COUPLED HIDDEN MARKOV MODELS FOR BIOSIGNAL INTERACTION MODELLING

Iead Rezek¹, Peter Sykacek², Stephen J. Roberts¹

1. Robotics Research Group, Department of Engineering Science, University of Oxford, UK
2. Austrian Research Institute for Artificial Intelligence, Vienna, Austria.

INTRODUCTION

Analysis of physiological systems frequently involves studying the interactions between them. In fact, the interactions themselves can give us clues about the state of health of the patient. For instance, interactions between cardio-respiratory variables are known to be abnormal in autonomic neuropathy [1]. Conversely, if the type of interaction is known a departure from this state is, by definition, an anomaly. For instance, in multichannel EEG analysis we would expect coupling between the channels. If the states of the channels are divergent one may suspect some recording error (electrode failure). Traditionally, such interactions would be measured using correlation or, more generally, mutual information measures [2]. There are, however, major limitations associated with correlation, namely, the linearity and stationarity assumptions. Mutual information, on the other hand, can handle non-linearity. However, it is unable to infer anti-correlation and it also assumes stationarity (for the density estimation).

Some recent work has focused on Hidden Markov models (HMM) in order to avoid some of the stationarity assumptions (the chain is still assumed to be ergodic) and model the dynamics explicitly [3]. HMMs handle the dynamics of the time-series through the use of a discrete state-space. These dynamics are taken from the state transition probabilities from one time step to the next. HMMs, however, typically have a single observation model and observation sequence and thus, by default, cannot handle the composite modelling of multiple time-series. This can be overcome through the use of multivariate observation models, such as multivariate autoregressive models [4]. The latter implies that interactions are modelled in the observation-space and not in state-space. This is unsatisfactory in some biomedical applications where the signals have completely different forms and thus spectra/bandwidths, e.g. heart-rate and respiration. Another approach models the interactions in state-space by coupling the state-spaces of each variate [5]. This allows interaction analysis to proceed on arbitrary multi-channel sequences.

HMMs: MARKOV CONDITION AND COUPLING

Classical HMMs

A standard HMM with \( K \) states and length \( N \) is defined by the parameter set \( \theta = \{ A, \mathbf{X}, \pi \} \), where \( \pi \) is the initial state probability parameter set¹, \( A \) is the set of observation probabilities conditioned on the state, and

¹This measure is often taken to be the probability of the state at time \( t = 0 \). This is of little use as it is not the probability of the states themselves which, for a homogenous Markov Chain can be estimated from the eigenvectors of the transition probabilities.
the set state transition probabilities. Graphically, it is represented as an unfolded mixture model whose states at time $t$ are conditioned on those at time $t - 1$ (see Figure 1[a]), also known as the first order Markov property. Estimation is often performed using the maximum likelihood framework [6, 7], where the hidden states are iteratively re-estimated by their expected values and, using these, the model parameters are re-estimated.

HMMs maybe extended in term of the dimensions of $A$ and $X$ or by changing the distributions of $A$ and $X$. In this paper, however, we are interested in systems with different topologies, and in particular, with extensions to more than one state-space variable.

**Coupled HMMS**

One possible extension to multi-dimensional systems is the coupled hidden Markov model (CHMM), as shown in Figure 1[b]. Here the current state is dependent on the states of its own chain ($X$) and that of the neighbouring chain ($Y$) at the previous time-step. These models have been used in computer vision [5] and digital communication [8]. Several things are of note:

- The model, though still a DAG (directed acyclic graph) contains a loop which forms the essential difficulty in estimating parameters in CHMMs. One approach to solve this problem is clustering [9]: all nodes of a loop are merged to form a single higher dimensional node. Though this may be an approach for the simple model shown in Figure 1[b] (the state variables $X_t, Y_t, X_{t+1}, Y_{t+1}$ would be merged in this case) it is infeasible once the “lag-link” (link from $X_t \rightarrow Y_{t+1}$ or $Y_t \rightarrow X_{t+1}$) is moved forward in time. For instance, the DAG shown in figure 1[c] requires merging of the state variables $X_t, Y_t, X_{t+1}, Y_{t+1}, X_{t+2}, Y_{t+2}$ and thus causes an immediate explosion of node dimensionality. Another approach, that of cut-sets [9], is also impractical for dynamic Bayes nets, since the number of cut-set variables increases with the chain length $N$.

- There are 2 chains with 2 possibly different observation models. Unless the observation models $A, B$ are identical, the chains cannot be simply transformed to a single HMM by forming a Cartesian product of the 2 observation-spaces.

While keeping close to the original model, estimation can be performed using variational learning (as shown in [5] although not recognized as such). For aesthetic reasons we have derived the full forward-backward equations for the maximum-likelihood estimation.
method. When doing so, it becomes apparent that a full approach without “explicit”-node merging is only possible for “lag-one” models. To be precise, a smoother (belief propagation both temporally causal and anti-causal) cannot be derived without node clustering, while filtering (only temporally causal belief propagation), can be derived for higher lags also. Sampling methods, however, do not rely on belief propagation and thus such difficulties can be avoided. All results shown in this paper use Gibbs sampling implemented in the BUGS software language [10].

