ISSN 1392–124X (print), ISSN 2335–884X (online) INFORMATION TECHNOLOGY AND CONTROL, 2013, Vol. 42, No. 3

Data Compression of EEG Signals for Artificial Neural Network Classification

Darius Birvinskas[1], Vacius Jusas[2], Ignas Martišius[1], Robertas Damaševičius[1]

[1] *Software Engineering Department, Kaunas University of Technology*
e-mail: darius.birvinskas@ktu.lt, ignas.martisius@ktu.lt, robertas.damasevicius@ktu.lt

[2] *Behavioral Medicine Institute, Lithuanian University of Health Sciences*
Software Engineering Department, Kaunas University of Technology,
e-mail: vacius.jusas@ktu.lt

crossref http://dx.doi.org/10.5755/j01.itc.42.3.1986

Abstract. Brain – Computer Interface (BCI) systems require intensive signal processing in order to form control signals for electronic devices. The majority of BCI systems work by reading and interpreting cortically evoked electro-potentials across the scalp via an electro encephalogram (EEG). Feature extraction and classification are the main tasks in EEG signal processing. In this paper, we propose a method to compress EEG data using discrete cosine transform (DCT). DCT takes correlated input data and concentrates its energy in just first few transform coefficients. This method is used as feature extraction step and allows reducing data size without losing important information. For classification we use feed forward artificial neural network. Experimental results show that our proposed method does not lose the important information. We conclude that the method can be successfully used for the feature extraction.

Keywords: brain – computer interface; discrete cosine transform; data compression.

1. Introduction

Brain – Computer Interface (BCI) is a fast emerging technology, which aim is to create communication channel between the human brain and an external device. This channel translates user's intentions into control commands for specific electronic device. There are many applications for this kind of communication including prosthesis control without using muscles, robot control and game development.

Like any communication and control system, a BCI system has an input, an output, and a translation algorithm that converts the former to the latter [1]. The algorithms are usually divided to feature extraction and classification. Feature extraction generates the feature vectors with data useful for classification, while noises are eliminated. Classification attempts to assign each input value to one of given set of classes.

Most BCI systems provide modes of communication based on electro encephalogram (EEG) [2]. The EEG signal has become the main data source of BCI study due to its low cost and non-invasive nature. It is possible to record different electrical potential from the nervous system of a human, depending on mental task. These potentials may be visual evoked, event related, motor evoked or others. Motor imagination is an often used mental task for BCI purposes. Its popularity is based on a fact that imagination of movements is a very intuitive task [3]. Because EEG signals are non-stationary and non-linear, and normally interfered by eye movements and muscle noises, it is difficult to differentiate the classes of mental tasks from EEG [4]. Besides that, amount of raw data needed for classification is impractical for most machine learning algorithm.

Therefore, feature extraction is necessary for successful classification. A variety of feature extraction methods exist for BCI applications, such as Fast Fourier Transformation (FFT) [5], Principal Component Analysis (PCA) [6], Independent Component Analysis (ICA) [7]. Usually the raw EEG data are being filtered with low-pass filters for de-noising before using the above-mentioned methods [8].

Another way to extract features is the use of Discrete Wavelet Transform (DWT). This transform was designed specifically to work with signals that have non-stationary properties, and, due to the characteristics of EEG, it is a very suitable tool for feature extraction in BCI designs. Qin *et al.* successfully used DWT to extract information relating to motor imagery brain activity from EEG signals using symmetric electrode pairs [8].

Discrete cosine transform (DCT) is a transformation method for converting a time series signal into basic frequency components. It is widely used for data compression. An important feature of DCT, the feature that makes it so useful in data compression, is that it takes correlated input data and concentrates its energy in just first few transform coefficients [4]. So far DCT has been widely used for audio and image compression [9] and signal filtering [10].

In the context of BCI, discrete cosine transform was used to compute maximum, minimum and mean value of EEG signal [11], [12]. This approach does not directly use DCT coefficients for classification and requires data filtering for high frequency noises.

