SINGLE-TRIAL EEG CLASSIFICATION FOR BRAIN-COMPUTER INTERFACE USING WAVELET DECOMPOSITION

Y. P. A. Yong, N. J. Hurley, and G. C. M. Silvestre

Department of Computer Science, University College Dublin
Belfield, Dublin 4, Ireland
web: www.ihl.ucd.ie, www.vimspri.ucd.ie

ABSTRACT

A classification system for EEG signals using wavelet decomposition to form the feature vectors is developed. Single-trial analysis loses the benefit of averaging to remove non-task related brain activity and makes it more difficult to pick out key features determining the execution of a task. Wavelet analysis is used here to localise the event-related desynchronization of voluntary movement. Classification of a self-paced typing experiment was made using wavelets for the feature selection and SVMs for the classification of feature vectors. Results of up to 91% classification accuracy were obtained, proving that wavelets are an effective tool, and the use of wavelets will be considered in more complex work.

1. INTRODUCTION

Research has shown that a large number of cognitive tasks have features that can be observed by electroencephalography (EEG), but only after averaging the results of several hundred replications of the same task. In order to be applicable to BCI systems, one must be able to detect cognitive activity from a single performance of a task, without the benefit of averaging. This is referred to as single-trial analysis. A single trial is obscured by non-task related activity in the brain and noise, making it more difficult to extract the relevant information from the signal

The investigation of features in the EEG signals requires a detailed time-frequency analysis. Wavelet analysis comes into play here since wavelets allow decomposition into frequency components while keeping as much time information as possible. Wavelet analysis have been used in many applications such as data compression, image processing and digital watermarking. Much work has been done on wavelets in the mathematical and theoretical areas, however, there is little transfer of this knowledge to the solution of real-life problems. Some work has been done with wavelets on medical data and EEG signals [6, 14], but there are more applications that can be explored and exploited.

Here, we apply the use of wavelets to the problem of EEG signal classification. Commonly used techniques in this area such as Fourier Transform (FT) have shortcomings when dealing with transitions and trends in the signal such as drifts and abrupt changes. Unfortunately, these are inherent in brain signals. The problem of feature extraction is further compounded by attempting to analyse the signals in a single-trial environment, losing the benefit of averaging out noise and other unrelated brain activity. Wavelets are designed especially for the analysis of non-stationary signals, and so are an ideal tool to work with EEG signals. We applied wavelet decomposition to a dataset of EEG signals obtained in experiment in order to extract the task-related features, achieving high classification accuracy with no preprocessing of the signals needed.

In this paper, we will present the present state of the art of BCI systems. Some background and technologies involved in this project will be explained. The experimental procedure and results are shown, and some discussion is made into future work and directions.

2. STATE OF THE ART

Due to the fact that EEG data can be easily obtained and recorded with comparatively inexpensive equipment, it is the main data source for many BCI programs in existence. EEG-based BCI systems can be subdivided into several groups. Dependent BCIs depend on the oculmotor control of the gaze direction: a subject has to be able to control what he wants to look at, such as Visual Evoked Potentials (VEPs). Lalor et al [11] have created an effective EEG-based BCI design for an immersive 3-D game using steady-state visual evoked potentials (SSVEPs). Independent BCI system types, which work regardless of vision control include Event Related Potentials (ERP), which occur as small fluctuations in EEG recordings in response to the presentation of a single stimulus, either sensory, motor or a mental event. The P300 evoked signal is used in Jessica Bayliss' work in a virtual reality environment [1]. Subjects learn to control μ and/or β-rhythm amplitudes in the Wadsworth Centre BCI Research and Development Program to move a cursor in one or two dimensions to choices on a computer screen [16]. Slow Cortical Potentials (SCP), which are voltage shifts generated in the cortex lasting over 0.5-10 seconds, are harnessed in the Thought Translation Device (TTD) developed by Hinterberger et al at the Max Planck Institute where with training, a patient can learn to generate the appropriate SCP signals to control cursor movements on a notebook computer running word-processing software as part of the TTD device [8].

