HMM and IOHMM Modeling of EEG Rhythms for Asynchronous BCI Systems

Silvia Chiappa and Samy Bengio
{chiappa,bengio}@idiap.ch

IDIAP, P.O. Box 592, CH-1920 Martigny, Switzerland

Abstract. We compare the use of two Markovian models, HMMs and IOHMMs, to discriminate between three mental tasks for brain computer interface systems using an asynchronous protocol. We show that the discriminant properties of IOHMMs give superior classification performance but that, probably due to the lack of prior knowledge in the design of an appropriate topology, none of these models are able to use temporal information adequately.

1 Introduction

Over the last 20 years, several research groups have shown the possibility to create a new communication system, called Brain Computer Interface (BCI), which enables a person to operate computers or other devices by using only the electrical activity of the brain, recorded by electrodes placed over the scalp, without involving the muscular activity [7]. Cognitive processing (e.g. arithmetic operations, language, etc.) and imagination of limb movements are accompanied by changes in oscillations of the electro-encephalographic (EEG) signal, known as EEG rhythms [5], which can be captured by classification systems. Up to now, most proposed works in BCI research used static classifiers, while only a few works attempted to model the dynamics of these changes. For instance, in [6] the authors used Hidden Markov Models (HMMs) to discriminate between two motor-related mental tasks: imagination of hand or foot movement. However, these experiments were based on EEG signals recorded with a synchronous protocol, in which the subject had to undertake the imagined movement after receiving a cue given by the machine.

In this paper we explore the use of Markovian models, in particular, HMMs and an extension of them - the Input-Output HMMs - for distinguishing between three cognitive and motor-related mental tasks, for BCI systems based

This work is supported by the Swiss National Science Foundation through the National Centre of Competence in Research on "Interactive Multimodal Information Management". The authors would like to thank J. del R. Millán for fruitful discussions.

on an asynchronous protocol [4]. In this protocol, the subject does not follow any fixed scheme but concentrates repetitively on a mental action for a random amount of time and switches directly to the next task, without passing through a resting state. Thus the signal associated to each mental task represents a continuous sequence of mental events without marked beginning or end, from which the Markovian models should extract some discriminant information about the underlying dynamics.

The rest of the paper is organized as follows. In Sec. 2 and Sec. 3 HMM and IOHMM models are presented. Sec. 4 describes the data and the protocol used in the experiments. Experimental results are presented in Sec. 5 and discussed in Sec. 6. Final conclusions are drawn in Sec. 7.

2 Hidden Markov Models

A Hidden Markov Model (HMM) is a probabilistic model of two sets of random variables $Q_{1:T} = \{Q_1, \ldots, Q_T\}$ and $Y_{1:T} = \{Y_1, \ldots, Y_T\}$ [1]. The variables $Q_{1:T}$, called states, represent a stochastic process whose evolution over time cannot be observed directly, but only through the realizations of the variables $Y_{1:T}$, which are, in our case, the EEG signal recorded from several electrodes. For the computations to be tractable, it is necessary to assume the existence of conditional independence relations among the random variables, which can generally be expressed by the graphical model of Fig. 1(a). Furthermore, the observations $Y_{1:T}$ are considered to be identically distributed given the state sequence. Thus, to completely define an HMM model it suffices to give[1]: the initial state probability vector π of elements $\pi_i = P(Q_1 = i)$; the state transition matrix A[2], with $a_{ij} = P(Q_t = i|Q_{t-1} = j)$ and the set of emission probability density functions $B = \{b_i(y_t) = p(y_t|Q_t = i)\}$, which are, in our case, modeled by Gaussian mixtures (GMMs) [2].
For classification, a different model with associated parameters $\Theta_c = \{\pi, A, B\}$ for each class $c \in [1, \ldots, C]$ is trained so that the likelihood $\prod_{m \in M_c} p(y^m_{1:T}|\Theta_c)$ of the observed training sequences is locally maximized using the Baum-Welch method, which is a particular case of the EM algorithm [3]. During testing, an unknown sequence is assigned to the class whose model gives the highest joint density of observations:

$$c^* = \arg\max_c p(y_{1:T}|\Theta_c)\,.$$

3 Input-Output Hidden Markov Models

An Input-Output Hidden Markov Model (IOHMM) is an extended HMM in which the distributions of the output variables $Y_{1:T}$ and the states $Q_{1:T}$ are

[1]We indicate with $P(Q_t = i)$ the probability that the variable Q_t takes the value $i \in [1, \ldots, N]$, and, to simplify the notations, with $p(y_t)$ the probability density function associated to the random variable Y_t.

