The Value of Observation for Monitoring Dynamic Systems

Eyal Even-Dar
Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104 USA
evendar@seas.upenn.edu

Sham M. Kakade
Toyota Technological Institute
Chicago, IL, USA
sham@tti-c.org

Yishay Mansour*
School of Computer Science
Tel Aviv University
Tel Aviv, 69978, Israel
mansour@cs.tau.ac.il

Abstract

We consider the fundamental problem of monitoring (i.e. tracking) the belief state in a dynamic system, when the model is only approximately correct and when the initial belief state might be unknown. In this general setting where the model is (perhaps only slightly) mis-specified, monitoring (and consequently planning) may be impossible as errors might accumulate over time. We provide a new characterization, the *value of observation*, which allows us to bound the error accumulation.

The value of observation is a parameter that governs how much information the observation provides. For instance, in Partially Observable MDPs when it is 1 the POMDP is an MDP while for an unobservable Markov Decision Process the parameter is 0. Thus, the new parameter characterizes a spectrum from MDPs to unobservable MDPs depending on the amount of information conveyed in the observations.

1 Introduction

Many real world applications require estimation of the unknown state given the past observations. The goal is to maintain (i.e. track) a *belief state*, a distribution over the states; in many applications this is the first step towards even more challenging tasks such as learning and planning. Often the dynamics of the system is not perfectly known but an approximate model is available. When the model and initial state are perfectly known then state monitoring reduces to Bayesian inference. However, if there is modelling error (e.g. the transition model is slightly incorrect), then the belief states in the approximate model may diverge from the true (Bayesian) belief state. The implications of such a divergence might be dire.

The most popular dynamic models for monitoring some unknown state are the Hidden Markov Model (HMM) [Rabiner & Juang, 1986] and extensions such as Partially Observable Markov Decision Process [Puterman, 1994], Kalman Filters [Kalman, 1960] and Dynamic Bayesian Networks [Dean & Kanazawa, 1989] — all of which share the Markov assumption. Naturally, one would like to provide conditions as to when monitoring is possible when modelling errors are present. Such conditions can be made on either the dynamics of the system (i.e. the transition between the states) or on the observability of the system. While the true model of the transition dynamics usually depends on the application itself, the observations often depend on the user, e.g. one might be able to obtain better observations by adding more sensors or just more accurate sensors. In this paper our main interest is in quantifying when the observations become useful and how it effects the monitoring problem.

Before we define our proposed measure, we give an illustrative example of an HMM, where the value of information can vary in a parametric manner. Consider an HMM in which at every state the observation reveals the true state with probability $1-\epsilon$ and with probability ϵ gives a random state. This can be thought of as having a noisy sensor. Intuitively, as the parameter ϵ varies from zero to one, the state monitoring become harder.

We introduce a parameter which characterizes how informative the observations are in helping to disambiguate what the underlying hidden state is. We coin this parameter the *value of observation*. Our value of observation criterion tries to quantify that different belief states should have different observation distributions. More formally, the L_1 distance between any two belief states and their related observation distributions is maintained up to a multiplicative factor (which is at least the *value of observation* parameter).

In this paper we use as an update rule a variant of the Bayesian update. First we perform a Bayesian update given our (inaccurate) model, and then we add some noise (in particular, we mix the resulting belief state with the uniform distribution). Adding noise is crucial to our algorithm as it ensures the beliefs do not become incorrectly overly confident, thus preventing the belief state from adapting fast enough to new informative information.

Our main results show that if the model is only approximate then our modified Bayesian updates guarantee that the true belief state and our belief state will not diverge — assuming the value of observation is not negligible. More specifically, we show that if the initial state is approximately accurate, then the expected KL-divergence between our belief state and the true belief state remains small. We also show

*Supported in part by a grant from ISF and BSF.

that if we have an uninformative initial state (e.g., arbitrary initial belief state) we will converge to a belief state whose expected KL-divergence from the true belief state is small and will remain as such from then on. Finally, we extend our results to the setting considered in [Boyen & Koller, 1998], where the goal is to compactly represent the belief state. The precision and rate of convergence depends on the value of observation.

