UvA-DARE (Digital Academic Repository)

Bayesian variable order Markov models: Towards Bayesian predictive state representations

Dimitrakakis, C.

Publication date
2009

Document Version
Final published version

UNIVERSITEIT
VAN
AMSTERDAM

Bayesian variable order Markov models: Towards Bayesian predictive state representations

Christos Dimitrakakis
Intelligent Systems Laboratory Amsterdam,
University of Amsterdam
The Netherlands

We present a Bayesian variable order Markov model that shares many similarities with predictive state representations. The resulting models are compact and much easier to specify and learn than classical predictive state representations. Moreover, we show that they significantly outperform a more straightforward Bayesian hierarchical Markov chain model and approach the performance of an oracle hidden Markov model. The simplicity of the approach makes it attractive for applications where the actual hidden state of the system does not need to be explicitly tracked, such as sequential prediction and decision making, while its fully Bayesian nature allows us to take into account the model uncertainty in decision making.
Keywords: Bayesian inference, reinforcement learning, variable order Markov models, predictive state representations

IAS

intelligent autonomous systems

Contents

Intelligent Autonomous Systems
Informatics Institute, Faculty of Science
University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam
The Netherlands

Tel (fax): +31 20 525 7461 (7490)
http://www.science.uva.nl/research/ias/

Corresponding author:

C. Dimitrakakis
tel: +31 20 525 7517
dimitrak@science.uva.nl
http://www.science.uva.nl/~dimitrak/

1   Introduction

We consider the problem of predicting a discrete sequence of observations arising from a discrete partially observable Markov process μ. When the state space and transition distribution of the process are unknown, this is not completely straightforward. One possibility is to explicitly estimate the process, as was done in the approach followed by Beal et al. [2]. Alternatively, one can ignore the underlying state structure and approximate the sequence of observations by a Markov chain. However, it is not known what order of the Markov chain might be suitable. The naive approach of maintaining a set of models of different order is inefficient, as the highest order models will usually be particularly sparse. We present a simple Bayesian construction that takes into account this sparseness, by creating a conditional hierarchy of predictive distributions. The approach can be seen as a Bayesian analogue to predictive state representations [c.f. 8, 13, 6], from which it was inspired.

More precisely, at each time step t, we observe the outcomes x_t of an unknown process μ. We denote the complete history of observations[1] to time t by $x^t = x_1, \ldots, x_t$ and a partial history by $x_{t-k}^t \triangleq x_{t-k}, \ldots, x_t$. When there is no need to specify a time index we shall use $\mathbf{x}$ to identify elements of $\mathcal{X}^k$. Finally, we write $\mu(\cdot|\cdot) \equiv \mathbb{P}_\mu(\cdot|\cdot) \equiv \mathbb{P}(\cdot|\cdot, \mu)$ to denote conditional distributions (as well as densities, when there is no ambiguity) under the process μ. We examine algorithms $\Lambda : \mathcal{X}^* \to \mathcal{X}$, mapping from any sequence of observations x^t to an inferred probability $\phi_t(x_{t+1}|x^t)$ over the subsequent outcomes $x_{t+1} \in \mathcal{X}$, and that make randomised predictions $\hat{x}_{t+1}$ by sampling from ϕ_t. Our goal is to find Λ that minimises the average loss over T steps

$$L_T(\Lambda) \triangleq \frac{1}{T} \sum_{t=1}^{T} \ell_t, \qquad\qquad \ell_t \triangleq \mathbb{I}\{\hat{x}_t \neq x_t\}, \qquad\qquad (1)$$

where $\mathbb{I}\{\cdot\}$ is an indicator function, that equals 1 when its argument is true and zero otherwise, and where $x_{t+1} \sim \mu$, and $\hat{x}_{t+1} \sim \phi_t$. We make the following assumption through out:

Assumption 1. *The unknown process μ is stationary.*

One particular type of process that matches our problem well is a hidden Markov model:

Definition 1 (HMM). *A hidden Markov model μ is a random process over $(\mathcal{S} \times \mathcal{X})^*$, the product space over sequences of states $s_t \in \mathcal{S}$, observations $x_t \in \mathcal{X}$, for all $t > 0$, with the following properties. Firstly, that the state distribution is Markov:*

$$\mu(s_{t+1}|s^t) = \mu(s_{t+1}|s_t), \qquad\qquad (2)$$

where we take s^t to mean $(s_1, \ldots, s_t)$. Secondly, that the observations only depend on the current state:

$$\mu(x_t|x^{t-1}, s^t) = \mu(x_t|s_t). \qquad\qquad (3)$$

When μ is unknown, one possibility is to use a Bayesian approach to estimate the correct model. Let $\mathcal{M}$ be a class of HMMs with common $\mathcal{X}, \mathcal{S}$, but unknown state observation distribution and let the true model be $\mu^* \in \mathcal{M}$. We equip the measurable space $(\mathcal{M}, \mathfrak{M})$, where $\mathfrak{M}$ is a suitable Borel set over $\mathcal{M}$, with a series of probability measures Ξ_t corresponding to our subjective belief. Thus, for any $M \in \mathfrak{M}$, $\Xi_t(M)$ is our subjective belief at time t that $\mu^* \in M \subset \mathcal{M}$.