The data acquired by Blankertz et al [2] in the Berlin BCI group that is used in this study consists of a self-paced typing exercise with no feedback, producing the ERP of a motor stimulus. They have worked on this dataset using Fast Fourier Transform (FFT) techniques for their preprocessing and feature selection and used a variety of techniques for the classification, obtaining very high results of up to 96.9%. Other groups have also used this data for their classification work. Kelly et al [9] has applied an autoregressive model approach to the classification problem, achieving accuracies of up to 70.7%, while Garrett et al [7] used genetic algorithms (GA) on this data to determine the subset of features of the data that are useful towards the classifcation process, obtaining 76% accuracy.

In work involving wavelet analysis and EEG signals, Der

Band	Frequency[Hz]	Occur while / Indicate
δ	0.5 - 3.5	Movement preparation
θ	3.5 - 8	Memory
α (μ)	8 - 13	Relaxation, sensory idling
β	13 - 22	Motor idling
γ	22 - 40	Feature binding

Table 1: Frequency bands of brain activity

et al investigate in detail the use of wavelet analysis for cognitive processes [6]. The use of wavelets in the medical field has mostly been confined to the area of noise reductions of acquired data, in EEG [13], and other signal types such as electromyography (EMG) [12].

3. BACKGROUND

3.1 Event-related synchronisation/desynchronisation

Physiologically meaningful signal features can be extracted from various frequency bands of recorded EEG. Movement preparation followed by execution, or even only motor imagination is usually accompanied by a power decrease in certain frequency bands, labeled as event-related desynchronization (ERD), in contrast, their increase after a movement indicates relaxation and is due to a syncrhonization in firing rates of large populations of cortical neurons (ERS). A negative "Bereitschaftspotential" (BP) precedes the voluntary initiation of the movement, with different scalp potential distribution reliably demostrated in a majority of experimental subjects. The BP is a time-locked response to the movement event. Table 1 shows the frequency bands and the neurological features that they are associated with [10].

3.2 The wavelet transform

Fourier analysis has the serious drawback that transitory information is lost in the frequency domain. This may be alleviated in the use of a short-time Fourier Transform (STFT), but when a time window is chosen, that window is the same for all frequencies of the signal being analysed, causing possibly essential information to be lost at very low or high frequencies. The essential advantage of the wavelet transform over Fourier transform or STFT is that the time-frequency window is flexible and adapts in such a way that there is always about the same number of periods of the frequency analysed in the time window. Wavelet analysis allows the use of long time-intervals for more precise low-frequency information, and shorter regions for more high-frequency information.

The wavelet transformation is achieved by the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet. A wavelet is a waveform of effectively limited duration with zero mean. Regardless of its mathematical properties, a basic requirement of the wavelet used is that it looks similar to the signal patterns to be localized. Local features can be described better with wavelets that have local extent. It minimises spurious effects in the reconstruction of the signal via the inverse wavelet transform.

Wavelets are able to determine if a quick transitory signal exists, and if so, can localise it. This feature makes wavelets very useful for the study of the EEG waveforms.

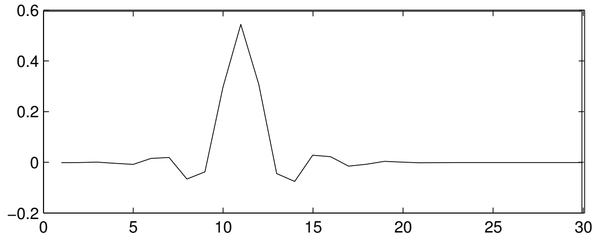

Figure 1: Wavelet (Coiflet order 5) used in decomposition

4. METHOD

4.1 Data Acquisition

In this EEG data from [2], execution of the typing is voluntary and without explicit external sensory input at a self-paced rate. The typing movement is actually executed to increase the BP signal strength, optimising the signal-to-noise ratio. The subject is also not engaged in the unnatural condition of vetoing the movement while preparing to execute the motor command.