The rest of the paper is organized as follows. We present a method of discrete cosine transform in Section 2. We discuss the data classification using artificial neural networks in Section 3. We present the materials and methods used for the experiment in Section 4. We report the results of the experiment in Section 5. We finish with conclusions and outline future work in Section 6.

2. Discrete Cosine Transform

DCT is a transformation method for converting a time series signal into basic frequency components. Low frequency components are concentrated in first coefficients and high frequency – in last ones. The one-dimensional DCT for a list of N real numbers is expressed by the following formula:

$$Y(u) = \sqrt{\frac{2}{N}}\alpha(u)\sum_{x=0}^{N-1} f(x)\cos\left(\frac{\pi(2\mathrm{x}+1)\cdot u}{2\mathrm{N}}\right) \qquad (1)$$

where

$$\alpha(j) = \begin{cases} \frac{1}{\sqrt{2}} & if\ \ j = 0 \\ 1 & if\ \ j \neq 0 \end{cases}$$

The input is a set of N data values (EEG samples, audio samples, or other data) and the output is a set of N DCT transform coefficients $Y(u)$. The first coefficient $Y(0)$ is called the DC coefficient and it holds average signal value. The rest coefficients are referred to as the AC coefficients (these terms have been inherited from electrical engineering) [4].

DCT exhibits good energy compaction for highly correlated signals. If the input data consist of correlated quantities, then most of the N transform coefficients produced by the DCT are zeros or small numbers, and only a few are large (normally the first ones). Compressing data with the DCT is therefore accomplished by quantizing the coefficients. The small ones are quantized coarsely (possibly all the way to zero) and the large ones can be quantized finely to the nearest integer.

Applying this feature for EEG signals allows compressing useful data to the first few coefficients. Therefore, only these coefficients can be used for classification using machine learning algorithms. This kind of data compression may dramatically reduce size of input vector and decrease time required for training.

In this paper, we propose to use first few per cents of DCT transform coefficients for EEG signal classification. This approach has several advantages:

- Reduces size of data set;
- Solves the problem of data filtering for high frequency noise;
- Decreases time for artificial neural network training.

3. Classification using artificial neural network

Artificial neural network (ANN) is a mathematical model that mimics some functional aspects of biological neuron network [13]. The ANN consists of an interconnected group of artificial neurons. These neurons are basic computational elements, often called either nodes or units. The node receives input from some other nodes or from an external source. Each input has an associated weight w_{ij}, which can be modified so as to model synaptic learning. The node computes some function f of the weighted sum of its inputs:

$$y_i = f(\sum_{j=0}^{m-1} w_{ij} x_j) \qquad (2)$$

where i is node number, x_j is input value, w_{ij} is element of weight matrix W, m is number of node inputs.

Single neuron is a relatively simple computational element so many neurons have to be connected into network in order to complete complex tasks. Two or more neurons can be combined in a layer, and a particular network could contain one or more such layers. Each output of one layer is interconnected to input of next layer through the weight matrix W. Usually ANN has one input layer, one output layer and one or more hidden neuron layers. Theoretically network with one hidden layer of neurons can solve task of any complexity [13].

In most cases, an ANN is an adaptive system that changes its internal structure by changing weight values. However, if neural network either has too many neurons or is being training too long, it may start to mimic input data. In this case, classification accuracy for unknown data is decreasing, as training time increases.

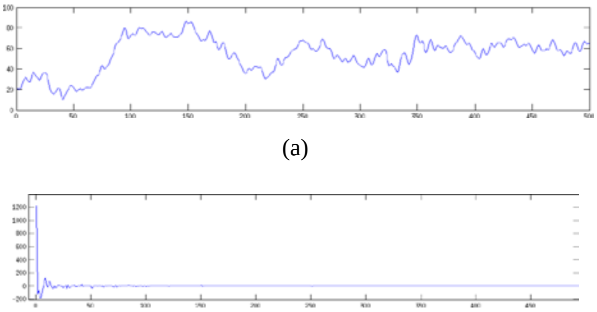
The classification accuracy in this paper is measured as a ratio of false results to the total number of trials:

$$err = \frac{n - n_{true}}{n} * 100 \qquad (3)$$

where n is the total number of trials, n_{true} is the number of trials with correct classification result.