---

[1] We maintain a general exposition, in that we do not consider a special structure in the space of observations, i.e. that x_t is a tuple (a_t, o_t, r_t) of actions, observations and rewards. Nevertheless, the proposed approach is adaptable to partially observable Markov decision processes with some additional work.

Assuming that the density $\xi_t \triangleq d\Xi_t$ over $\mathcal{M}$ exists for all t, we can write the following update:

$$\xi_{t+1}(\mu) \triangleq \xi_t(\mu|x_{t+1}, x_t), \qquad\qquad \xi_t(\mu) = \sum_{s^t} \xi_t(\mu, s^t). \qquad (4a)$$

Inference in such a domain is not trivial, and it becomes harder when $\mathcal{S}$ is unknown. Nonparametric methods such as the infinite hidden Markov model [2] can be used in that case. However, if we are not interested in the state *per se*, but only in the observations, we may be able to predict x equally well in some other way.

Predictive state representations [8] and observable operator processes [5]), do not explicitly model the state. Rather, they create a model over next observations, conditioned on histories of observations and past and future actions $\mathbb{P}(x_{t+1}^{t+k}|x^t, a^{t+k})$. We shall employ a similar device, with a Bayesian approach that considers all possible contexts (in practice up to some maximum order) all the time. We do not explicitly discuss actions in this paper, however it is not difficult to adapt the approach to take them into account.[2]

Them main contribution of the paper is that the difficulty in learning such representations vanishes, as the process can be implemented as a simple hierarchical prior over conditional models. This enables us to perform full Bayesian inference for discrete observations.

The paper is organised as follows. We discuss related work in Sec 1.1. Section 2 discusses the models used in this paper to predict observations. We first examine a simple Bayesian Markov chain model of order k in Sec. 2.1, which we later extend with a prior over orders k in Sec. 2.2. Finally, Bayesian predictive state representations are introduced in Sec. 2.3. Experimental comparisons and results are presented in Sec. 3 and we conclude with Sec. 4.

1.1   Related work

The suggested approach is inspired by predictive state representations (PSR) [8], and the closely related observable operator processes (OOP) [5] and variable length Markov chains [4, 3]. Such representations use a set of *contexts* $\{M_i\}$ on observations (called *tests* in the reinforcement learning literature), over which a probability p_t is maintained at any given moment t. Jointly, the set of tests and the probability of each test given the history, then assign a probability $P_t(x_{t+1}|x^t) = \sum_i M_i(x_{t+1})p_t(M_i|x^t)$ to the next observation.[3] Many approaches for learning the set of *core tests*, which is the set of tests necessary to predict future outcomes, as well as the required probability model p_t have been proposed in the past [13, 12, 6].

To our knowledge, so far there have not been any Bayesian approaches for learning such representations. This is an important issue, as non-Bayesian approaches appear difficult to adapt to the online learning case. Using a fully Bayesian framework, there is no set of "core" tests, in contrast to the previous approaches. Rather, we have a different amount of certainty in the predictions of each different test. Furthermore, the usefulness of each context changes as more data is acquired, something which is not taken into account at all in previous related approaches.

A closely related approach are context tree weightings [15]. Thsese are related to variable order Markov models, and they employ a Dirichlet prior at each context. However, in those models the representation is not updated.

Conceptually, our model is very similar to CTW, especially in its use of recursive computation to simplify inference and prediction. The main difference is that in the CTW model, the weight of each model class $w(M)$ are defined in a non-Bayesian way. Thus, they only depend upon

---

[2]As our approach does not employ Monte Carlo sampling to perform estimation, we need no special assumptions on the sampling distribution.

[3]Again note that we have simplified matters somewhat by only considering the next observation and no actions. However, we feel that this difference is tangential to the topic of this paper.

the size of the model and not on the number of data the model has seen. The main advantage of the Bayesian approach is that the weights of larger contexts become higher once more data becomes available.