EEG signals were recorded from one subject in 3 sessions with some minutes' breaks in between. The subject was sitting in a normal chair, relaxed arms resting on the table, fingers in the standard typing position at the computer keyboard (index fingers at 'f','j' and little fingers at 'a',';'). The task was to press the aforementioned keys with the corresponding fingers in a self-chosen order and timing ('self-paced key typing'). A total of 516 keystrokes was done at an average speed of 1 key every 2.1 seconds. 3 events have been rejected due to heavy measurement artifacts. Brain activity was measured with 27 Ag/AgCl electrodes referenced to nasion at 1000 Hz using a band-pass filter from 0.05 to 200 Hz. The timing of keystrokes was stored along with the EEG data. EMG and EOG signals were also recorded (but are not supplied).

The supplied data consists of 27 EEG channels in 516 single trials. Windows 1500 ms long were cut out of the continuous raw signals each ending at 120 ms before the respective keystroke, as from that point on there is EMG activity in an significant number of trials, which produce serious artifacts in the data.

4.2 Signal Classification

4.2.1 Feature Extraction

The signal used in the experiment was of 1000 Hz frequency. Only 6 channels out of the full available set of 27 are used, as they were found to be sufficient to ensure a high level of classification, and they are in the centro-parietal region (C3, CP3, P3, C4, CP4 and P4).

The wavelet used for the decomposition was Coiflets order 5 (Fig. 1), which was deemed to be closest in resemblance to the signal waveforms under consideration. This family of wavelets was built by Daubechies at the request of Coifman, the wavelet function has $2N$ moments equal to 0 and the scaling function has $2N-1$ moments equal to 0 [5].

To create the feature vectors for each trial, wavelet decomposition was performed on each signal A sample wavelet decomposition of a signal is shown in figure 2. Each decomposition level (detail) corresponds a breakdown of the main signal to a frequency bandwidth. Various decomposition levels were selected to form the feature vector, and if more than

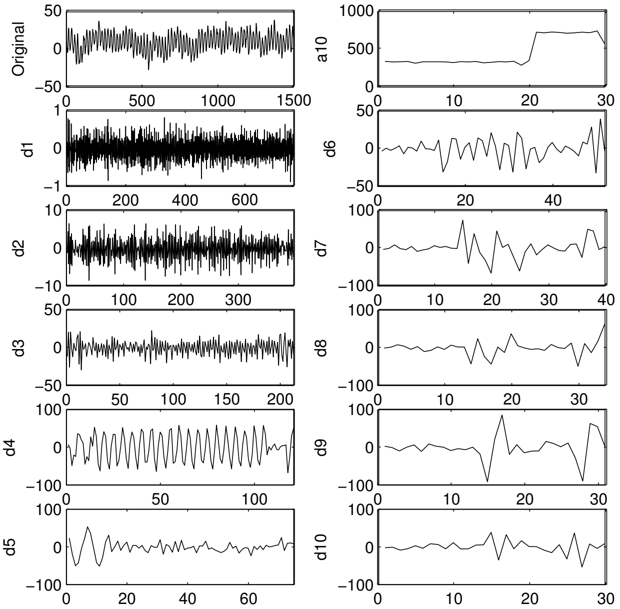

Figure 2: Co-ordinates of a sample wavelet decomposition

one level was selected, the coefficients of those levels were combined. The results from each channel were then combined to form an overall feature vector for one trial in the classification.

4.2.2 *Classification Algorithm*

A primary motivation behind Support Vector Machines (SVMs) is to directly deal with the objective of good generalisation by simultaneously maximising the performance of the machine while minimising the complexity of the learned model.

An important property of SVMs is that their ability to learn can be independent of the dimensionality of the feature space. With SVMs, the complexity of the optimization problem is based on the margin with which they separate the data, not with the number of features. SVMs also have the potential to handle large feature spaces since they use overfitting protection.

Another advantage of SVMs is the ability to learn different kernel functions for classification. In their basic form, SVMs learn the linear threshold function. However, by changing to an appropriate kernel function, they can be used to learn polynomial classifiers, radial basic function (RBF) networks and three-layer sigmoid nets.

