N-1)/2 decision boundary. Another criterion of analyzing the data is that the results of previous steps give some knowledge about results of upcoming steps. Let's take an example: channel pair which is highly correlated should give low misclassified area and should be either neighborhood channels or channels which are connected in a long path way. After breaking the EEG signal data trial wise, the new data is stored in 278 files corresponding to 278 trials. Correlation coefficient matrix is a symmetric matrix because the correlation between channel X and channel Y (say) will be the same as that between channel Y and channel X. Two channels are highly correlated (anti-correlated) if correlation coefficient between then is near to +1 (-1). If correlation coefficient is near zero then these two channels are uncorrelated. If correlation coefficient between two channels is higher than that signifies they are changing together. After getting correlation coefficient among channels, their mean is calculated for each class separately and in the same way standard deviation of correlation values of each channel pair is obtained for each class. Once these parameters of the normal distribution are obtained, a curve is plotted between the normal distributions of the correlation values one channel pair of class +1 and that of class -1. Some results of plotting these normal distributions have been shown in Figure 3 to Figure 5. Results show that channels having higher value of correlation coefficient have less misclassified area while channels those have value of correlation coefficient near about zero have more misclassified area.



(a)



(b)

Figure 3. **(a)** Plot of normal distribution of correlation value between two channels of different class. **(b)** Gaussian plot having more than one decision boundary.

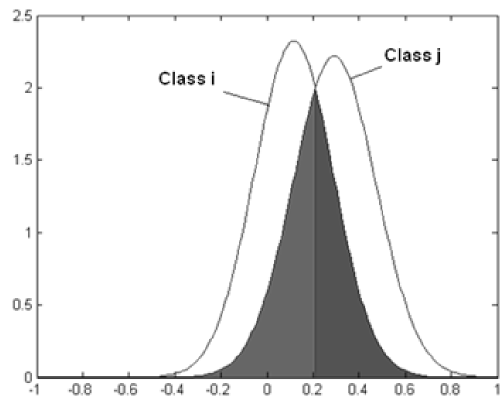


Figure 4. Misclassified area.

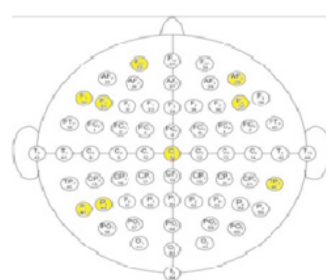


Figure 5. Shows the position of channels or electrode over scalp and the highlighted blocks indicates the electrodes selected as features.

Table 1. List of channel pairs selected as a feature

37	46
11	22
11	31
11	30
29	46
37	48
37	47

Table 2. Accuracy in classification using different number of feature

Number of Feature selected	% Accuracy
7 features	80.95
6 features	82.143
5 features	78.571
4 features	61.90
3 features	79.76
2 features	66.52

3.1 Result of Feature Selection

Decision boundary between curves corresponding to channels, is found out so it is a matrix of 64 X 64. Curves in which there are more than one decision boundary is there, their values are ignored and put zero in place of that in the matrix. We did not find a single channel pair that gave a perfect separation between the two classes. Therefore, we applied the method explained in the previous section to select the best seven features. The error due to a single feature is the area of overlap between the two normal distributions and calculated using Equation (7). The error measured only for clearly separable curves means those curves for which decision boundary is clear. The error values are arranged in increasing order of error and channel pairs which have minimum value of error are selected as a feature. We have selected 7 channels pair as a feature, shown in Table 1.

3.2 Result of Classification

We found accuracy in classification using combination of different number of features and for different marginal value. Table 2 shows the accuracy in classification using different number of features. The Table 2 shows maximum accuracy in classification by using six features.

4. Conclusion

The paper proposes the classification of EEG signal using correlation coefficient among channels as a method of features. It emphasizes on the physiological aspect of EEG signal. Here we calculate the correlation among channel pair followed by plotting Gaussian curve among channel pairs of same class. Then on the basis of misclassified area some features are selected and used for SVM training. Results truly encourages for experiments on other datasets including multi class datasets.

5. Future Scope

In future we want to experiments the same methodology on other binary class datasets and multi class datasets. Data can be more refined by improving the preprocessing. So we are looking to find out improvement in preprocessing step.

6. References

1. Alomari MH, AbuBaker A, Turani A, Baniyounes AM, Manasreh A. EEG mouse: A machine learning-based brain computer interface. International Journal of Advanced Research in Computer Science and Applications. 2014; 5(4):1–6.
2. Lotte F, Congedo M, Lecuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain–computer interfaces. Journal of Neural Engineering. 2007 Jun; 4(2):1–13.
3. Soman S. High performance EEG signal classification using classifiability and the twin SVM. Applied Soft Computing. 2015 May; 30:305–18.
4. Xiao D, Mu Z, Hu J. Classification of motor imagery eeg-signals based on energy entropy. IEEE Proceedings of the International Symposium on Intelligent Ubiquitous Computing and Education; 2009 May. p. 61–4.
5. Ke L, Li R. Classification of EEG signals by multi-scale filtering and PCA. Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems; 2009 Nov. p. 362–6.
6. Oveisi F. EEG signal classification using nonlinear independent component analysis. Proceedings of ICASSP IEEE International Conference on Acoustics, Speech and Signal; 2009 Apr. p. 361–4.
7. Skinner BT, Nguyen HT, Liu DK. Classification of EEG signals using a genetic-based machine learning classifier. Proceedings of IEEE Conference of Engineering Medicine Biology Society. 2007 Aug. p. 3120–3.

8. Anderson C, Stolz E, Shamsunder S. Multivariate autoregressive models for classification of spontaneous electroencephalogram during mental tasks. IEEE Transactions on Biomedical Engineering. 1998 Mar; 45(3):277–86.
9. Vatankhah M, Yaghubi M. Adaptive neuro-fuzzy inference system for classification of EEG signals using fractal dimension. 3rd UKSim European Symposium on Computer Modeling and Simulation; 2009 Nov. p. 214–8.
10. Yazdani A, Ebrahimi T, Hoffmann U. Classification of EEG signals using dempster shafer theory and a K-nearest neighbor classifier. Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering Antalya; Turkey. 2009 Apr-May. p. 327–30.
11. Zhovna I, Shallom ID. Automatic detection and classification of sleep stages by multichannel EEG signal modeling. Proceedings of 30th Annual International IEEE EMBS Conference; Vancouver, British Columbia, Canada. 2008Aug. p. 2665–8.
12. Blankertz B, Muller K, Krusienski D, Schalk G, Wolpaw J, Schlogl A. The BCI competition III: Validating alternative approaches to actual BCI problems. IEEE Transaction on Neural System Rehability Energy. 2006 Jun; 14(2):153–9.