Other related work includes the infinite hidden Markov model[2] and the infinite Markov model [10]. However, these models employ sampling, while the presented approach is closed form. Another well-known closed-form approach is Polya trees [7, 9], which define a tree of partition on $[0, 1]$, but can be trivially extended to $\mathcal{X}^*$. The main difference between Polya trees and the method proposed herein is our approach takes into account the quality of the predictions at each context.

The main contribution of this paper is a construction that allows us to compactly represent and update belief over all possible contexts. This belief can then be considered as a *Bayesian predictive representation of state*. In the sequel, we shall develop the model and demonstrate its predictive ability.

2   Predictive Bayesian models

One may usually predict the observations well by using a Markov chain of sufficiently high order k. Bayesian inference for a Markov chain of known order with discrete observations is simple and is summarily described in Sec. 2.1. Unfortunately, not only is the required order not known, but the best-approximating order depends on the amount of available data. For this reason, we consider a simple hierarchical prior over model classes of different order in Sec. 2.2. The problem with this approach, (ignoring switching time considerations such as those analysed in [14]) is that, since k order models have a predictive distribution condition on k observations, different contexts (partial histories) will have been observed different amounts of times. Thus, for some contexts it might be better to switch to a lower order model. We may also need to switch to a lower order model if for a specific previous observation x_{t-k}, the next observation x_{t+1} no longer depends on $x_{t-k'}$ for $k' > k$.

The main insight of this paper is that this can be achieved very easily with a Bayesian formulation of predictive state representations. Inference in such models is much simpler than inference in hidden Markov models and computationally efficient. The method is fully described in Sec. 2.3. However, we begin by introducing Bayesian inference over the set of Markov chains of a specific order and extend this to a collection of sets of various orders, before introducing the full model.

2.1   Bayesian inference for a single Markov chain

We restrict ourselves to the set $M_k \subset \mathcal{M}$ of Markov chains (MC) of order k with observation set $\mathcal{X}$. Each Markov chain $\mu \in M_k$ corresponds to a probability distribution conditioned on the k previous states. More specifically, we use a Dirichlet distribution for each $\mathbf{x} \in \mathcal{X}^k$, with density

$$\xi_t(\tau_{\mathbf{x}} = u) = \frac{\Gamma(\psi^{\mathbf{x}}(t))}{\prod_{i \in \mathcal{X}} \Gamma(\psi_i^{\mathbf{x}}(t))} \prod_{i \in \mathcal{X}} u_i^{\psi_i^{\mathbf{x}}(t)}, \tag{5}$$

with $\tau_x \triangleq \mathbb{P}(x_{t+1} \mid x_t{=}x)$, $u \in \mathbb{R}^{|\mathcal{X}|}$, $\|u\|_1 = 1$, $u \geq 0$. We will denote by $\Psi(\xi_t)$ the matrix of state transition counts at time t, with $\Psi(\xi_0)$ being the matrix defining our prior Dirichlet distribution.

Thus, for any belief ξ, we have Dirichlet parameters $\{\psi_i^x(\xi) : i \in \mathcal{X}, \mathbf{x} \in \mathcal{X}^k\}$. These values are initialised to $\Psi(\xi_0)$ and are updated via simple counting:

$$\psi_i^{\mathbf{x}}(\xi_{t+1}) = \psi_i^{\mathbf{x}}(\xi_t) + \mathbb{I}\left\{x_{t+1} = i \wedge x_{t-k}^t = \mathbf{x}\right\}. \tag{6}$$

We now need to move from the distribution of a single context vector $\mathbf{x}$ to the set of transition distributions for the whole chain. In order to do this easily, we shall make the following simplifying assumption.

Assumption 2. *For any* $\mathbf{x}, \mathbf{x}' \in \mathcal{X}^k$ $p(\tau_{\mathbf{x}}, \tau_{\mathbf{x}'}) = p(\tau_{\mathbf{x}})p(\tau_{\mathbf{x}'})$.

