A COMPARATIVE STUDY OF TIME DELAY NEURAL NETWORKS AND HIDDEN MARKOV MODELS FOR ELECTROENCEPHALOGRAPHIC SIGNAL CLASSIFICATION

AMIN FAZEL
Dept. of CS and Electrical Eng.
School of Computing and Eng.
University of Missouri Kansas City
Kansas City, Missouri

REZA DERAKHSHANI
Dept. of CS and Electrical Eng.
School of Computing and Eng.
University of Missouri Kansas City
Kansas City, Missouri

ABSTRACT

In this paper, we analyze the performance of Time Delay Neural Networks (TDNN) and Hidden Markov Models (HMM) for Electroencephalogram (EEG) signal classification. The specific focus of this study is Brain-Computer Interfacing (BCI), where near-real time detection of mental commands during a multi-channel EEG recording is desired. We argue that HMM and TDNN should be preferred over the rigid, one-size-fits-all methods of the more traditional EEG signal classifiers. To analyze the utility of modern classification methods for BCI, we compare and discuss the performance of our suggested TDNN and HMM EEG classifiers with the reported best results on BCI 2003 EEG benchmark dataset Ia.

INTRODUCTION

The goal of Brain Computer Interfaces (BCI), also known as Thought Translation Devices (TTD), is to provide individuals with a new communication channel that conveys commands directly from the brain (Wolpaw 2002). Other areas that may benefit from this area of research include diagnosis of memory problems, cognitive development of the children (Taylor and Baldeweg 2002), diagnosis and treatment of attention-deficit disorder, and many other medical conditions that need tools for classification of brain's electrophysiological activities.

Classification of Electroencephalogram (EEG) is an important part of current BCI research. Using EEG-based BCI is arguably superior to other related modalities such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), since the brain signals of interest happen at a rate that is only within the temporal resolution of EEG. Moreover, compared to PET and fMRI, EEG is convenient, portable, and affordable.

In this paper we examine the application of two nonlinear intelligent signal analysis tools to non-invasively collected EEG signals in order to classify the spatiotemporal signatures of imagined commands; as while the researchers in neurobiology have mostly been utilizing linear discriminant for BCIs, one can argue that such approach is mathematically justifiable only if the dimension of the feature space is high enough.

Many attempts have been made to build an EEG-based BCI system (Wolpaw 2002). Generally speaking, the most important steps of these BCI systems are feature extraction and classification. For feature extraction, adaptive auto regressive models, Hjorth parameters, power spectrum, and principle component features have been used (Obermaier 2001, Keirn and Aunon 1990). Various classifiers have also been used, however Linear Discriminant Analysis (LDA), Artificial Neural Networks (ANN), Linear Dynamical Systems, and recently Hidden Markov Models (HMM) (Zhong and Ghosh 2002) have received especial attention.

METHODS

Slow Cortical Potential (SCP) signals are considered to be caused by shifts in the depolarization level of the upper cortical dendrites. Negative SCPs are the sum of synchronized ultra-slow excitatory postsynaptic potentials of apical dendrites. Positive SCPs are the result of a reduction in synchronized neuronal inflow at apical dendrites. They can also be generated by inhibitory activity or by outbound excitatory flow from cell bodies in cortex layers IV and V. To operate an SCP-based BCI systems, subjects learn to increase the negative SCP (Birbaumer 1990).

In this paper, first we demonstrate the utility of HMMs in modeling and classifying SCP signals. We do so by feeding the EEG signals directly to the HMM. This approach avoids the troublesome feature extraction step and thus the need for expert knowledge for selecting suitable EEG features. We then show the results of TDNN EEG classification using linear prediction code (LPC) features. We demonstrate that TDNN, by virtue of considering a history of input LPC features, can improve upon the reported LPC-MLP BCI classifiers (Garrett 2003). The proposed methods are tested on the dataset Ia of the BCI Competition 2003.

HMM-based BCI System: HMMs are stochastic models and have been used for model-based signal processing, where the model of a signal is first assumed and then its parameters are estimated (training). Classification of an unknown signal (i.e. testing) is based on the best fit among the available signal models. Usually we don't have the precise model of the stochastic process under study. However, in many cases the benefits of having an assumed model outweighs the inaccuracies of initial assumptions. The basic principles of HMM are briefly discussed below. A more detailed description can be found in (Rabiner 1989).