4. Materials and methods

For experiments, we used Data et IV from BCI competition II datasets, provided by Berlin: Fraunhofer-FIRST, Intelligent Data Analysis Group and Freie Universitat Berlin, Department of Neurology, Neurophysics Group (http://www.bbci.de/competition/ii/). The data sets were taken from a healthy subject during no-feedback session. During recording EEG potentials of motor imaginary activity were taken. The task was to press the corresponding keyboard keys with left or right hand index fingers in a self-chosen order and timing. Data set consists of 416 trials of 500 ms length (sampling rate 1000 Hz) each ending 130 ms before a key press. 316 epochs are for training (labelled 0 for upcoming left hand movements and 1 for upcoming right hand movements). The remaining 100 trials are for testing.



(a)

(b)

Figure 1. a) sample of raw EEG data, b) sample of EEG data after DCT transform.

In Figure 1a, an example of raw EEG data in time domain is shown. The signal is non-linear and non-periodic. This means that no data can be excluded from classification using ANN. Figure 1b shows discrete cosine transform of EEG data. Resulting signal in frequency domain is less complex, useful data are compressed into first few coefficients while last ones are close to zero. These small coefficients can be omitted from classification.

Classification of data was performed using the MATLAB Neural Network Toolbox. A feed forward network with a pure line transfer function in last neuron was created (Figure 2) and used for the experiment.

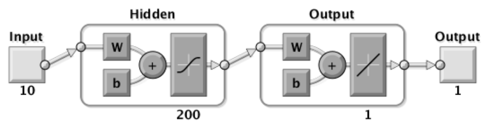


Figure 2. Architecture of artificial neural network

Since neural network is initialized with random weight values, experiments must be performed several times and average result must be taken.

The classification accuracy depends on the number of neurons in the hidden layer and the size of input vector.

In this experiment, we chose first 10 DCT coefficients for classification, ignoring the rest of DCT data. DCT data plot (Figure 1b) shows that all the other coefficients are less than our predefined accuracy. Since signal has 500 samples, we are using only 2% of original. Also, this method takes advantage of DCT feature to concentrate low frequency components in first coefficients. It means that there is no need of using any filters for high frequency noises.

The number of neurons in hidden layer was set to 10, 20, 50, 100, 150, 200 and 250. With each chosen value, ten artificial neural networks were created and average classification accuracy measured.

5. Results of experiment

All computations in this experiment were carried out using MATLAB. The method of DCT was applied to convert the time domain data to frequency domain. Then the first 10 coefficients of every trial were taken for further investigation. This procedure effectively compressed data of the original set.

Table 1. Results of experiment

No. of neurons	Average error	Minimum error
10	32.943	27.848
20	31.854	21.202
50	31.044	22.468
100	26.708	18.670
150	21.392	14.924
200	20.586	14.873
250	21.316	14.940

Resulting data set was divided into subsets for training (70%), validation (15%) and testing (15%). Afterward, training using different topologies of feed forward artificial neural network with respect to variations in number of neurons in hidden layer was carried out.

One can observe in Table 1 that classification results of test subset are constantly increasing until the best possible value (200 hidden neurons) is achieved. Afterwards results start to decrease, when the number of neurons increases. This occurrence is called overfitting. Overfitting is a phenomenon when a neural network gets very good at dealing with one data set at the expense of becoming very bad at dealing with other data sets.

Column "Minimum error" shows result of best performing network in accordance to the number of hidden neurons.