One natural setting with an inaccurate model is when the underlying environment is not precisely Markovian. For example, it might be that the transition model is slightly influenced by some other extrinsic random variables. Given these extrinsic variables, the true transition model of the environment is only a slightly different model each time step. This is a case where we might like to model the environment as Markovian, even at the cost of introducing some error, due to the fact that transition model is not entirely Markovian. Our results apply also to this setting. This is an encouraging result, since in many cases the Markovian assumption is more of an abstraction of the environment, then the a precise description.

Related Work. The work most closely related to ours is that of Boyen and Koller (1998), where they considered monitoring in a Hidden Markov Model. In their setting, the environment is (*exactly*) known and the agent wants to keep a *compact* factored representation of the belief state (which may not exactly have a factored form). Their main assumption is that the environment is mixing rapidly, i.e the error contract by geometric factor after each time we apply the transition matrix operator. In contrast, we are interested in monitoring when we have only an *approximate* environment model. Both our work and theirs assume some form of contraction where beliefs tend to move closer to the truth under the Bayesian updates — ours is through an assumption about the value of observation while their is through assumption about the transition matrix. The main advantage of our method is that in many applications one can improve the quality of its observations, by adding more and better sensors. However, the mixing assumption used by Boyan and Koller may not be alterable. Furthermore, in the final Section, we explicitly consider their assumption in our setting and show how a belief state can be compactly maintained when both the model is approximate and when additional error accumulates from maintaining a compact factored representation.

Particle Filtering [Doucet, 1998] is a different monitoring approach, in which one estimates the current belief state by making a clever sampling, where in the limit one observes the true belief state. The major drawback with this method is in the case of a large variance where it requires many samples. A combination of the former two methods was considered by [Ng et al., 2002].

Building on the work of [Boyen & Koller, 1998] and the trajectory tree of [Kearns et al., 2002], McAllester and Singh (1999) provides an approximate planning algorithm. Similar extensions using our algorithm may be possible.

Outline. The outline of the paper is as follows. In Section 2 we provide notation and definitions. Section 3 is the main section of the paper and deals with monitoring and is composed from several subsections; Subsection 3.1 describes the algorithm; Subsection 3.2 provides the main monitoring theorem; Subsection 3.3 proves the Theorem. In Section 4, we show how to extend the results into the dynamic Bayesian networks.

2 Preliminaries

An Hidden Markov Model (HMM) is 4-tuple, (S, P, Ob, O), where S is the set of states such that $|S| = N$, P is the transition probability form every state to every state, Ob is the observations set and O is the observation distribution in every state. A *belief state* b is a distribution over the states S such that $b(i)$ is the probability of being at state s_i. The transition probabilities of the belief states are defined according to HMM transition and observation probability, using a Bayesian update.

For a belief state $b(\cdot)$, the probability of observing o is $O(o|b)$, where

$$O(o|b) = \sum_s O(o|s)b(s).$$

After observing an observation o in belief state $b(\cdot)$, the updated belief state is:

$$(U_o^O b)(s) = \frac{O(o|s)b(s)}{O(o|b)}$$

where U_o^O is defined to be the observation update operator. Also, we define the transition update operator T as,

$$(T^P b)(s) = \sum_{s'} P(s', s)b(s').$$

We denote by b_t the belief state at time t, where at time 0 it is b_0. (We will discuss both the case that the initial belief state is known and the case where it is unknown.) After observing observation $o_t \in Ob$, the inductive computation of the belief state for time $t+1$ is:

$$b_{t+1} = T^P U_{o_t}^O b_t,$$

where we first update the belief state by the observation update operator according to the observation o_t and then by the transition update operator. It is straightforward to consider a different update order. Therefore, b_{t+1} is the distribution over states conditioned on observing $\{o_0, o_1, \dots o_t\}$ and on the initial belief state being b_0.

3 Approximate Monitoring

We are interested in monitoring the belief state in the case where either our model is inaccurate or we do not have the correct initial belief state (or both). Let us assume that an algorithm has access to a transition matrix $\widehat{P}$ and an observation distribution $\widehat{O}$, which have error with respect to the true models. The algorithm's goal is to accurately estimate the belief state at time t, which we denote by $\hat{b}_t$.

For notational simplicity, we define $E_{o\sim b} = E_{o\sim O(\cdot|b)}$. When P and $\widehat{P}$ are clear from the context, we define T to be T^P and $\widehat{T}$ to be $T^{\widehat{P}}$. When O and $\widehat{O}$ are clear from the context, we define U_o to be U_o^O and $\widehat{U}_o$ to be $U_o^{\widehat{O}}$.