HMM is a stochastic finite state automata which models a system by a Markov process defined by parameters $\lambda = (\Pi, A, B)$, where Π is the initial state probability, $A = \{a_{i,j}\}$ is the state transition probability matrix, and $B = \{b_j(O_k))$ is the observation probability density function for each state. Given a training dataset, parameters of the HMM may be estimated by Baum-Welch algorithm (Rabiner and Juang 1993), which is an iterative Expectation-Maximization (EM) method to find maximum

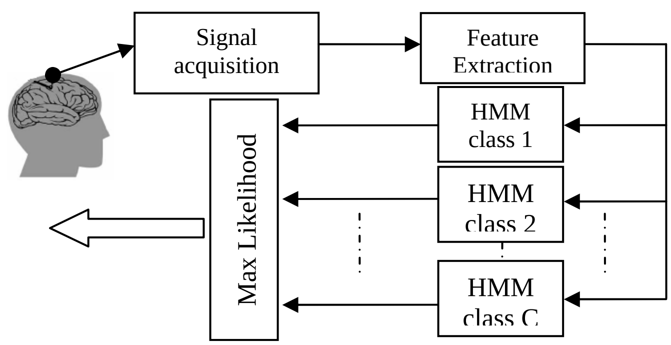

Figure 1: Block diagram of an HMM-based BCI system for C commands (here C=2)

likelihood. Given an HMM with parameters λ_i and a sequence of observations $O = O_1, ..., O_T$, we can find the probability of *O* being generated by an HMM model defined by λ_i, $p(O \mid \lambda_i)$. To classify an unknown signal *O* using HMMs characterized by λ_i, we find the most probable model of the unknown signal, that is

$$j = \arg \max_i \{ p(O \mid \lambda_i) \} \tag{1}$$

Since the state sequence is hidden, we need to find the cumulative probability of all possible observation sequences, hence

$$\begin{gathered} p(O \mid \lambda_i) = \sum_{\forall I} p(O \mid I, \lambda_i)\, p(I \mid \lambda_i) = \\ \sum_{i_1, i_2, ..., i_T} \pi_{i_1} b_{i_1}(O_1) a_{i_1, i_2} b_{i_2}(O_2) ... a_{i_{T-1}, i_T} b_{i_T}(O_T) \end{gathered} \tag{2}$$

However, the above has a computational complexity of $2TN^T$, thus the following efficient iterative process is used to find the probabilty of observation *O* given model λ:

$$\alpha_t(i) = p(O_1, O_2, ..., O_t, i_t = i \mid \lambda) \tag{3}$$

$\alpha_t(i)$ yields the probability of the partial observation vector for the time period [0 t], with t being the time of state i. The forward variable may be iteratively calculated as

$$\alpha_1(i) = \pi_i b_i(O_1); \qquad \mathrm{i} = 1, 2, ..., \mathrm{N} \tag{4}$$

$$\alpha_{t+1}(j) = \left[\sum \alpha_t(i) a_{i,j} \right] b_j(O_{t+1}); \qquad \mathrm{j} = 1, 2, ..., \mathrm{N} \tag{5}$$

$$p(O \mid \lambda) = \sum_{j=1}^{N} \alpha_T(j) \tag{6}$$

The above has computational complexity of the order TN^2. Similar to the forward variable, one can define an efficient backward variable which is used for training:

$$\beta_t(i) = p(O_{t+1}, O_{t+2}, ..., O_T \mid i_t = i, \lambda) \tag{7}$$

The above yields the probability of HMM output being *O* for time t+1 onward, with t being the time of state i. The output probability via backward variable is given by

$$p(O \mid \lambda) = \sum_{j=1}^{N} \beta_1(j) b_j(O_1) \tag{8}$$

According to the aforementioned procedures, and using the training data, we first created an HMM model for each class of EEG signals (i.e. training phase, please see Rabiner 1989 for details), with each class corresponding to a specific imagined cursor movement. During the EEG recognition (test phase), unknown SCP signal were passed to the trained HMMs. The probability of each model was calculated and the model with maximum output was chosen as the class of the unknown signal (cursor up/down motor command). We used the first 3.5 seconds of the feedback phase of channels 1 through 4 (out of 6) for this HMM classification study (see next section for the details of this dataset). Our HMM classifier uses continuous observation density HMM (CDHMM), a 5-state left-to-right model, one skip between the states, two Gaussian mixtures per state, and the conventional Viterbi beam search to calculate HMM class likelihoods. Our choice of EEG channels, as well HMM configuration, were based on trial and error.

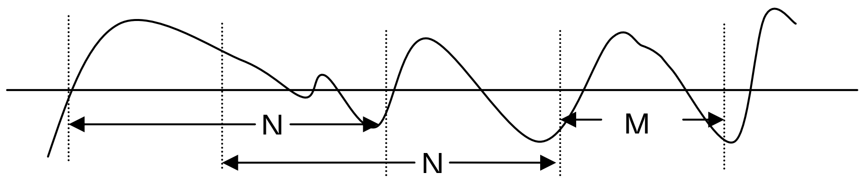

Figure 2: EEG signal frames

TDNN-based BCI System: Neural network techniques have been successfully applied to a wide range of pattern recognition tasks. Time Delay Neural Network (TDNN) (Waibel 1989) refers to a set of neural network architectures particularly suited to the classification of time series. This capability has been demonstrated for a variety of signal classification tasks such as phoneme recognition, and on-line gesture and handwriting recognition (Guyon 1991)(Vo 1993). In this study, we used focused TDNNs, where a history of the input signals is kept in input tapped-delay lines.