The SVM theory was developed by Vapnik [15]. The fundemental idea behind SVMs is to separate data $X \subset \ {}^n$ from two classes by finding a weight vector $w \in \ {}^n$ and an offset $b \in$ of a hyperplane such that

$$\begin{aligned} H: \quad & {}^n \rightarrow \{-1,1\} \\ & x \mapsto sign(w \cdot x + b) \end{aligned} \tag{1}$$

with the largest possible margin. Given a training set of instance-label pairs (x_i, y_i), $i = 1, ..., l$ where $x_i \in \ {}^n$ and $y \in \{-1,1\}^l$, the SVM consists of solving the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi} \quad & \tfrac{1}{2}\|w\|_2^2 + C \ {}_{i=1}^{l} \xi_i \\ \text{subject to} \quad & y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0. \end{aligned} \tag{2}$$

The parameters ξ_i are called slack variables and ensure that the problem has a solution in case the data is not linearly separable. $C > 0$ is the penalty parameter of the error term, which controls the trade-off between a low training error eg ξ_i^2, and a large margin $\gamma(X,Y,C) = 1/\|w\|_2^2$. Cover's theorem on the separability of patterns [4] says that data cast nonlinearly into a high-dimensional feature space is more likely to be linearly spearable than in a lower-dimensional space. The SVM still produces a linear decision function even though the function is now linear in the feature space and not in the input space, and is capable of producing arbitrary decision functions in the input space, depending on the kernel function $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$.

For classification into right and left hand movement, the implementation called LIBSVM by Chang and Lin [3] was used. The classification is implicitly two-class, and the SVMs are considered on a radial basis where the kernel function is $K(x_i, x_j) = exp(-\gamma \times |x_i - x_j|^2)$. The optimal values for γ and C are manually fine-tuned to give the best known classification for each test set. The training set size used to train the SVM was 413, and the classifier tested out on a set of 100 trials.

5. RESULTS

The results are shown in table 2. Classification is successfully achieved within a whole range of decomposition levels and thus their frequency band equivalents, with rates ranging from 57% to 91%, even with the use of only 6 channels out of the possible 27. The most effective results were found to correspond to the frequency band of 1-8 Hz, which is lower than the α and β bands and corresponds well with the observation of BP occurance.

In comparison, when the algorithm as described in [2] was used, with preprocessing and downsampling to obtain the features, and using SVMs as the classifier yielded an accuracy of 94% while using all 27 channels, while using only the 6 channels gave a classification accuracy of 91%.

Other groups who have worked on this material using different methods include Kelly et al [9], who applied an autoregressive model in the feature extraction stage and linear discriminants for the classification, achieving accuracies of up to 70.7%.

Garrett et al [7] used genetic algorithms (GA) to determine the subset of features of the data (broken up into frequency bands and time windows) that are useful towards the classifcation process, using SVMs to test and train the features. They obtained 76% accuracy.

Table 3 provides a summary of the comparative accuracy of classification results.

6. CONCLUSION

Wavelet decomposition has been shown to be effective in creating feature vectors for the single-trial classification of EEG signals. It has been able to isolate effectively the features that are otherwise obscured by noise and other artifacts without

Decomp. Level	Freq. Range[Hz]	Classification
D6	16.13 - 32.25	57%
D7+D8	4.03 - 16.13	61%
D7+D8+A8	0 - 16.13	66%
D8	4.03 - 8.06	69%
D7	8.06 - 16.13	71%
D8+D9	2.02 - 8.06	78%
D8+D10	1.01 -2.02, 4.03 - 8.06	83%
D9+D10	1.01 - 4.03	85%
D8+D9+D10	1.01 - 8.06	91%

Table 2: Frequency range vs. accuracy of classification

Group	Classification Accuracy
Yong et al	91% [a]
Blankertz et al [2]	91% [a], 94% [b]
Kelly et al [9]	70.7% [b]
Garrett et al [7]	76% [a]

Table 3: Summary of accuracy of classification results ([a]: 6 channels, [b]: all 27 channels)

the benefit of averaging, allowing for effective classification. The properties of wavelets are highly suited toward their use in this area of brain signal analysis. The effectiveness of a single-trial classification system is an important step towards online classification of signals, allowing for direct feedback and interaction with the user and the development of a feasible BCI system. This is definitely an important area to be developed further.

