CLASSIFICATION OF MULTICHANNEL EEG DATA USING LENGTH/ENERGY TRANSFORMS

David Gutiérrez, Fabián García-Nocetti, and Julio Solano-González

Department of Computer Systems Engineering and Automation,
Institute of Research in Applied Mathematics and Systems,
Apdo. Postal 20-726 Admon. No. 20, 01000, Mexico City, Mexico.
{david, fabian, julio}@uxdea4.iimas.unam.mx

ABSTRACT

We propose the use of length and energy transforms in the classification of multichannel EEG data to identify different cognitive activity using a reduced set of recording electrodes. The length transform (LT) represents a temporarily smoothed time course of the data, while the energy transform (ET) can be interpreted as a short-term energy estimate. The transformation of the data in the length/energy domain allows to effectively preserve important data features when autoregressive (AR) models are used to reduce the dimension of the classification problem. We evaluate the performance of the LT and ET on the classification of real cognitive EEG data for the case when the optimal AR model is selected under the Schwarz's Bayesian criterion (SBC) and a Mahalanobis distance-based classifier is used. Our results show that accurate classification is achieved when the data is transformed through the LT or ET even for low-order AR models, having the LT slightly better performance.

1. INTRODUCTION

Brain-computer interfaces based on multichannel electroencephalographic (EEG) data rely on accurate classification methods in order to use brain's electrical activity to control computerized devices in real-time [1]. Recently, many linear and non-linear classifiers have been proposed in the literature, some of which report excellent performance in classifying different cognitive tasks [2]. However, most of these methods fail to justify the preprocessing techniques used to make the data suitable for the classification.

One of the most commonly used preprocessing techniques is the autoregressive (AR) modeling of the data. An AR model of order p reduces the dimensionality of the classification problem by approximating the signal with the response of filtering unit variance white noise by means of an FIR filter of order p [3]. Clearly, the order of the reduction as well as the accuracy of the classifier depend on the order of the AR model selected. However, most of the time *ad hoc* models are selected based on prior information or obtained through experimentation. Then, misclassifications produced by under- or over- estimating the model may occur. On way to choose an optimal AR model is using the Schwarz's Bayesian criterion (SBC), but this criterion does not warranty that the resulting model will be suitable for accurate signal classification.

In this paper, we propose a robust method for EEG signal modeling and classification based on the length transform (LT) or the energy transform (ET) of the data. The LT represents a temporarily smoothed time course of the data, while the ET can be interpreted as a short-term energy estimate. These transforms appear in the context of multi-channel signal processing, specifically in the problem of QRS detection from electrocardiographic (ECG) signals [4], and real-time magnetocardiographic (MCG) data analysis [5]. However, these transforms are also used for single-channel analysis, as in our case.

We show that better performance is achieved when classifying transformed data rather than raw-data, even for the case when the optimal AR model of the raw-data is used. Hence, we still can use the SBC to select a model, but using transformed data makes the classification process more robust. For this reason, we propose a classification method with the following processing steps: (i) adaptive artifact removal, (ii) preprocess data using the LT or EL, (iii) AR modeling of the transformed data, and (iv) classification based on the Mahalanobis distance.

In Section 2, we introduce the LT and ET in the context of single-channel preprocessing. Section 3 presents the classification method with all its processing steps (from AR modeling to the Mahalanobis-based classifier). In Section 4 we show the applicability of our method though numerical examples using real cognitive EEG data. In Section 5 we discuss the results and future work.

0-7803-9323-6/05/$20.00 ©2005 IEEE.

2. LENGTH AND ENERGY TRANSFORMS

In this section, we define the LT and ET for the case of single-channel signal processing. The original expressions of these transforms for the case of multi-channel processing can be found in [4].

Let us denote the signal at the mth channel as $x_m(n)$ for $n = 1, 2, \ldots, N$ time samples and $m = 1, 2, \ldots, M$ channels. Then, the LT is given by

$$l_m(n) = \sum_{k=n}^{n+q-1} |x_m(k) - x_m(k-1)| \quad (1)$$

where q is the length of the processing window. The LT represents a temporarily smoothed time course of the data, and can be interpreted as a measure of the path's length taken by signal to go from $x_m(1)$ to $x_m(N)$.

In a similar way, we can define the EL as

$$e_m(n) = \sum_{k=n}^{n+q-1} (x_m(k) - x_m(k-1))^2 \quad (2)$$

which can be interpreted as a short-term energy estimate.

