IDIAP RESEARCH REPORT

HMM and IOHMM Modeling of EEG Rhythms for Asynchronous BCI Systems

Silvia Chiappa, Samy Bengio

IDIAP–RR 03-49

December 2003

submitted for publication

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 − 27 − 721 77 11
fax +41 − 27 − 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

IDIAP Research Report 03-49

HMM and IOHMM Modeling of EEG Rhythms for Asynchronous BCI Systems

Silvia Chiappa, Samy Bengio

December 2003

submitted for publication

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 IOHMMs outperform HMMs but that, probably due to the lack of any prior information on the state dynamics, no practical advantage in the use of these models over their static counterparts is obtained.

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 [8, 13]. 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 [9], 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 [10] the authors used Hidden Markov Models (HMMs) to discriminate between two motor-related mental tasks: imagination of hand or foot movement. These experiments were based on EEG signals recorded with a *synchronous* protocol, in which the subject had to follow a fixed scheme before undertaking a movement under the instruction of the machinery.

This paper contributes to exploring 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 on an *asynchronous* protocol [8]. In this protocol, the subject does not follow any fixed scheme but concentrates repetitively on a mental task for a random amount of time and switches directly to the next, 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 document 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 $Y_{1:T} = \{Y_1, \ldots, Y_T\}$ [12]. 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 $Y_{1:T}$.

In order to make the related computations tractable, the following conditional independence relations over the random variables are assumed:

$$\{Q_t\} \perp \{Q_{1:t-2}, Y_{1:t-1}\} | Q_{t-1}\,, \tag{1}$$

and

$$\{Y_t\} \perp \{Q_{1:t-1}, Q_{t+1:T}, Y_{1:t-1}, Y_{t+1:T}\} | Q_t\,, \tag{2}$$

for each $t \in [1, \ldots, T]$, where the symbol $X \perp Y | Z$ indicates the conditional independence of X and Y, given Z. We can express these conditional independence relations by the help of a graphical model, as shown in Fig. 1.
The states $Q_{1:T}$ are discrete and can take a finite number of values $1, \ldots, N$, while the random variables $Y_{1:T}$ are continuous and represent the EEG signal recorded from several electrodes in the time interval $1, \ldots, T$. To simplify the notations, we will indicate with $P(q_t)$ the probability that the variable Q_t takes the value $q_t \in [1, \ldots, n]$ (the probability of being in the state q_t at time t), and with $p(y_t)$ the probability density function associated to the random variable Y_t.
From the independence relations in (1) it follows that:

$$P(q_t | q_{1:t-1}) = P(q_t | q_{t-1})\,,$$

that is, the states form a discrete time first order Markov chain. In addition, in order to reduce the number of parameters, $P(q_t | q_{t-1})$ is considered to be independent of time t, that is, the chain is *homogeneous.*

Another important property that can be derived from (2) is the following:

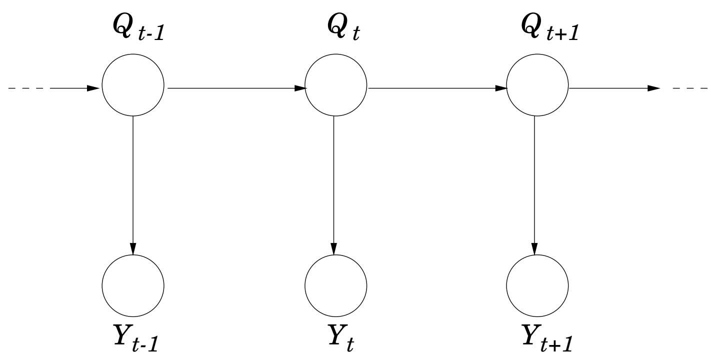

Figure 1: Graphical model specifying the conditional independence properties for a Hidden Markov Model. The nodes represent the random variables, while the arrows express direct dependencies between variables.

$$p(y_t|q_{1:t}, y_{1:t-1}) = p(y_t|q_t)\,.$$

From (2) it can also be seen that each observation Y_t is independent from every other observation given the current state, that is:

$$p(y_{t:t+h}|q_{t:t+h}) = \prod_{\tau=t}^{t+h} p(y_\tau|q_\tau)\,.$$

It is assumed that the observations are also identically distributed given the state sequence.
Given the above assumptions, to completely define an HMM model it suffices to give:

- the initial state probabilities $\pi_i = P(Q_1 = i), \quad i \in [1, \ldots, N]$,
- the state-transition probabilities matrix A, where $a_{ij} = P(Q_t = i|Q_{t-1} = j), \quad i, j \in [1, \ldots, N]$,
- the emission probability density functions $b_i(y_t) = p(y_t|Q_t = i), \quad i \in [1, \ldots, N]$.