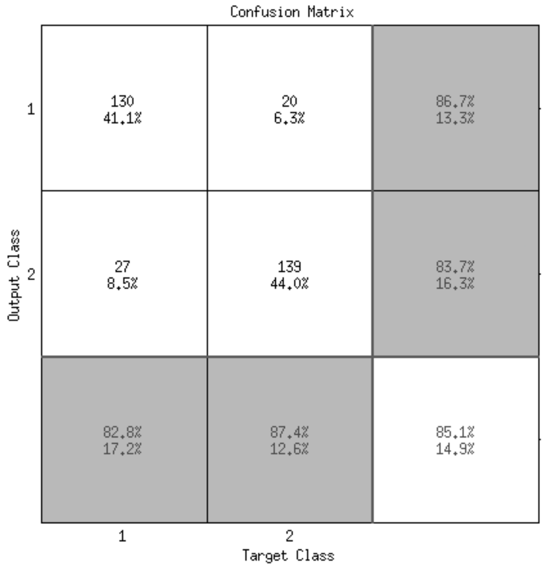


Figure 3. Confusion matrix

In Figure 3, individual classification accuracy and overall accuracy of the best performing network are shown. The obtained results show that data compression, using discrete cosine transform, can be used for BCI feature extraction.

During original BCI Competition II the best classification accuracy with this data set was obtained 16%. Second and third places were 19% and 23%, respectively. Comparing these results with ours, we can claim that our method performs better and it can compress EEG signal data without losing important information.

6. Conclusions and future work

In this paper, we proposed to use data compression as a feature extraction step for BCI systems. This method allows reducing size of data vector without losing important information. Classification was carried out using feed forward artificial neural network. Several architectures with different number of hidden neurons were used to achieve the best possible classification result.

Future research work may include using fast DCT computation algorithms to decrease computation time. Also, other classification algorithms may increase the classification accuracy.

References

[1] **J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, T. M. Vaughan**. Brain–Computer Interface technology: a review of the first international meeting. In: *IEEE Trans. Rehabil. Eng*., 2000, Vol. 8, No. 2, 164-173.

[2] **A. Vuckovic, F. Sepulveda.** A four-class BCI based on motor imagination of the right and the left hand wrist. In: *First International Symposium on Applied Sciences on Biomedical and Communication Technologies, ISABEL '08*. Aalborg, DK, 25-28 Oct. 2008, 4 pages. DOI: 10.1109/ISABEL.2008.4712628

[3] **L. Guo, Y. Wu, L. Zhao, T. Cao, W. Yan, X. Shen**. Classification of mental task from EEG signals using immune feature weighted support vector machines. In: *IEEE Trans. on Magnetics*, 2011, Vol. 47, No. 5, 866-869.

[4] **D. Salomon**. *Data compression. The complete reference*, 3rd Edition. Springer, 2004.

[5] **S. Lemm, B. Blankertz, G. Curio, K.-R. Müller**. Spatio-spectral filters for improving classification of single trial EEG. In: *IEEE Trans. Biomed. Eng., Sep*. 2005, Vol. 52, No. 9, 1541–1548.

[6] **G. Gage, K. Ludwig, K. Otto, E. Ionides, D. Kipke**. Naive coadaptive cortical control. *J. Neural Eng.*, 2005, Vol. 2, 52–63.

[7] **L. Qin, L. Ding, B. He.** Motor imagery classification by means of source analysis for brain computer interface applications. *J. Neural Eng.*, 2004, Vol. 1, 135–141.

[8] **L. Qin, B. He**. A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications, *J. Neural Eng.*, 2005, Vol. 2, No. 4, 65–72.

[9] **D. Kancelkis, J. Valantinas**. A new le Gall wavelet-based approach to progressive encoding and transmission of image blocks. *Information Technology and Control*, 2011, Vol. 41, No. 3, 239-247.

[10] **P. Orlowski**. Simplified design of low-pass, linear parameter-varying, finite impulse response filters. *Information Technology and Control*, 2010, Vol. 39, No. 2, 130-137.

[11] **Ch. Aparna, J. V. R. Murthy, B. R. Babu**. Energy computation for BCI using DCT and moving average window for noise smoothening. *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, 2012, Vol. 2, No. 1, 15–21.

[12] **Ch. Aparna, J. V. R. Murthy, B. R. Babu, Ch. S. Rao**. Reducing dataset size in frequency domain for brain computer interface motor imagery classification. *International Journal of Computer Science and Engineering (IJCSE)*, 2010, Vol. 2, No. 9, 2924-2927.

[13] **S. O. Haykin**. *Neural networks and learning machines*. Prentice Hall, 2008.

Received June 2012.