[2]It is assumed that the state transitions are independent of time t.

conditioned on a set of input variables $X_{1:T}$ [1]. For classification, the input variables represent the observed sequences and the output variables represent the classes. As shown in Fig. 1(b), independence properties analogue to the HMM case are assumed. Thus to parameterize an IOHMM we need the same set of distributions, which in this case are conditioned on the input variables. To model these distributions, we define a Multilayer Perceptron (MLP) [2] state network N_j and an MLP output network O_j for each state $j \in [1, \ldots, N]$. Each state network N_j has to predict the next state distribution:

$$P(Q_t = i | Q_{t-1} = j, x_t)\,,$$

while each output network O_j computes the current class distribution:

$$P(Y_t = c | Q_t = j, x_t)\,.$$

Training maximizes the likelihood $\prod_{m \in M} P(y^m_{1:T} | x^m_{1:T}, \Theta^3)$ using a generalized EM algorithm, which is an extension of the EM algorithm where, in the maximization step, the expected value of the likelihood cannot be analytically maximized and is instead increased by gradient ascent [3]. Note that, as opposed to the HMM framework where for each class a different model is trained on examples of that class only, here a unique IOHMM model is trained. During testing, we assign an unknown sequence[4] to the class c^* such that:

$$c^* = \arg\max_c P(Y_1 = c, \ldots, Y_T = c | x_{1:T}, \Theta)^5\,.$$

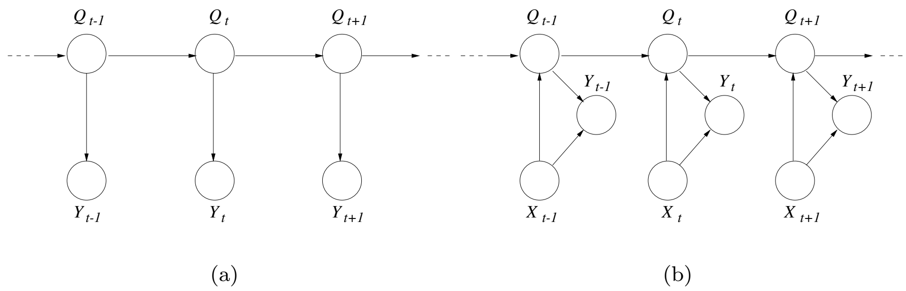

Figure 1: Graphical model specifying the conditional independence properties for an HMM (a) and an IOHMM (b). The nodes represent the random variables, while the arrows express direct dependencies between variables.

[3]Here Θ represents the set of MLP network parameters.

[4]In our case the whole test sequence belongs to one class.

[5]Another way to use the model is to assign class label only at the end of the sequence, modifying the likelihood maximization. Experiments carried out with this method gave worse performance, and thus are not reported here.

4 Data Acquisition

The EEG potentials were recorded with a portable system using 32 electrodes located at standard positions of the 10-20 International System, at a sample rate of 512 Hz. The raw potentials (without artifact rejection or correction) were spatially filtered using a surface Laplacian computed with a spherical spline [5]. Then the power spectral density over 250 milliseconds of data was computed with a temporal shift of 31.2 milliseconds, in the band 4-40 Hz and for the following 19 electrodes: F3, FC1, FC5, T7, C3, CP1, CP5, P3, Pz, P4, CP6, Cp2, C4, T8, FC6, FC2, F4, Fz and Cz.

Data was acquired from two healthy subjects without any experience with BCI systems during three consecutive days. Each day, the subjects performed 5 recording sessions lasting 4 minutes, with an interval of around 5 minutes in-between. During each recording session the subjects had to concentrate on three different mental tasks: imagination of repetitive self-paced left and right hand movements and mental generation of words starting with a given letter. The subjects had to change every 20 seconds between one mental task and another under the instruction of an operator[6]. In this study we have analyzed the performance on the last two days of recording, when the subjects had already acquired some confidence with the mental tasks.

5 Experiments

The HMM and IOHMM models have been trained on the EEG signal of the first three sessions of recordings of each day, while the following two sessions were used as validation and test sets. In the HMM model, the validation set was used to choose the number of states, in the range from 2 to 7, and the number of Gaussians (between 3 and 15). In the IOHMM, the validation set was used to choose the number of states (from 2 to 7), the number of iterations and the number of hidden units (between 25 and 200) for the MLP transition and emission networks. The MLP networks had one hidden layer. For reasons explained in the next section, we used a fully connected topology in which each hidden state could be reached by any other state. We split each recording session into signal segments of 1, 2 and 3 seconds, with a shift of half a second, obtaining a number of examples between 360 and 420.