Thus the complete parameter set can be denoted as $\Theta = (\pi, A, B)$, where B indicates the parameters corresponding to the emission density functions.
Given these parameters, we can compute the likelihood of an observed output sequence $y_{1:T}$, by considering all the possible state sequences $q_{1:T}$:

$$p(y_{1:T}) = \sum_{q_{1:T}} p(y_{1:T}|q_{1:T}) P(q_{1:T})\,,$$

where:

$$p(y_{1:T}|q_{1:T}) = \prod_{t=1}^{T} p(y_t|q_t) = b_{q_1}(y_1) \ldots b_{q_T}(y_T)\,,$$

and:

$$P(q_{1:T}) = \pi_{q_1} \prod_{t=2}^{T} P(q_t|q_{t-1}) = \pi_{q_1} a_{q_1 q_2} \ldots a_{q_{T-1} q_T}\,.$$

However, the number of calculations required in the above formulas is of the order of $2T \cdot n^T$, thus a recursive procedure, based on dynamic programming, which requires a number of computations of the order of $n^2 T$ is used in practice [1, 3].

For classification, a different model with associated parameters Θ_i for each class i is trained so that the likelihood:

$$\prod_{m \in M_c} p(y_{1:T}^m | \Theta_c)$$

is locally maximized over the set of training sequences using the Baum-Welch method, which is a particular case of the EM algorithm [5].
Once the HMMs have been trained, we assign an unknown test sequence $y_{1:T}$ to the class whose model gives the highest likelihood:

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

We assume that in each state $j \in [1, \ldots, N]$ our observations are generated by a Gaussian mixture model (GMM), that is:

$$b_j(\cdot) = \sum_{k=1}^{K} c_{jk} N(\cdot, \mu_{jk}, \Sigma_{jk}) \, ,$$

where $N(\cdot, \mu_{jk}, \Sigma_{jk})$ is a Gaussian distribution, with mean vector μ_{jk} and diagonal covariance matrix Σ_{jk}, and c_{jk} is the mixture coefficient for the k-th mixture of state j.

3 Input Output Hidden Markov Models

An Input-Output Hidden Markov Model (IOHMM) is an extension of an HMM to the case in which the distribution of the output variables $Y_{1:T}$ and the states $Q_{1;T}$ are conditioned on a set of input variables $X_{1:T}$ [2]. For classification, the input variables are associated to the observed sequences and the output variables to the classes.

As shown in Fig. 2, independence properties analogue to the HMM case are assumed, from which the following principal relations can be derived:

$$P(q_t | q_{1:t-1}, x_{1:t}) = P(q_t | q_{t-1}, x_t) \, ,$$

and

$$P(y_t | q_{1:t}, y_{1:t-1}, x_{1:t}) = P(y_t | q_t, x_t) \, .$$

Thus to parametrize an IOHMM we need:

- the initial state probabilities $\pi_i = P(Q_1 = i | x_1), \quad i \in [1, \ldots, N]$,
- the state-transition probabilities $P(Q_t = i | Q_{t-1} = j, x_t), \quad i, j \in [1, \ldots, N]$, $t \in [2, \ldots, T]$,
- the emission probabilities $P(Y_t = c | Q_t = j, x_t)$, $c \in [1, \ldots, C]$, $\quad j \in [1, \ldots, N]$, $t \in [2, \ldots, T]$.