Our main interest is the behavior of

$$E\left[KL(b_t||\hat{b}_t)\right]$$

where the expectation is taken with respect to observation sequences $\{o_0, o_1, \dots o_{t-1}\}$ drawn according to the true model, and b_t and $\hat{b}_t$ are the belief states at time t, with respect to these observation sequences.

In order to quantify the accuracy of our state monitoring, we must assume some accuracy conditions on our approximate model. The KL-distance is the natural error measure. The assumptions that we make now on the accuracy of the model will later be reflected in the quality of the monitoring.

Assumption 3.1 *(Accuracy) For a given HMM model (S, P, Ob, O), an (ϵ_T, ϵ_O) accurate model is an HMM (S, P, Ob, O), such that for all states $s \in \mathbb{S}$,*

$$\begin{aligned} KL(P(\cdot|s)||\widehat{P}(\cdot|s)) &\leq \epsilon_T \\ KL(O(\cdot|s)||\widehat{O}(\cdot|s)) &\leq \epsilon_O \ . \end{aligned}$$

Next we define the value of observation parameter.

Definition 3.1 *Given an observation distribution O, let M^O be the matrix such that its (o, s) entry is $O(o|s)$. The* Value of Observation, *γ, is defined as $\inf_{x:\|x\|_1=1} \|Mx\|_1$ and it is in $[0, 1]$.*

Note that if the value of observation is γ, then for any two belief states b_1 and b_2,

$$\|b_1 - b_2\|_1 \geq \|O(\cdot|b_1) - O(\cdot|b_2)\|_1 \geq \gamma\|b_1 - b_2\|_1 \,.$$

where the first inequality follows from simple algebra.

The parameter γ plays a critical rule in our analysis. At the extreme, when $\gamma = 1$ we have $\|b_1 - b_2\|_1 = \|O(\cdot|b_1) - O(\cdot|b_2)\|_1$. Note that this definition is very similar to definition of the Dobrushin coefficient, $\sup_{b_1,b_2} \|P(b_1, \cdot) - P(b_2, \cdot)\|_1$ and it is widely used in the filtering literature [Moral, 2004]. We now consider some examples.

Let γ be 1 and consider b_1 having support on one state and b_2 on another state. In this case $\|b_1 - b_2\|_1 = 2$ and therefore $\|O(\cdot|b_1) - O(\cdot|b_2)\|_1 = 2$, which implies that we have a different observations from the two states. Since this holds for any two states, it implies that given an observation we can uniquely recover the state. To illustrate the value observation characterization, in POMDP terminology for $\gamma = 1$ we have a fully observable MDP as no observation can appear with positive probability in two states. At the other extreme, for an unobservable MDP, we can not have a value of $\gamma > 0$ since $\|O(\cdot|b_1) - O(\cdot|b_2)\|_1 = 0$ for any two belief states b_1 and b_2.

Recall the example in the Introduction where at every state the observation reveals the true state with probability $1-\epsilon$ and with probability ϵ gives a random state. Here, it is straightforward to show that γ is $1 - \epsilon$. Hence, as ϵ approaches 0, the value of observation approaches 1.

We now show that having a value of γ bounded away from zero is sufficient to ensure some guarantee on the monitoring quality, which improves as γ increases. Throughout, the paper we assume that $\gamma > 0$.

3.1 The Belief State Update

Now we present the belief state update. The naive approach is to just use the approximate transition matrix $\hat{P}$ and the approximate observation distribution $\hat{O}$. The problem with this approach is that the approximate belief state might place negligible probability on a possible state and thus a mistake may be irreversible.

Consider the following update operator $\tilde{T}$. For each states $s \in S$,

$$(\tilde{T})(s) = (1 - \epsilon_U)(\hat{T})(s) + \epsilon_U \texttt{Uni}(s),$$

where `Uni` is the uniform distribution. Intuitively, this update operator $\tilde{U}$ mixes with the uniform distribution, with weight ϵ_U, and thus always keeps the probability of being in any state bounded away from zero. Unfortunately, the mixture with the uniform distribution is an additional source of inaccuracy in the belief state, which our analysis would latter have to account for.