Now we shall denote the matrix of transition *probabilities* for MC μ as $\mathcal{T}^\mu$ and let $\tau_{\mathbf{x},i}^\mu \triangleq \mu(x_{t+1} = i \mid x_{t-k}^t = \mathbf{x})$. Then

$$\xi_t(\mu) = \xi_t(\mathcal{T}^\mu) = \xi_t(\tau_{\mathbf{x}} = \tau_{\mathbf{x}}^\mu \forall s \in \mathcal{S}, a \in \mathcal{A}) \tag{7a}$$

$$= \prod_{s \in \mathcal{S}} \prod_{a \in \mathcal{A}} \xi_t(\tau_{\mathbf{x}} = \tau_{\mathbf{x}}^\mu), \qquad \text{(from Ass. 2)} \tag{7b}$$

$$= \prod_{s \in \mathcal{S}} \prod_{a \in \mathcal{A}} \frac{\Gamma(\psi^{\mathbf{x}}(\xi_t))}{\prod_{i \in \mathcal{S}} \Gamma(\psi_i^s(\xi_t))} \prod_{i \in \mathcal{S}} \left(\tau_{s,i}^\mu\right)^{\psi_i^s(\xi_t)}. \tag{7c}$$

Thus Ψ is a sufficient statistic for expressing the density over M_k. To fully specify the model, we need to set the prior Dirichlet parameters. For this model and throughout the paper, these are always initialised to 1.

We can now employ ξ_t, the posterior over the parameters of M_k at time t, to predict the next data point:

$$\xi_t(x_{t+1}|x^t, M_k) \triangleq \int_{M_k} \mu(x_{t+1}|x^t)\xi_t(\mu)\,\mathrm{d}\mu. \tag{8}$$

It is common and straightforward to add another prior over model order, allowing us to switch to more complex models when more data is available. This is described in the next section.

2.2   A hierarchical prior over Markov chain orders

Let a collection of sets of models $\mathcal{W} = \{M_k\}$, equipped with a prior distribution ϕ_0 over $M \in \mathcal{W}$. Each model set M_k contains all Markov chains of order k for a fixed observation set $\mathcal{X}$ and thus admits a conjugate prior such as the Dirichlet prior outlined in the previous section. The belief over model sets can be updated as follows:

$$\phi_{t+1}(M_k) \triangleq \phi_0(M_k|x^{t+1}) = \frac{\xi_t(x_{t+1}|x^t, M_k)\phi_t(M_k)}{\sum_{M \in \mathcal{W}} \xi_t(x_{t+1}|x^t, M)\phi_t(M)}. \tag{9}$$

The posteriors over the models in each set M_k are updated according to (7), so $\xi_t(x_{t+1}|x^t, M_k)$, given by (8), is the predictive distribution of the k-th order model, conditioned on the history, and resulting from the posterior ψ_t obtained after seeing t observations. The only remaining question is how to set the prior ϕ_0. In this paper, we simply use the Akaike information criterion [1] and set it to $\phi_0(M_k) \propto \exp(-|\mathcal{X}|^{k+1})$. This is not ideal, since explicit switch time distributions have better performance [14], but it is good enough for our purposes.

We can now use the posterior over M_k to form a distribution over next observations:

$$\phi_t(x_{t+1}|x^t) \triangleq \sum_{M_k \in \mathcal{W}} \xi_t(x_{t+1}|x^t, M_k). \tag{10}$$

The main problem with this setting is that it will take a long time for $\phi_t(M_{k+1})$ to become greater than $\phi_t(M_k)$ because the number of possible contexts for order $k+1$ is larger by a factor of $|\mathcal{X}|$. Furthermore, for t such that $\phi_t(M_{k+1}) > \phi_t(M_k)$, there will exist some histories x^t for which M_{k+1} will be making much poorer predictions than M_k because of the possibility that $\mathbb{P}_\mu(x_{t+1}|x_{t-k-1}^t) \approx \mathbb{P}_\mu(x_{t+1}|x_{t-k}^t)$. Thus, intuitively, we could do better by switching to larger order models for some contexts only. This can be achieved if we allow our belief over model order to depend on the history.

2.3 Bayesian predictive state representations

We can test the hypothesis that higher order models are only better for some context vectors, by using a *conditional prior* over model orders. In order to do this, we now consider model classes M_i that are only active for specific subsets of histories.

More specifically, let M_i denote a conjugate model class predicting the next observation x_{t+1}. Letting our belief over model parameters at time t be ξ_t as usual, we define the predictive distribution of M_i at time t as:

$$M_i^t(x_{t+1}) \triangleq \int_{M_i} \mu(x_{t+1})\xi_t(\mu)\,\mathrm{d}\mu. \tag{11}$$