Tables 1 and 2 show the performance of the two subjects over the second and third day of recording, using HMM and IOHMM models and their static counterparts, that is, GMM and MLP models respectively. GMMs and MLPs correspond to HMMs and IOHMMs with only one hidden state and thus can be used to test whether there is an advantage in using dynamical models over static ones. For each day, the columns give the error rate for different window lengths.

[6]During the real operation of the system the changing of mental task is performed as soon as the task has been recognized by the system.

Subject A	Second Day			Third Day		
	1 s	2 s	3 s	1 s	2 s	3 s
HMM	40.0%	36.4%	29.5%	24.3%	15.8%	09.0%
GMM	41.7%	34.3%	32.7%	22.4%	14.3%	12.1%
IOHMM	39.6%	32.8%	28.9%	19.6%	13.3%	09.3%
MLP	40.5%	29.4%	27.0%	19.3%	14.5%	09.8%

Table 1: Error rate of Subject A on the second and third day of recording, using HMMs and IOHMMs and their static counterparts: GMMs and MLPs.

Subject B	Second Day			Third Day		
	1 s	2 s	3 s	1 s	2 s	3 s
HMM	47.2%	46.2%	43.8%	49.1%	40.0%	36.3%
GMM	50.1%	45.9%	40.7%	45.7%	43.4%	34.9%
IOHMM	34.5%	29.4%	28.6%	36.7%	33.0%	27.5%
MLP	36.2%	29.7%	24.5%	40.0%	35.9%	31.4%

Table 2: Error rate of Subject B on the second and third day of recording.

6 Discussion

From the results presented in Tables 1 and 2 it can be seen that the classifiers perform significantly better than chance (66.7%), even with almost no user' s training, and that there is a great improvement when increasing the window length from 1 up to 3 seconds (which would still correspond to a reasonable speed for a BCI system, because of the short window shift).

We can also observe the superior performance of IOHMMs and MLPs compared to HMMs and GMMs. This can theoretically be explained by the fact that, when using HMMs, a separate model is trained for each class on examples of that class only. As a consequence, the training focuses on the characteristics of each class and not on the differences among them. On the contrary, in the IOHMM framework, a single model is trained using the examples from all the classes. This type of learning seems particularly appropriate when dealing with highly variable and noisy signal such as EEG, giving also more stable performance in different runs of the same experiments.

Another important result, shown in Tables 1 and 2, is the impossibility to choose between dynamical models and their static counterparts, which can be due to several reasons. The use of an asynchronous protocol in which the subject performs repetitive self-paced mental actions makes impossible to determine the beginning of each mental event. This fact, together with the lack of prior information about the dynamics of the rhythms hinders the selection of a state topology more appropriate than the fully connected one (which is known to have weak learning capabilities) and the modeling through an appropriate, and often crucial, state initialization. Furthermore, the high variability of the EEG signal recorded during different sessions, even if recorded very close in time, often makes the hyper-parameters chosen from an independent validation set not suitable for the test set.

7 Conclusions

This work pointed out two important aspects in the Markovian modeling of EEG, which are arousing growing interest in BCI research: first, the superiority of more discriminant models like IOHMMs over generative ones like HMMs; second, the lack of a practical advantage in using these models, as opposed to static ones, when no prior information can be used to build an appropriate structure for the hidden states.

A future research direction could be the application of other graphical models more accurately designed for the particular application, while maintaining a discriminant approach as in the IOHMM framework. An example could be a model that takes into account the high level of noise in the EEG, which includes artifacts and all the EEG activity which is not relevant for the discrimination of the mental task. In this case, the hidden structure should be designed to model the noisy component separately, in such a way that only the discriminant part is used for classification.

References

[1] Y. Bengio. Markovian models for sequential data. *Neural Computing Surveys*, 2:129–162, 1999.

[2] C Bishop. *Neural Networks for Pattern Recognition*. Oxford Univ., 1995.

[3] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley and Sons, 1997.

[4] J. del R. Millán. Brain-computer interfaces. In *The Handbook of Brain Theory and Neural Networks*, pages 178–181, 2002.

[5] P. L. Nunez. *Neocortical Dynamics and Human EEG Rhythms*. Oxford Univ., 1995.

[6] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller. Hidden Markov models for online classification of single trial EEG data. *Pattern Recognition Letters*, 22:1299–1309, 2001.

[7] T. M. Vaughan, W. J. Heetderks, L. J. Trejo, W. Z. Rymer, M. Weinrich, M. M. Moore, A. Kubler, B. H. Dobkin, N. Birbaumer, E. Donchin, E. W. Wolpaw, and J. R. Wolpaw. Brain-computer interface technology: A review of the second international meeting. *IEEE Transactions on Rehabilitation Engineering*, 11:94–109, 2003.