The belief state update is as follows.

$$\hat{b}_{t+1} = \tilde{T}\hat{U}_{o_t}\hat{b}_t, \qquad (1)$$

where $\hat{b}_t$ is our previous belief state.

3.2 Monitoring the belief state

In this subsection we present our main theorem, which relates the accuracy of the belief state to our main parameters: the quality of the approximate model, the value of observation, and the weight on the uniform distribution.

Theorem 3.2 *At time t let $\hat{b}_t$ be the belief state updated according to equation (1), b_t be the true belief state, $Z_t = E\left[KL(b_t||\hat{b}_t)\right]$, and γ be the value of observation. Then*

$$Z_{t+1} \leq Z_t + \epsilon - \alpha Z_t^2,$$

where $\epsilon = \epsilon_T + \epsilon_U \log N + 3\gamma\sqrt{\epsilon_O}$ and $\alpha = \frac{\gamma^2}{2\log^2 \frac{N}{\epsilon_U}}$.

Furthermore, if $\|\hat{b}_0 - b_0\|_1 \leq \sqrt{\frac{\epsilon}{\alpha}}$ then for all times t:

$$Z_t \leq \sqrt{\frac{\epsilon}{\alpha}} = \frac{\log \frac{N}{\epsilon_U}}{\gamma}\sqrt{2\epsilon}$$

Also, for any initial belief states b_0 and $\hat{b}_0$, and $\delta > 0$, there exists a time $\tau(\delta) \geq 1$, such that for any $t \geq \tau(\delta)$ we have

$$Z_t \leq \sqrt{\frac{\epsilon}{\alpha}} + \delta = \frac{\log \frac{N}{\epsilon_U}}{\gamma}\sqrt{2\epsilon} + \delta$$

The following corollary now completely specifies the algorithm by providing a choice for ϵ_U, the weight of the uniform distribution.

Corollary 3.3 *Assume that $\|\hat{b}_0 - b_0\|_1 \leq \sqrt{\frac{\epsilon}{\alpha}}$ and $\epsilon_U = \frac{\epsilon_T}{\log N}$. Then for all times t,*

$$Z_t \leq \frac{6\log \frac{N}{\epsilon_T}}{\gamma}\sqrt{\epsilon_T + \gamma\sqrt{\epsilon_O}}$$

Proof: With the choice of ϵ_U, we have:

$$\sqrt{2\epsilon} = \sqrt{4\epsilon_T + 6\gamma\sqrt{\epsilon_O}} \leq 3\sqrt{\epsilon_T + \gamma\sqrt{\epsilon_O}}.$$

And,

$$\log \frac{N}{\epsilon_U} = \log \frac{N \log N}{\epsilon_T} \leq 2\log \frac{N}{\epsilon_T},$$

which completes the proof. □

3.3 The Analysis

We start by presenting two propositions useful in proving the theorem. These are proved later. The first provides a bound on the error accumulation.

Proposition 3.4 *(Error Accumulation) For every belief states b_t and $\hat{b}_t$ and updates $b_{t+1} = TU_{o_t} b_t$ and $\hat{b}_{t+1} = \tilde{T}\widehat{U}_{o_t}\hat{b}_t$, we have:*

$$\begin{aligned} E_{o_t}\left[KL(b_{t+1}||\hat{b}_{t+1})\right] &\leq KL(b_t||\hat{b}_t) + \epsilon_U \log N + \epsilon_O \\ &\quad - KL(O(\cdot|b_t)||\widehat{O}(\cdot|\hat{b}_t)) \end{aligned}$$

The next proposition lower bounds the last term in Proposition 3.4. This term, which depends in the value of observation, enables us to ensure that the two belief states will not diverge.

Proposition 3.5 *(Value of Observation) Let γ be the value of observation, b_1 and b_2 be belief states such that $b_2(s) \geq \mu$ for all s. Then*

$$\begin{aligned} KL(O(\cdot|b_1)||\widehat{O}(\cdot|b_2)) &\geq \frac{1}{2}\left(\frac{\gamma KL(b_1||b_2)}{\log\frac{1}{\mu}}\right)^2 \\ &\quad -3\gamma\sqrt{\epsilon_O} + \epsilon_O \end{aligned}$$

Using these two propositions we can prove our main theorem, Theorem 3.2.