Furthermore, let $C_k \triangleq \{M_i : i = 2^k, \dots, 2^{k+1} - 1\}$, be a collection of k-order Markov models. Different models in the collection predict x_{t+1} given different context history vectors x_{t-k}^t. More precisely, we associate a vector $\mathbf{x}_i \in \mathcal{X}^k$ with each model class $M_i \in C_k$, such that $\bigcup_{i=2^k}^{2^{k+1}-1} \mathbf{x}_i = \mathcal{X}^k$ and $\mathbf{x}_i \cap \mathbf{x}_j = \emptyset$ for all $i \neq j$. Let us write $x^t \succeq \mathbf{x}$, for $\mathbf{x} \in \mathcal{X}^k$ if $x_{t-k}^t = \mathbf{x}$, and denote the set of active models M_i for a given history x^t by $\mathscr{M}(x^t) = \{M_i : x^t \succeq \mathbf{x}_i\}$. Now, note that we can use the collection C_k to define a distribution over next observations for all $t \geq k$:

$$C_k(x_{t+1}|x^t) = \sum_{M_i \in C_k} \mathbb{I}\left\{M_i \in \mathscr{M}(x^t)\right\} M_i^t(x_{t+1}) \tag{12}$$

The set C_k is analogous to the "uniform" set M_k used by the hierarchical model. All that remains is to define an appropriate distribution over models in $\mathscr{M}(x^t)$. In order to do this efficiently, we take advantage of the following construction.

Let $C^k \triangleq \bigcup_{j=0}^k C_j$ be the set of all models of order at most k, and denote the event that the order of M is at most k by $B_k \triangleq \mathbb{I}\left\{M \in C^k \wedge (M \notin C_{k'} \ \forall k' > k)\right\}$. Then we can write a recursion relating the prediction given that the model is at most order k, with the prediction given that the model is at most order $k - 1$, for the particular context x^t:

$$\begin{aligned}
\mathbb{P}(x_{t+1}|x^t, B_k) = {} & \mathbb{P}(M \in C_k|x^t, B_k)\,\mathbb{P}(x_{t+1}|x^t, M \in C_k) \\
& + [1 - \mathbb{P}(M \in C_k|x^t, B_k)]\,\mathbb{P}(x_{t+1}|x^t, B_{k-1}).
\end{aligned} \tag{13}$$

The above recursion allows us to efficiently store our belief over models using different contexts. Let us now see how to update this belief and make predictions. For compactness, let $\phi_t(\cdot|\cdot) \triangleq \mathbb{P}(\cdot|\cdot, x^t, \phi_0)$ denote any conditional distribution under our belief at time t. In addition, with a slight abuse of notation, let M_k denote the event that $M \in C_k$. Then, we can write the following update for our belief:

$$\phi_{t+1}(M_k|B_k) = \frac{\phi_t(x_{t+1}|x^t, M_k)\phi_t(M_k|B_k)}{\sum_{i=1}^k \phi_t(x_{t+1}|x^t, M_i)\phi_t(M_i|B_k)}. \tag{14a}$$

$$= \frac{\phi_t(x_{t+1}|x^t, M_k)\phi_t(M_k|B_k)}{\phi_t(x_{t+1}|x^t, M_k)\phi_t(M_k|B_k) + \phi_t(x_{t+1}|x^t, B_{k-1})[1 - \phi_t(M_k|B_k)]} \tag{14b}$$

Note that for any ϕ:

$$\phi(M_{k-1}|B_k) = [1 - \phi(M_k|B_k)]\phi(M_{k-1}|B_{k-1}), \tag{15}$$

which also allows us to write the following expression for the predictive distribution:

$$\phi(x_{t+1}|x^t, B_k) = \phi(x_{t+1}|x^t, M_k)\phi(M_k|x^t, B_k) + \phi(x_{t+1}|x^t, B_{k-1})[1 - \phi(M_k|x^t, B_k)]. \tag{16}$$

Let us now put everything together.

2.3.1   Implementation

Each collection C_k contains models of order k. Let $M_k^t \in C_k$ such that $\phi(M_k^t, x^t) > 0$, i.e. $M_k^t \triangleq C_k \cap \mathscr{M}(x^t)$. By construction, there is only one such model in C_k. We then use $p_k^{t'} \triangleq \phi_{t'}(M_k^t | B_k)$ to denote the probability that the correct model is M_k^t, given that the correct model's order is at most k, under belief $\phi_{t'}$ at time t'. Then note that B_t is trivially true at time t and thus $\phi(M_t^t) = \phi(M_t^t | B_t)\phi(B_t) + \phi(M_t^t | \neg B_t)\phi(\neg B_t) = \phi(M_t^t | B_t)$. So $\phi(M_t^t) = p_t^t$ and $\phi(M_{t-1}) = p_{t-1}^t(1 - p_t^t)$ and

$$\phi(M_k^t) = p_k^t \prod_{j=k+1}^{t}(1 - p_j^t).$$

In order to make predictions we must calculate (16), thus we must calculate

$$\alpha_k \triangleq \phi(x_{t+1} | x^t, B_k)$$

for all k. Note that

$$\alpha_k^t = p_k^t M_k^t(x_{t+1}) + (1 - p_k^t)\alpha_{k-1}^t.$$

Finally, we can calculate the posterior for each conditional model via

$$p_k^{t+1} = \frac{p_k^t M_k^t(x_{t+1})}{\alpha_k^t}.$$

This quantity only needs to be calculated for the models in $\mathscr{M}(x^t)$. With an efficient sparse matrix implementation, it is possible to store the coefficients p_k^t with little overhead.