The LT and ET have the advantage of being simple to implement and suitable for on-line processing, which is a desired characteristic in BCI applications.

3. THE PROPOSED METHODS

In this section, we describe each of the steps involved in the proposed classification method. These steps are:

- eye blinking artifact removal using an adaptive filter,
- data transformation using LT/ET,
- AR modeling, and
- Mahalanobis distance-based classification.

3.1. Artifact removal

The dominant artifact during EEG data recording is eye blinking. A common practice in EEG studies is to reject data when this kind of artifact is identified in measurements. However, rejecting contaminated trials causes substantial data loss, and restricting eye movements/blinks limits the experimental designs [6]. For this reason, we decided to use an adaptive filter to remove blinking artifacts. The adaptive filter uses electrooculogram (EOG) measurements, obtained simultaneously with the EEG, as an external measurement of the artifact. Then, the correlation of the measured artifact to each EEG channel is estimated in order to cancel it. This problem falls into one of the applications of the Wiener filter, which is usually referred to as *noise cancellation*. Further information about the implementation of this filter can be found in [7]. In our case, we used the least-mean-square (LMS) adaptive algorithm to implement the eye blinking artifact removal with on-line capabilities.

3.2. Data Transformation

After artifact removal, the filtered EEG data is transformed using either (1) or (2), and the results are arranged in data vectors such that

$$\boldsymbol{l}_m = [l_m(1), l_m(2), \ldots, l_m(N-q+1)]^T \quad (3)$$

and

$$\boldsymbol{e}_m = [e_m(1), e_m(2), \ldots, e_m(N-q+1)]^T. \quad (4)$$

If this procedure is repeated for each of the M EEG channels, we can collect the corresponding transforms $\boldsymbol{l}_m$ and $\boldsymbol{e}_m$ in long vectors containing the information of the full array, i.e. $\boldsymbol{l} = [\boldsymbol{l}_1^T, \boldsymbol{l}_2^T, \ldots, \boldsymbol{l}_M^T]^T$ and $\boldsymbol{e} = [\boldsymbol{e}_1^T, \boldsymbol{e}_2^T, \ldots, \boldsymbol{e}_M^T]^T$, respectively.

3.3. AR Modeling

In order to reduce the dimension of our classification problem, we compute the AR models of $\boldsymbol{l}_{(\cdot)}$ and $\boldsymbol{e}_{(\cdot)}$. The order of the model that provides an optimal approximation (denoted as p_{opt}) is selected from the raw-data according to the SBC. We choose the SBC as our model selection criterion because it has been proven to be effective in different mixture estimation and clustering applications [8].

Under theses conditions, we can reduce the dimension of the data from size $(N-q+1)M \times 1$ to $p_{\text{opt}}M \times 1$. Then, the processed data to be used in the classifier is represented as

$$\tilde{\boldsymbol{l}} = [\text{AR}(\boldsymbol{l}_1^T)_{p_{\text{opt}}}, \text{AR}(\boldsymbol{l}_2^T)_{p_{\text{opt}}}, \ldots, \text{AR}(\boldsymbol{l}_M^T)_{p_{\text{opt}}}]^T \quad (5)$$

and

$$\tilde{\boldsymbol{e}} = [\text{AR}(\boldsymbol{e}_1^T)_{p_{\text{opt}}}, \text{AR}(\boldsymbol{e}_2^T)_{p_{\text{opt}}}, \ldots, \text{AR}(\boldsymbol{e}_M^T)_{p_{\text{opt}}}]^T, \quad (6)$$

respectively, where $\text{AR}(\cdot)_{p_{\text{opt}}}$ indicates the AR model of order p_{opt}.

3.4. Mahalanobis Distance-based Classifier

The Mahalanobis distance is defined as the quantity

$$d_{\text{M}} = (\boldsymbol{y} - \bar{\boldsymbol{y}}_i)^T C_i^{-1} (\boldsymbol{y} - \bar{\boldsymbol{y}}_i), \quad (7)$$

where $\boldsymbol{y}$ is a feature vector (in our case can be either the $\tilde{\boldsymbol{l}}$ or $\tilde{e}$ data to be classified), $\bar{\boldsymbol{y}}_i$ and C_i are the mean vector and covariance matrix, respectively, of the ith class, where