For TDNN-based BCI, we divided each of the 6 input EEG channels into 13 overlapping frames in time. Then we extracted k Linear Predictive Coefficient (LPC) features from each frame, and repeated with k = 6, 9, 12, and 15. We used LPC of sliding time widows since it is reported that multivariate autoregressive parameters are effective EEG features for mental imagery task detection (Anderson 1998). Having 6 input channels, we used a focused TDNN with 6k input nodes, one hidden layer, and one neuron per class in the output layer. We chose a time span of 0.5 seconds as the length of each input sliding window. Given the 256 samples/second rate of the input signals and a 50% adjacent frame overlap, we had N = 128 samples/window with an overlap of M = 64 samples, leading to 13 input frames per each trial. Thus the 268 training samples resulted in a 6×k×268 input training matrix. This process is shown in Figure 2.

We chose the number of training epochs for each trial to be 250, as early stopping did not provide us with any significant advantage during our experiments. We used zero, one, and two delay steps for our TDNNs, as shown in Figure 3.

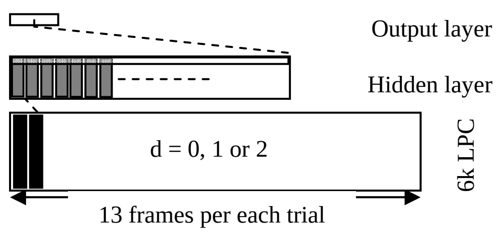

Figure 3: TDNN architecture

EXPERIMENTAL RESULTS

The dataset Ia was provided courtesy of Dr. Birbaumer and his team at University of Tuebingen, Germany (Blankertz 2004) for BCI 2003 competition. Six EEG channels were recorded from a healthy subject and sampled at 256Hz. The subject was asked to imagine moving a cursor up or down on a computer screen while his/her SCPs were being recorded. The subject received visual feedback of his/her SCPs (feedback phase). The dataset was divided into training (268 trials) and test set (293 trials), according to the BCI 2003 Ia set up. We used the EEG data of the feedback phase in each case. The direct current (DC) offset of SCP is typically an artifact of the recording process, thus we subtracted the mean of each signal to remove this DC component.

HMM Results: Table I shows the performance of our HMM-based classifier. It also shows the reported best results using LDA for classification of BCI 2003 SCP data (Mensh 2004). As it can be seen, our results are better on both training and test sets (85.82% vs. %70.9 and 85.3% vs. %82.6 on the training and test sets, respectively).

TABLE I
Results of our HMM vs. the best reported LDA

Classifier	Training set (% correct)	Test set (% correct)
HMM	85.82	85.3
LDA	70.9	82.6

Our results with HMM are better than the previously reported best results with LDA (BCI 2003 Ia SCP dataset). This shows the utility of HMMs in the challenging case of single-trial EEG classification.

TDNN Results: The following figure shows the results of TDNN with a one-step input delay on the training set.

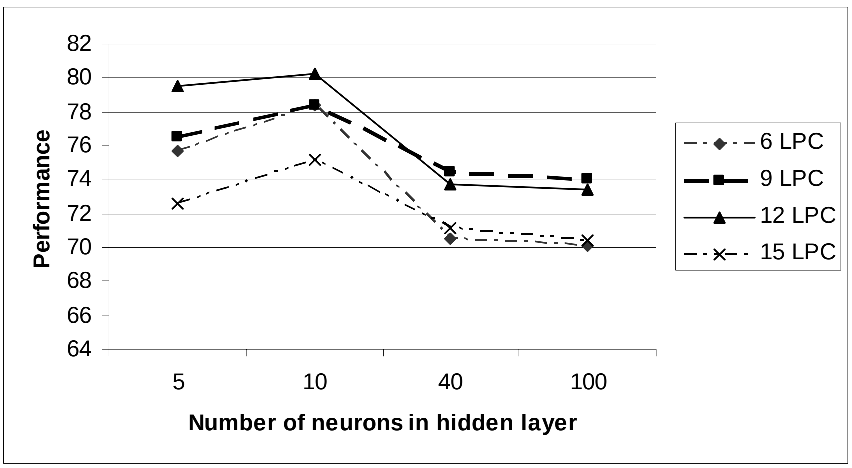

Figure 4: TDNN performance with a one-step input delay

We also tested the TDNN with two as well as no input delays (i.e. a simple feed forward Multi-Layer Perceptron or MLP) and with 6, 9, 12, and 15 LPC feature vectors, but non of them yielded better results. As shown in Figure 4, the best result with neural network is produced by TDNN with one-step input delay and 12 LPC feture vector. This shows that including the history of dynamically produced LPC features with TDNN improves upon similar systems with MLP classifiers.