Proof of Theorem 3.2: Due to the fact that $\tilde{U}$ mixes with the uniform distribution, we can take $\mu = \epsilon_U/N$. Combining Propositions 3.4 and 3.5, and recalling the definition of ϵ, we obtain that:

$$\begin{aligned} &E_{o_t}\left[KL(b_{t+1}||\hat{b}_{t+1})|o_{t-1}, ..., o_0\right] \leq \\ &KL(b_t||\hat{b}_t) + \epsilon - \alpha\left(KL(b_t||\hat{b}_t)\right)^2 \end{aligned}$$

By taking expectation with respect to $\{o_0, o_1, \ldots o_{t-1}\}$, we have:

$$\begin{aligned} Z_{t+1} &\leq Z_t + \epsilon - \alpha E\left[\left(KL(b_t||\hat{b}_t)\right)^2\right] \\ &\leq Z_t + \epsilon - \alpha Z_t^2, \end{aligned}$$

where the last line follows since, by convexity,

$$E\left[\left(KL(b_t||\hat{b}_t)\right)^2\right] \geq \left(E\left[KL(b_t||\hat{b}_t)\right]\right)^2,$$

which proves the first claim in the theorem.

We proceed with the case where the initial belief state is good in the sense that $\|\hat{b}_0 - b_0\|_1 \leq \sqrt{\frac{\epsilon}{\alpha}}$. Then we have that Z_t is always less than $\sqrt{\frac{\epsilon}{\alpha}}$. The function $Z_t - \alpha Z_t^2 + \epsilon$ has derivative $1 - 2Z_t\alpha$, which is positive when $Z_t \leq \sqrt{\frac{\epsilon}{\alpha}}$. Since Z_t at $\sqrt{\frac{\epsilon}{\alpha}}$ is mapped to $\sqrt{\frac{\epsilon}{\alpha}}$, then every $Z_t \leq \sqrt{\frac{\epsilon}{\alpha}}$ is mapped to a smaller value. Hence the Z_t will always remain below $\sqrt{\frac{\epsilon}{\alpha}}$.

We conclude with the subtle case of unknown initial belief state, and define the following random variable

$$Z'_t = \begin{cases} Z_t, & Z_t > \sqrt{\frac{\epsilon}{\alpha}} \\ \sqrt{\frac{\epsilon}{\alpha}}, & \text{Otherwise} \end{cases}$$

Note that Z'_t is a positive super-martingale and therefore converges with probability 1 to a random variable Z'. The expectation of Z' cannot be larger than $\sqrt{\frac{\epsilon}{\alpha}}$, since whenever Z' is larger than $\sqrt{\frac{\epsilon}{\alpha}}$ its expectation in the next timestep is strictly less than its expected value. Since by definition $Z'_t \geq Z_t$ then, regardless the our initial knowledge on the belief state, the monitoring will be accurate and results in error less than $\sqrt{\frac{\epsilon}{\alpha}}$. □

Error Accumulation Analysis

In this subsection we present a series of lemmas which prove Proposition 3.4. The lemmas bound the difference between the updates in the approximate and true model.

We start by proving the Lemma 3.8 provided at the beginning of the Subsection

Lemma 3.6 *For every belief states b_1 and b_2,*

$$KL(Tb_1||\widehat{T}b_2) \leq KL(b_1||b_2) + \epsilon_T$$

Proof: Let us define the joint distributions $p_1(s', s) = P(s, s')b_1(s)$ and $p_2(s', s) = \hat{P}(s, s')b_2(s)$. Throughout the proof we specifically denote s to be the (random) 'first' state, and s' to be the (random) 'next' state. By the chain rule for relative entropy, we can write:

$$\begin{aligned} KL(p_1||p_2) &= KL(b_1||b_2) + E_{s\sim b}[KL(T(\cdot|s)||\widehat{T}(\cdot|s)] \\ &\leq KL(b_1||b_2) + \epsilon_T \end{aligned}$$

where the last line follows by Assumption 3.1.