3   Experiments

In order to test the efficacy of the proposed approach, we compared the Bayesian predictive state representation (BVMM) model, described in Sec. 2.3, with the Bayesian hierarchical model over Markov chains (BHMC) described in Sec. 2.2.

Each experiment was performed by generating data from an underlying class of hidden Markov models M, with $|\mathcal{S}|$ states and $|\mathcal{X}|$ observations, as well a specified maximum order $k_{\max}$ of the BVMM and BHMC models. Each experiment consisted of 100 runs of length $T = 10^4$. At the start of the n-th run, we randomly created a hidden Markov model μ_n and generated x^T observations.

Each of the models under evaluation calculated a history-dependent probability distribution ϕ_t, for $t = 0, \ldots, T-1$, from which we generated a series of predictions $\hat{x}^T$, by sampling $\hat{x}_{t+1} \sim \phi_t$. We then calculated the instantaneous loss of each model, ℓ_t. In addition to the BVMM and BHMC models, we also evaluated an oracle and an HMM oracle.

The **HMM oracle** selects $\hat{x}_{t+1}$ with probability $\beta_t(x_{t+1}) = \sum_{s_{t+1}, s_t} \mu_n(x_{t+1}, s_{t+1} | s_t)\beta_t(s_t)$. This is done by maintaining a belief $\beta_t(s_t)$ over states[4] with the initial belief β_0 being uniform. Thus, the predictions of this model are the best we could do if we knew the correct model μ_n.

The **oracle** actually observes s_t and predicts x_{t+1} with probability $\mu_n(x_{t+1} | s_t) = \sum_{s_{t+1}} \mu_n(x_{t+1}, s_{t+1} | s_t)$. Its performance is that obtainable under perfect state estimation.

Figure 1 presents some experiments with $|\mathcal{S}| = 4$, $|\mathcal{X}| = 4$ and for $k_{\max} \in \{2, 4, 8\}$. The results on the left column show L_t, the average loss to time t, averaged over 100 runs. The rightmost columns show the cumulative *regret* of each algorithm Λ compared to the HMM oracle Λ'

$$R_T(\Lambda, \Lambda') = \sum_{t=1}^{T} \ell_t(\Lambda) - \ell_t(\Lambda'),$$

---

[4]Using the standard updates $\beta_t(s_t) = \beta_{t-1}(s_t | x_t) = \mu_n(x_t | s_t)\beta_{t-1}(s_t)/\beta_{t-1}(x_t)$ and $\beta_{t-1}(s_t) = \sum_{s_{t-1}} \mu_n(s_t | s_{t-1})\beta_t(s_{t-1})$.