Further work is underway to refine the method, using further data, a downsampled version at 100 Hz and exploring some preprocessing techniques to refine the feature extraction process in order to develop a system which is fast to calculate and reliable. Ultimately, it is hoped that distinct features characteristic of each movement can be extracted that will forecast whether a movement is actually taking place in a real-time situation and reliably determine what that movement is.

7. ACKNOWLEDGEMENT

The authors would like to acknowledge the use of the data acquired by Blankertz et al at Fraunhofer FIRST and the Neurophysics Group of the Free University in Berlin [2].

This work was partially funded by European Commission through the IST Programme under Contract IST-2002-507609 SIMILAR.

REFERENCES

[1] J. D. Bayliss and D. H. Ballard. Single trial p300 recognition in a virtual environment. In *CIMA'99 (Soft Computing in Biomedicine)*, Rochester, NY, June 22-25 1999. CIMA'99.

[2] B. Blankertz, G. Curio, and K. R. Müller. Classifying single trial EEG : Towards brain computer interfacing. In T. G. Diettrich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems (NIPS 01)*, volume 14. MIT Press, 2002.

[3] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[4] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions of Electronic Computing*, EC-14:326 –334, 1965.

[5] I. Daubechies. Ten lectures on wavelets. In *CMBS*, pages 258–261. SIAM, 1994.

[6] R. Der and U. Steinmetz. Wavelet analysis of eeg signals as a tool for the investigation of the time architecture of cognitive processes. Technical Report Report Nr.4/1997, des Instituts fuer Informatik der Universitaet Leipzig, 1997.

[7] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut. Comparison of linear, nonlinear, and feature selection methods for eeg signal classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):141 – 144, 2003.

[8] T. Hinterberger, B. Wilheim, R. Veit, N. Weiskopf, T. N. Lal, and N. Birbaumer. Neural mechanisms underlying control of a brain-computer-interface (bci): Simultaneous recording of bold-response and eeg. (poster), 2004. (submitted).

[9] S. Kelly, D. Burke, P. de Chazal, and R. Reilly. Parametric models and spectral analysis for classification in brain-computer interfaces. In *Proceedings of 14th International Conference on Digital Signal Processing*, Greece, July 2002.

[10] R. Krepki, B. Blankertz, G. Curio, and K.-R. Müller. The berlin brain-computer interface (bbci): towards a new communication channel for online control of multimedia applications and computer games. In *9th International Conference on Distributed Multimedia Systems (DMS'03)*, pages 237 – 244, 2003.

[11] E. Lalor, S. P. Kelly, C. Finucane, R. Burke, R. Smith, R. B. Reilly, and G. McDarby. Steady-state vep-based brain computer interface control in an immersive 3-d gaming environment. *Journal of Applied Signal Processing*, 2004. in press.

[12] R. L. Ortolan, R. N. Mori, Jr. R. R. Pereira, C. M. N. Cabral, J. C. Pereira, and Jr. A. Cliquet. Evaluation of adaptive/nonadaptive filtering and wavelet transform techniques for noise reduction in emg mobile acquisition equipment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(1):60 – 69, March 2003.

[13] R. Q. Quiroga. Obtaining single stimulua evoked potentials with wavelet denoising. *Physica D*, 145:278 – 252, 2000.

[14] R. Q. Quiroga, O. W. Sakowitz, E. Basar, and M. Schürmann. Wavelet transform in the analysis of the frequency composition of evoked potentials. *Brain Research Protocols*, 8:16 – 24, 2001.

[15] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 2000.

[16] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and G. W. Forneris. An eeg-based brain-computer interface for cursor control. In *Electroencephalog. Clin. Neurohysiol.*, volume 78, pages 252 – 259, March 1991.