RESULTS

Parameter Estimation

The likelihood function of the CHMM with state space dimensionalities \( K_x \) and \( K_y \), is given as

\[
\mathcal{L}(\theta) = P(X_0) P(Y_0) \prod_{t=2}^{T} P(A_t|X_t) P(B_t|Y_t) P(X_t|X_{t-1}, Y_{t-1}) P(Y_t|X_{t-1}, Y_{t-1})
\]

where the parameter vector \( \theta \) contains the transition probabilities \( TP_x, TP_y \) of \( P(X_t|X_{t-1}, Y_{t-1}) \) and \( P(Y_t|X_{t-1}, Y_{t-1}) \), the prior probabilities \( P_x \) and \( P_y \) of \( P(X_0) \) and \( P(Y_0) \), and the parameters of the observation densities \( P(A_t|X_t) \) and \( P(B_t|Y_t) \). Here, the latter are taken to be multivariate Gaussians with mean vectors \( \mu_x, \mu_y \) and covariance matrices \( \Sigma_x, \Sigma_y \). To obtain the full conditionals required for Gibbs sampling, we choose conjugate priors for all the coefficients:

- \( \mu_x \sim N(\xi_x, \kappa_x) \),
- \( \mu_y \sim N(\xi_y, \kappa_y) \),
- \( \Sigma_x^{-1} \sim \text{Wishart}(R_x, \nu_x) \),
- \( \Sigma_y^{-1} \sim \text{Wishart}(R_y, \nu_y) \),
- \( TP_x \) and \( P_x \sim \text{Dirichlet}(\alpha_{x1}, \ldots, \alpha_{xK_x}) \),
- \( TP_y \) and \( P_y \sim \text{Dirichlet}(\alpha_{y1}, \ldots, \alpha_{yK_y}) \),

where \( \mu_x \sim N(\xi_x, \kappa_x) \) denotes that \( \mu_x \) is drawn from a normal distribution with mean \( \xi_x \) and covariance matrix \( \kappa_x \), which are hyperparameters of the distribution. The hyperpriors were such such that all priors were vague or flat [11].

Synthetic Data

Synthetic data was generated by drawing 200 samples from a CHMM. Figure 2 shows the sampled state sequence and the sampled observation sequences. To highlight the benefits of the interaction model, the figure also shows the correlation function for the two observation sequences - clearly, no structure can be seen. The CHMM was then trained on the synthetic data set in an unsupervised fashion. The chain is allowed a burn-in time of 10000 iterations, after which we use the autocorrelation and visually check for adequate mixing. Subsequently, we monitor the variables for approximately 2000 iterations. The chain converged in almost all runs. Figure 3 shows the difference between the resultant state sequence of the CHMM and the true state sequence. The estimated state-labels were wrong on average for 3 out of 200 state variables.

Cheyne Stokes Data

We applied the CHMM to features extracted from a section of Cheyne Stokes Data\(^3\), which consisted of one EEG recording and a simultaneous respiration recording sampled at 128Hz. The feature, fractional spectral radius (FSR) measure [12], was computed from consecutive non-overlapping windows of two seconds length. Thus, respiration FSR then formed one channel while EEG FSR the second channel. The data has clearly visible

\(^3\)A breathing disorder in which bursts of fast respiration is interspersed with breathing absence.
states which can be used to verify the results of the estimated model. Figure 4 shows the FSR traces superimposed with the joint Viterbi path sequence. The CHMM states were set to $K_x = 2$ and $K_y = 2$. Essentially, of the four total states the CHMM splits the sequence such that two correspond to respiration and EEG in phase and two correspond to the transitions states, i.e. respiration on, EEG off and vice versa.

### Sleep EEG Data

We applied a 5-th order Bayesian auto-regressive (AR) model to a full night EEG recording. The AR-coefficients were extracted from two channels, namely C3 and C4, using sliding non-overlapping windows of 5s length (sampling rate 200Hz). The coefficients then formed the observation sequence for a CHMM. For each of the two chains the number of hidden states was set to 3, corresponding to three sleep-stages [13]: Deep Sleep, REM and Wake. The training data were state labels which were marked in the hypnogram as sleep state S4, REM and Wake. Thus, where the EEG was marked as S4 the CHMM states of both chains were set 0, where the EEG was marked as REM the CHMM states were set to 1, and so on.

Figure 5 depicts the hypnogram together with the estimated state sequence. The states, provided during training, are marked with crosses. Despite a range of 9 possible states, the chain spends most of its time in 3 stages. Movement artefacts are classified by the chain primarily as “wake”, while sleep stages S4 and S3 are classified as “Deep Sleep”.

The investigation of the transition probabilities shows a strong coupling of the chains whenever they agree on the state $(P(S_{t+1} = 0/0|S_t = 0/0) = 0.9664, P(S_{t+1} = 1/1|S_t = 1/1) = 0.9230, P(S_{t+1} = 2/2|S_t = 2/2) = 0.7889)$. The degree of coupling however changes depending on the joint state. Thus, the coupling is least when the chains reside in state 2/2, and most in state 0/0. This corresponds with the intuition, in that coupling is expected to be minimal during periods of wakefulness and maximal during deep sleep. As the two chains correspond to signal information from different sides of the brain this indicates a greater interhemispheric coupling during deep sleep.
CONCLUSIONS & FUTURE WORK

The stability of the CHMM sampled estimates is very encouraging. Moreover, the smoothness of the state sequence is much higher than that obtained using single chain HMMs [14]. The main difficulty is that it is not obvious by looking at the observations alone, whether the model will give meaningful results in an unsupervised training scheme.

The results presented here are clearly only preliminary. The CHMMs themselves may be expanded to more than just two coupled chains and have continuous hidden state spaces or other observation models. One other possible direction is that of trimming unnecessary transition probabilities with the use of entropy priors [15]. More important, however, is the question of lag-estimation. Future work will focus on model-lag selection for which we envisage the use of reversible jump sampling methods [16].

References


Acknowledgements

The authors would like to thank Will Penny for helpful discussions. This work is funded by the European Community Commission, DG XII (project SIESTA, Biomed-2 PL962040) whose support we gratefully acknowledge.