To model the above conditional distributions, we define a Multilayer Perceptron (MLP) [4] 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:

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

based on the current input and on the previous state, while each output network O_j computes the distribution of the current output of the system:

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

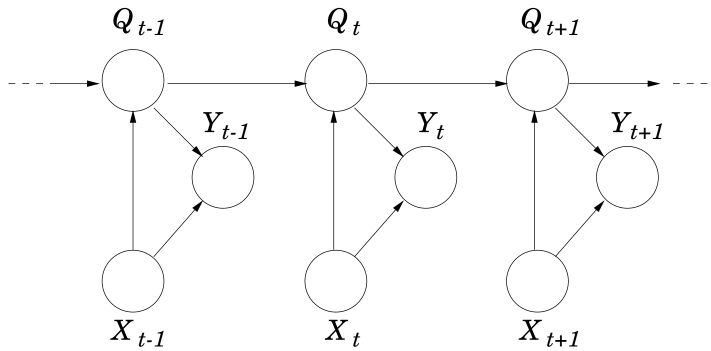

Figure 2: Graphical model specifying the conditional independence properties for an Input-Output Hidden Markov Model.

based on the current state and input. Thus each output of the state network N_j is associated to one of the successor i of the state j, while each output of the output network O_j is associated to one of the classes.

Training maximizes the likelihood over the M training sequences:

$$\prod_{m \in M} P(y_{1:T}^m | x_{1:T}^m, \Theta^1)\,,$$

using a generalized EM algorithm [7]. As in the standard EM algorithm, the expectation step estimates the expectation of the joint likelihood of the data and the state variables, given the data and the old model parameters, while in the maximization step the MLP networks are trained to find a new set of parameters which increases the expectation using gradient ascent techniques.
Note that, as opposed to the HMM framework where for each each class a different model is trained on examples of that class only, here a unique IOHMM model is trained, thus yielding more discriminant properties.
Once the IOHMM model has been trained, to assign an unknown test sequence to a class we compare the probabilities of observing a sequence of outputs of the same class[2]:

$$c^* = \arg\max_c P(Y_1 = c, \ldots, Y_T = c | x_{1:T}, \Theta)^3\,. \tag{3}$$

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 [9]. 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

[1] Here Θ represents the set of MLP network parameters.
[2] In our case the whole test sequence belongs to one class, as explained in the next section.
[3] 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.

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[4].

In this study we have analyzed the performance on the last two days of recording, when the subjects 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 the 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 segments of signal lasting 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 can thus serve as direct verification of the advantage obtained using dynamical models. For each day, the columns give the error rate for different window lengths.

Subject	Second Day			Third Day		
A	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	Second Day			Third Day		
B	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.

[4]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.

6 Discussion

From the results presented in Tables 1 and 2 it can be seen that the correct classification of the three mental tasks is significantly better than chance, even with almost no user' s training, and shows a great improvement when increasing the window length from 1 up to 3 seconds (which would correspond to a still 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 turned out to be more appropriate for a high variable signal such as the 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 very close in time, often makes the hyper-parameters chosen from an independent validation set not suitable for the test set. This problem can eliminate, in practice, the eventual benefit in using dynamical models, as we have noticed in our experiments, where we have compared both the results obtained selecting the hyper-parameters in the test set directly and using an independent validation set. The first case (not shown here) corresponds to the unrealistic situation in which the correct parameters would be known a priori and thus can be used only for analyzing the change in performance of dynamical models.

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 practical advantage in using sequential models when no prior information can be used to build an appropriate structure, like in the case of modeling EEG rhythms with asynchronous BCI systems.

A future direction could be to analyze the use of IOHMMs to model the switching between mental tasks, in which case it would be possible to initialize properly the model (this requires an appropriate protocol with fast switching between mental tasks). Furthermore, given the high presence of noise in the EEG, another direction could be the use of IOHMMs for modeling regimes corresponding to part of the signal which is not discriminant, after constructing an appropriate model for the noise.

8 Acknowledgments

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.

References

[1] L.E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3:1–8, 1972.

[2] Y. Bengio and P. Frasconi. Input-output HMMs for sequence processing. *IEEE Transactions on Neural Networks*, 7:1231–1249, 1996.

[3] J. Bilmes. A gentle tutorial on the em algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical Report ICSI-TR-97-021, University of Berkeley, 1997.

[4] C Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

[6] H. H. Jasper. Ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–373, 1958.

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

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

[9] P. L. Nunez. *Neocortical Dynamics and Human EEG Rhythms.* Oxford University Press, 1995.

[10] 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.

[11] F. Perrin, J. Pernier, O. Bertrand, and J. Echallier. Spherical spline for potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, 72:184–187, 1989.

[12] L. R. Rabiner and B. H. Juan. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 1986.

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