Let $p_1(s|s')$ and $p_2(s|s')$ denote the distributions of the first state given the next state, under p_1 and p_2 respectively. Again, by the chain rule of conditional probabilities we have,

$$\begin{aligned} KL(p_1||p_2) &= KL(Tb_1||\widehat{T}_a b_2) \\ &\quad + E_{s'\sim Tb}[KL(p_1(s|s')||p_2(s|s')] \\ &\geq KL(Tb_1||\widehat{T}b_2), \end{aligned}$$

where the last line follows from the positivity of the relative entropy. Putting these two results together leads to the claim. □

The next lemma bounds the effect of mixing with uniform distribution.

Lemma 3.7 *For every belief states b_1 and b_2*

$$KL(Tb_1||\tilde{T}b_2) \leq (1-\epsilon_U)KL(Tb_1||\hat{T}b_2) + \epsilon_U \log N$$

Proof: By convexity,

$$\begin{aligned} KL(Tb_1||\tilde{T}b_2) &\leq (1-\epsilon_U)KL(Tb_1||\hat{T}b_2) \\ &\quad + \epsilon_U KL(Tb_1||\texttt{Uni}(\cdot)) \\ &\leq (1-\epsilon_U)KL(Tb_1||\hat{T}b_2) \\ &\quad + \epsilon_U \log N, \end{aligned}$$

where the last line uses the fact that the relative entropy between any distribution and the uniform one is bounded by $\log N$. □

Combining these two lemmas we obtain the following lemma on the transition model.

Lemma 3.8 *For every belief states b_1 and b_2,*

$$KL(Tb_1||\tilde{T}b_2) \le KL(b_1||b_2) + \epsilon_T + \epsilon_U \log N$$

After dealing with transition model, we are left to deal with the observation model. We provide an analog lemma with regards to the observation model.

Lemma 3.9 *For every belief states b_1 and b_2,*

$$E_{o\sim O(\cdot|b_1)}\left[KL(U_o b_1||\widehat{U}_o b_2)\right] \le KL(b_1||b_2) + \epsilon_O - KL(O(\cdot|b_1)||\widehat{O}(\cdot|b_2))$$

Proof: First let us fix an observation o. We have:

$$\begin{aligned} KL(U_o b_1||\widehat{U}_o b_2) &= \sum_s U_o b_1(s) \log \frac{U_o b_1(s)}{\widehat{U}_o b_2(s)} \\ &= \sum_s (U_o b_1)(s) \log \frac{O(o|s) b_1(s)/O(o|b_1)}{\widehat{O}(o|s) b_2(s)/\widehat{O}(o|b_2)} \\ &= \sum_s (U_o b_1)(s) \log \frac{b_1(s)}{b_2(s)} - \log \frac{O(o|b_1)}{\widehat{O}(o|b_2)} \\ &\quad + \sum_s (U_o b_1)(s) \log \frac{O(o|s)}{\widehat{O}(o|s)} \end{aligned}$$

where the last line uses the fact that $O(o|b_1)$ and $\widehat{O}(o|b_2)$ are constants (with respect to s).

Now let us take expectations. For the first term, we have:

$$\begin{aligned} &E_{o\sim b_1}\left[\sum_s (U_o b_1)(s) \log \frac{b_1(s)}{b_2(s)}\right] \\ &= \sum_o O(o|b_1)\left[\sum_s \frac{O(o|s)}{O(o|b_1)} b_1(s) \log \frac{b_1(s)}{b_2(s)}\right] \\ &= \sum_{s,o} O(o|s) b_1(s) \log \frac{b_1(s)}{b_2(s)} = KL(b_1||b_2) \end{aligned}$$

where the last step uses the fact that $\sum_o O(o|s) = 1$. Similarly, for the third term, it straightforward to show that:

$$\begin{aligned} &E_{o\sim b_1}\left[\sum_s (U_o b_1)(s) \log \frac{O(o|s)}{\widehat{O}(o|s)}\right] \\ &= \sum_o O(o|b_1) \sum_s \frac{O(o|s) b_1(s)}{O(o|b_1)} \log \frac{O(o|s)}{\widehat{O}(o|s)} \\ &= E_{s\sim b}\left[KL(O(\cdot|s)||\widehat{O}(\cdot|s))\right] \end{aligned}$$

For the second term,

$$E_{o\sim b_1}\left[-\log \frac{O(o|b_1)}{\widehat{O}(o|b_2)}\right] = KL(O(\cdot|b_1)||\widehat{O}(\cdot|b_2))$$

directly from the definition of the relative entropy. The lemma follows from Assumption 3.1. □

Now we are ready to prove Proposition 3.4.