(a) Average loss,$|\mathcal{S}| = 4$

(b) Total regret, $|\mathcal{S}| = 4$

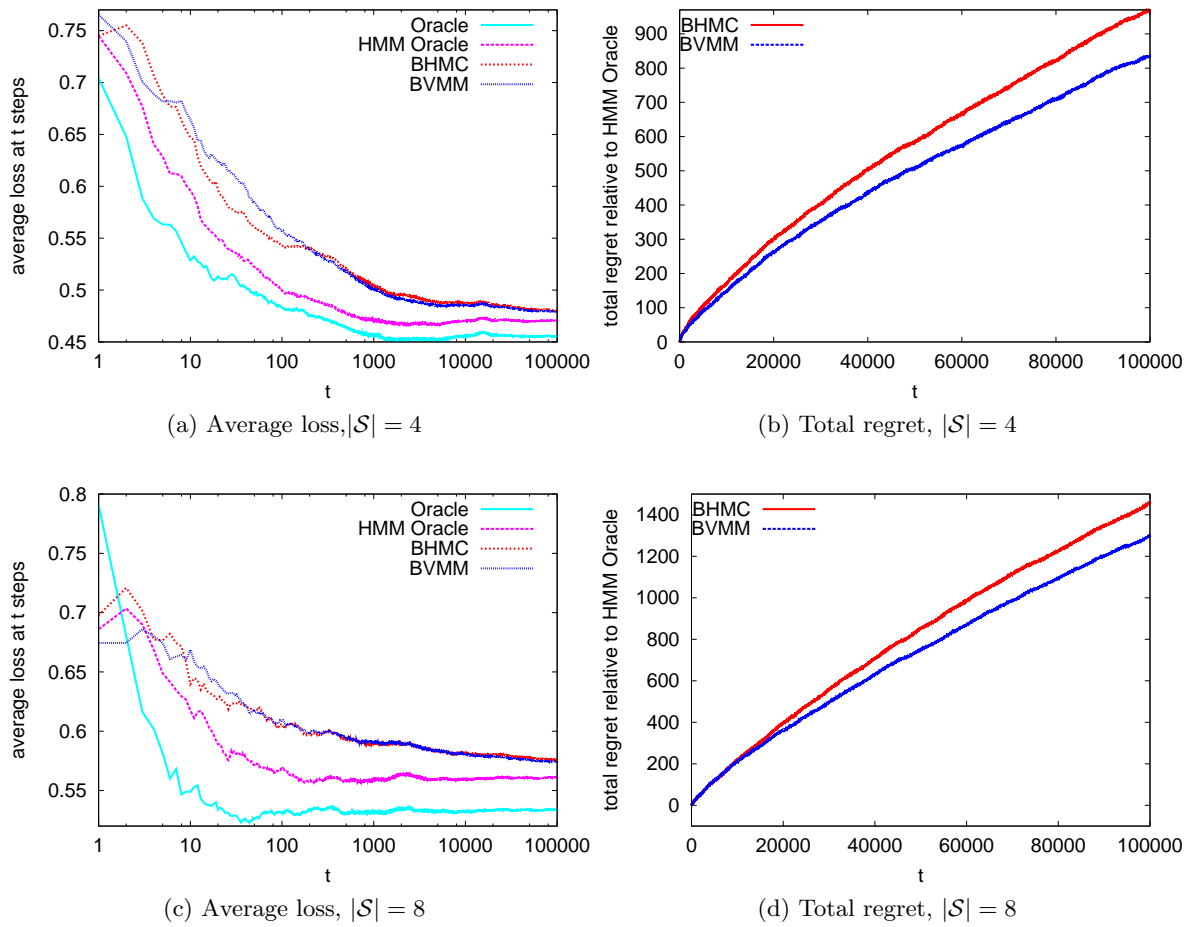(c) Average loss, $|\mathcal{S}| = 8$

(d) Total regret, $|\mathcal{S}| = 8$

Figure 1: The figures depict average loss at t time steps for all models with $k_{\max} = 8$, and cumulative regret with respect to the HMM oracle, for the two estimated models, with the underlying HMM having an observation set with $|\mathcal{X}| = 4$. The results are averaged over 100 runs.

where Λ is either the BVMM or the BHMC model. It can easily be seen that both models start at the same level of performance, but BHMC reaches a plateau very quickly. This fits the hypothesis that the conditional prior over models is more suitable for prediction. Overall, we see that the cumulative regret of BVMM is consistently smaller than that of BHMC. However, the overall gain, while signifcant, is not very large.

4 Conclusion

We presented a simple extension of the simpler Bayesian hierarchical Markov chain, by allowing our posterior over model orders to be conditioned on the history. This allows us to switch between higher and lower order models depending on the recent observations. The fully Bayesian approach allows us to treat the learning and prediction problem in a unified framework.

Experimentally, it appears as though the BVMM model consistently outperforms the naive hierarchical approach and suffers only a small amount of regret compared to the HMM oracle. We conjecture that a more classical PSR learning scheme, such as [13], can perform similarly to the BVMM approach for a fixed amount of data and with the right choice of core tests. However we think that the question of selecting the right core tests has not been satisfactorily addressed. Most methods extend the approach suggested in [8, 12], which relies on having a known POMDP model, to the case when the POMDP model is unknown. That requires performing tests of conditional independence, which in our view, not only lacks the elegance afforded by the fully Bayesian approach, but is also difficult to implement as it requires the definition of a threshold for accepting conditional independence.

The presented construction is similar to the one used in predictive state representations, though the two approaches are not directly equivalent. It is, however, easy to obtain a partial equivalence by replacing the space $\mathcal{X}$ with the product space of POMDP observations and actions $\mathcal{O} \times \mathcal{A}$. Then each outcome is actually $x_t = (o_t, a_t)$. Then, instead of maintaining a distribution $\mathbb{P}(x_{t+1}|x^t)$, we maintain $|\mathcal{A}|$ distributions, $\mathbb{P}(o_{t+1}|a_{t+1}, x^t)$, which fully characterise the system.