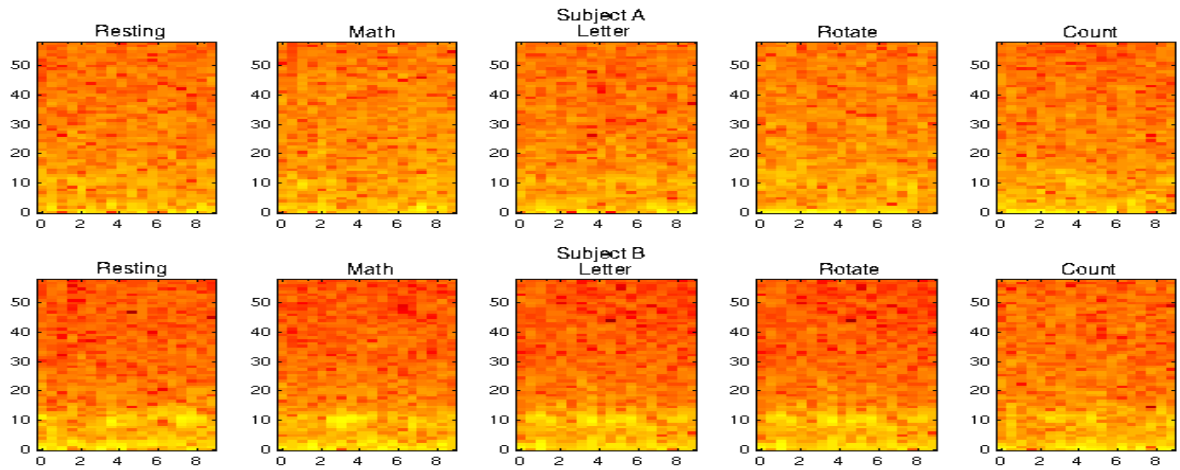

Fig. 1. Spectrograms of EEG data for two subjects performing the five cognitive tasks described in Section 4. The vertical axes correspond to frequency (in Hz), the horizontal axes to time (in seconds), and the color scale indicates energy. The figure is courtesy of C. W. Anderson.

$i = 1, 2, \ldots, I$ classes. In our case, i corresponds to different cognitive activities and the pair $\{\bar{\boldsymbol{y}}_i, C_i\}$ is computed from a set of training data at independent experiments (trials) of the corresponding cognitive activity.

It has been shown in [9] that Mahalanobis distance-based classifiers are able to detect EEG activity with affordable accuracy even when a reduced set of channels is used. Hence, we use (7) in our minimum-distance classifier as follows: After calculating $\{\bar{\boldsymbol{y}}_i, C_i\}$ for all classes using training data, we classify a feature vector $\boldsymbol{y}$ by measuring the Mahalanobis distance d_{M} to each of the classes, and assign $\boldsymbol{y}$ to the class for which d_{M} is minimum.

4. NUMERICAL EXAMPLES

We conducted a classification experiment using real EEG data from seven subjects performing five different cognitive tasks: resting (baseline measurement), compute a nontrivial multiplication problem, mentally compose a letter to a friend, rotate a three-dimensional solid, and visualize a sequence of numbers being written on a blackboard. The data used in this study is from the work of Keirn and Aunon [10] and it was collected using the following procedure: Subjects were placed in a dim, sound controlled room and electrodes were placed at positions C3, C4, P3, P4, O1, and O2 as defined by the 10–20 system of electrode placement [11] and referenced to two electrically linked mastoids at A1 and A2. The data was recorded at a sampling rate of 250 Hz with a Lab Master 12-bit A/D converter mounted in an IBM-AT computer. Eye blinks were detected by means of a separate channel of data recorded from two electrodes placed above and below the subject's left eye. Subjects were asked to perform the five separate mental tasks described above (these tasks were chosen to invoke hemispheric brainwave asymmetry). Each task was recorded for 10 seconds and the experiment was repeated five times. As an example of the data collected, Figure 1 shows the corresponding spectrograms for two subjects at the first trial. Note the difficulty in distinguishing frequency patterns for different cognitive processes, as well as the difference in the response between subjects.

Then, our classification problem is for each subject to identify the task he/she is performing given an EEG feature measurement. For this particular experiment we did not count with separate set of training data. Then, we used the *Jack-knife Method* (leave-one-trial out) for the classification process, while computing $\{\bar{\boldsymbol{y}}_i, C_i\}$, for $i = 1, \ldots, 5$, from the data at the remaining trials.