CONCLUSIONS AND FUTURE WORK

We have demonstrated the utility of using HMM and TDNN for EEG signal classification by improving upon LDA and MLP-based BCI methods. Of especial importance to the field of BCI is the data-driven nature of HMM and TDNN-based classifiers. Individuals experiencing similar mental states often express these states differently in EEG recordings. Hence, such adaptable, data-driven classification systems are needed for detecting differently expressed targeted mental states of issuing commands to external devices. Another important issue in BCI is real-time or near real-time operation. After training our classifiers and using commercial personal computers, both the HMM and TDNN classifiers showed negligible time complexity during the test phase, so they can be used in online BCIs.

In this study we used time-domain information. However, using other features such as frequency-domain information and their fusion with our suggested systems may yield better performance. Thus for future work, we wish to incorporate a multi-modal approach through fusion of different methods, either at feature or classifier levels. We also plan to incorporate more EEG channels, especially in conjunction with the HMM classifier, as high-resolution EEGs can yield a better topographical detail of the cognitive processes and localized cognitive functions (McEvoy 2000).

REFERENCES

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M., 2002, “Brain-computer interfaces for communication and control,” Electroenceph. Clin. Neurophysiol., Vol. 113, pp. 767–791.

Taylor, M. J., and Baldeweg, T., 2002 “Application of EEG, ERP and intracranial recordings to the investigation of cognitive functions in children,” Developmental Science 5:3, pp. 318-334.

Obermaier, B., Guger, C., Neuper C., and Pfurtscheller G., 2001, “Hidden Markov Models for Online Classification of Single Trial EEG Data,” Pattern Recognition Letters, Vol. 22, pp. 1299-1309.

Keirn, Z. A., and Aunon, J. I., 1990, “A new mode of communication between man and his surroundings,” IEEE Trans. Biomed. Eng., Vol. 37, pp. 1209–1214.

Zhong, S., and Ghosh, J., 2002 “HMMs and coupled HMMs for multi-channel EEG classification,” In Proceeding IEEE Int. Joint Conference on Neural Networks, pp. 1154-1159.

Garrett, D., Peterson D. A., Anderson C. W., and Thaut M. H., 2003, “Comparison of Linear and Nonlinear Methods for EEG Signal Classification,” IEEE Trans. on Neural Systems and Rehabilitative Engineering, Vol. 11, pp. 141-144.

Birbaumer, N., Elbert, T., Canavan, A., and Rockstroh, B., 1990, “Slowpotentials of the cerebral cortex and behavior,” Physiol. Rev., Vol. 70, pp. 1-41.

Blankertz, B., Mueller, K. R., Curio, G., Vaughan, T. M., Schalk, G., Wolpaw, J. R., Schloegl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T., Schroeder, M., and Birbaumer, N., 2004, “The BCI Competition 2003,” IEEE Trans. on Biomedical Engineering, Vol. 51, pp. 1044–1051.

Rabiner, R., 1989, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proceedings of the IEEE, Vol. 77.

Rabiner, L. R., and Juang B-H., 1993, “Fundamentals of Speech Recognition”, Prentice Hall, N.J.

Waibel, A., Hanazawa, T., Hinton, G., Shiano, K., and Lang, K., 1989, “Phoneme Recognition using Time - Delay Neural Net-works,” IEEE Trans. on Acou., Speech, and Sig. Proc., Vol. 31.

Guyon, I., Albrecht, P., LeCun, Y., Denker, W., and Hub-bard, W., 1991 “Design of a Neural Network Character Recognizer for a Touch Terminal,” Pattern Recognition.

Vo, M.T. and Waibel, A., 1993, “A Multimodal Human Computer Interface: Combination of Gesture and Speech Recognition,” Adjunct Proc. InterCHI-93.

Anderson, C. W., Stolz, E. A., and Shamsunder, S., 1998, “Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks,” IEEE Trans. Biomed. Eng., Vol. 45, pp. 277-286.

Mensh, B. D., Werfel, J., and Seung, H. S., 2004, “BCI Competition 2003 - Data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals,” IEEE Trans. on Biomed. Eng., Vol. 51, pp. 1052-1056.

McEvoy, L.K, Smith, M. E., and Gevins, A., 2000 “Test-retest reliability of cognitive EEG,” Clinical Neurophysiology, Vol. 111, pp. 457-463.