Compared to PSRs, the suggested approach makes use of the fact that the set of useful tests changes as we acquire more data. This is an extremely important aspect of the problem of learning to act in a large POMDP. Then, even if we knew the "right" core tests, it would be improper to use them from the start, since they are initially poorly estimated. Rather, estimating simpler tests initially and more complex tests as more data is acquired is a much more efficient use of the data.

In the future, we would like to address the following issues. Firstly, it would be important to perform further experiments on larger problems and with higher order models. Secondly, it is necessary to apply the model to actual POMDP problems, explicitly taking actions into consideration. Because the approach is fully Bayesian, it would be also theoretically possible to perform Bayes-optimal exploration [c.f. 11] in this framework. In fact, using a BVMM, inference is much simpler, since it is no longer required to perform elaborate sampling procedures. Finally, it would be extremely interesting to examine the performance gain of an explicit switching time prior [14] and to perform a theoretical analysis of the regret.

References

[1] H. Akaike. An information criterion (AIC). *Math Sci*, 14(5), 1976.

[2] Matthew J. Beal, Zoubin Ghahramani, and Carl Edward Rasmussen. The infinite hidden Markov model. In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani, editors, *NIPS*, pages 577–584. MIT Press, 2001.

[3] Ron Begleiter, Ran El-Yaniv, and Golan Yona. On prediction using variable order Markov models. *Journal of Artificial Intelligence Research*, pages 385–421, 2004.

[4] Peter Bühlmann and Abraham J. Wyner. Variable length Markov chains. *The Annals of Statistics*, 27(2):480–513, 1999.

[5] H. Jaeger. Observable operator processes and conditioned continuation representations. *Neural computation*, 12(6):1371–1398, 2000.

[6] M. R. James, T. Wessling, and N. Vlassis. Improving approximate value iteration using memories and predictive state representations. In *Proceedings of the National Conference on Artificial Intelligence*, 2006.

[7] Michael Lavine. Some aspects of polya tree distributions for statistical modelling. *The Annals of Statistics*, pages 1222–1235, 1992.

[8] M. L. Littman, R. S. Sutton, and S. Singh. Predictive representations of state. In *Advances in Neural Information Processing Systems 14*, 2001.

[9] R. Daniel Mauldin, William D. Sudderth, and S. C. Williams. Polya trees and random distributions. *The Annals of Statistics*, 20(3):1203–1221, 1992. ISSN 00905364. URL `http://www.jstor.org/stable/2242009`.

[10] D. Mochihashi and E. Sumita. The infinite Markov model. In *Advances in Neural Information Processing Systems*, pages 1017–1024. MIT Press, 2008.

[11] Stephane Ross, Brahim Chaib-draa, and Joelle Pineau. Bayes-adaptive POMDPs. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, Cambridge, MA, 2008. MIT Press.

[12] Matthew R. Rudary and Satinder Singh. A nonlinear predictive state representation. In *NIPS*, 2004.

[13] Satinder Singh, Michael L. Littman, Nicholas K. Jong, David Pardoe, and Peter Stone. Learning predictive state representations. In *Proceedings of the Twentieth International Conference on Machine Learning*, August 2003.

[14] T. van Erven, P. D. Grünwald, and S. de Rooij. Catching up faster by switching sooner : a prequential solution to the AIC-BIC dilemma. *arXiv*, 2008. A preliminary version appeared in NIPS 2007.

[15] F.M.J. Willems, Y.M. Shtarkov, and T.J. Tjalkens. The context tree weighting method: basic properties. *IEEE Transactions on Information Theory*, 1984.

Acknowledgements

IAS reports

This report is in the series of IAS technical reports. The series editor is Bas Terwijn (`bterwijn@science.uva.nl`). Within this series the following titles appeared:

[16] C. Dimitrakakis, A. Mitrokotsa *Statistical decision making for authentication and intrusion detection* Technical Report IAS-UVA-09-03, Informatics Institute, University of Amsterdam, The Netherlands, June 2009.

[17] P. Oude, G. Pavlin *Dependence discovery in modular Bayesian networks.* Technical Report IAS-UVA-09-02, Informatics Institute, University of Amsterdam, The Netherlands, 2009.

[18] C. Dimitrakakis *Complexity of Stochastic Branch and Bound for Belief Tree Search in Bayesian RL* Technical Report IAS-UVA-09-01, Informatics Institute, University of Amsterdam, The Netherlands, April 2009.

All IAS technical reports are available for download at the ISLA website, `http://www.science.uva.nl/research/isla/MetisReports.php`.