We repeated the classification experiment for raw data, LT, and ET, using the procedure described in Section 3 with a value of $q = 5$. The averaged results for all the subjects in terms of percentage of correct classification over 25 trials (5 trials per task) and 50 trials (10 trials per task) are shown in Figure 2. The graphics show the results for different values of p, and indicate in a vertical dotted-line the result at p_{opt}. We note that better performance is achieved when the data is transformed either by the LT or ET, while the optimal model obtained from the raw-data underperformed at the classification stage. These results were more evident when a larger set of training data was used in the classification procedure. Still, the LT appeared to have better performance in most cases.

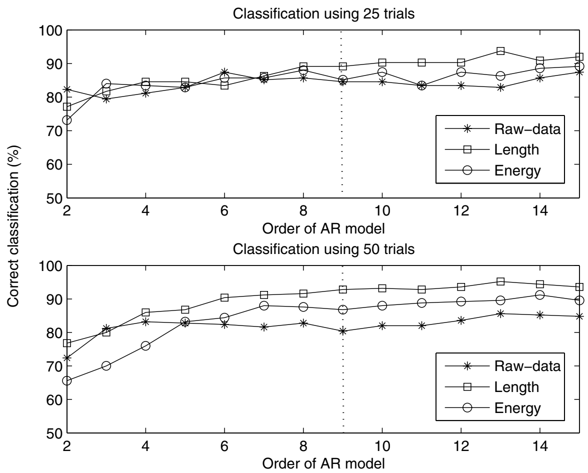

Fig. 2. Percentage of correct classification for all subjects using raw-data, LT, and ET, and for different AR models. The vertical dotted-line indicates the optimal model for the raw-data.

5. CONCLUSIONS

We presented a classification method for multichannel EEG data based on the LT and ET. The classification is made in a reduced dimension using an AR model. The advantage of our procedure is that the AR model is more likely to lead to an accurate classification when it is obtained through transformed data rather than raw-data and a sufficiently large set of training data is used. Furthermore, the SBC can be used to select an optimal AR model without loosing necessary features for signal classification. Future work in this area includes more extensive application to real EEG data of different cognitive activities, as well as evaluating other model selection criteria, like the Akaike's information criterion, or Sawa's Bayesian information criterion.

6. ACKNOWLEDGMENT

The authors would like to thank Dr. C. W. Anderson from Colorado State University for providing the data used in this paper. The work of D. Gutiérrez was supported by the National Autonomous University of Mexico (UNAM) Postdoctoral Fellowship Program.

7. REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, pp. 767–791, 2002.

[2] B. Blankertza, K. R. Muller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroder, and N. Birbaumer, "The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1044–1051, 2004.

[3] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, Inc., New York, 1996.

[4] F. Gritzali, G. Frangakis, and G. Papakonstantinou, "A comparison of the length and energy transformations for the QRS detection," in *Proceedings of the 9th Annu. Conf. IEEE EMBS*. 1987, pp. 549–550, Boston, MA.

[5] S. Waldert, D. Gutiérrez, A. Nehorai, D. McKenzie, C. L. Lowery, P. Murphy, H. Eswaran, and H. Preissl, "Real-time access of magnetoencephalographic/cardiographic data: technical realization and application to online fetal heart rate recording," in *Proceedings of the 27th Annu. Conf. IEEE EMBS*. 2005, Shangai, China.

[6] G. Gratton, "Dealing with artifacts: The EOG contamination of the event-related brain potential," *Behavior Research Methods, Instruments, and Computers*, vol. 30, pp. 44–53, 1998.

[7] B. Widrow, J.R. Glover, J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr., and R.C. Goodlin, "Adaptive noise cancelling: principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.

[8] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[9] F. Babiloni, L. Bianchi, F. Semeraro, J. R. Millán, J. Mouri no, A. Cattini, S. Salinari, M. G. Marciani, and F. Cincotti, "Mahalanobis distance-based classifierss are able to recognize EEG patterns by using few EEG electrodes," in *Proceedings of the 23th Annu. Conf. IEEE EMBS*. 2001, pp. 651–654, Istambul, Turkey.

[10] Z.A. Keirn and J.I. Aunon, "A new mode of communication between man and its surroundings," *IEEE Transactions on Biomedical Engineering*, vol. 37, pp. 1209–1214, 1990.

[11] H. Jasper, "The ten twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 371–375, 1958.