Proof of Proposition 3.4: Using the definitions of updates and the previous lemmas:

$$\begin{aligned} &E_{o_t\sim b_t}\left[KL(b_{t+1}||\hat{b}_{t+1})\right] \\ &= E_{o_t\sim b_t}\left[KL(TU_{o_t} b_t||\tilde{T}\widehat{U}_{o_t}\hat{b}_t)\right] \\ &\le E_{o_t\sim b_t}\left[KL(U_{o_t} b_t||\widehat{U}_{o_t}\hat{b}_t)\right] + \epsilon_T + \epsilon_U \log N \\ &\le KL(b_t||\hat{b}_t) + \epsilon_O - KL(O(\cdot|b_t)||\widehat{O}(\cdot|\hat{b}_t)) \\ &\quad + \epsilon_T + \epsilon_U \log N \end{aligned}$$

where the first inequality is by Lemma 3.8 and the second is by Lemma 3.9 This completes the proof of Proposition 3.4. □

Value of Observation Proposition - The Analysis

The following technical lemma is useful in the proof and it relates the L_1 norm to the KL divergence.

Lemma 3.10 *Assume that $\hat{b}(s) > \mu$ for all s and that $\mu < \frac{1}{2}$. Then*

$$KL(b||\hat{b}) \le \|b - \hat{b}\|_1 \log \frac{1}{\mu}$$

Proof: Let $\mathbb{A}$ be the set of states where b is greater than $\hat{b}$, i.e. $\mathbb{A} = \{s|b(s) \ge \hat{b}(s)\}$. So

$$\begin{aligned} KL(b||\hat{b}) &= \sum_s b(s) \log \frac{b(s)}{\hat{b}(s)} \le \sum_{s\in\mathbb{A}} b(s) \log \frac{b(s)}{\hat{b}(s)} \\ &= \sum_{s\in\mathbb{A}} (b(s) - \hat{b}(s)) \log \frac{b(s)}{\hat{b}(s)} + \sum_{s\in\mathbb{A}} \hat{b}(s) \log \frac{b(s)}{\hat{b}(s)} \\ &\le \log \frac{1}{\mu} \sum_{s\in\mathbb{A}} (b(s) - \hat{b}(s)) \\ &\quad + \sum_{s\in\mathbb{A}} \hat{b}(s) \log \left(1 + \frac{b(s) - \hat{b}(s)}{\hat{b}(s)}\right) \\ &\le \log \frac{1}{\mu} \sum_{s\in\mathbb{A}} (b(s) - \hat{b}(s)) + \sum_{s\in\mathbb{A}} (b(s) - \hat{b}(s)) \\ &= \frac{1}{2}\left(\log \frac{1}{\mu} + 1\right) \|b - \hat{b}\|_1 \end{aligned}$$

where we have used the concavity of the log function and that $\sum_{s\in\mathbb{A}} (b(s) - \hat{b}(s)) = \frac{1}{2}\|b - \hat{b}\|_1$. The claim now follows using the fact that $\mu < \frac{1}{2}$. □

We are now ready to complete the proof. We will make use of Pinsker's inequality, which relates the KL divergence to the L_1 norm. It states that for any two distributions p and q

$$KL(p||q) \ge \frac{1}{2}(\|p - q\|_1)^2 . \tag{2}$$

Proof of Proposition 3.5: By Pinsker's inequality and Assumption 3.1, we have

$$\|O(\cdot|s) - O(\cdot|s)\|_1 \le \sqrt{2KL(O(\cdot|s)||\widehat{O}(\cdot|s))} \le \sqrt{2\epsilon_O}$$

for all states $s \in S$. Using the triangle inequality,

$$\|O(\cdot|b) - O(\cdot|\hat{b})\|_1 \le \|O(\cdot|b) - \widehat{O}(\cdot|\hat{b})\|_1 + \|\widehat{O}(\cdot|\hat{b}) - O(\cdot|\hat{b})\|_1 .$$

Therefore,

$$\begin{aligned}\|O(\cdot|b) - \widehat{O}(\cdot|\hat{b})\|_1 &\geq \|O(\cdot|b) - O(\cdot|\hat{b})\|_1 \\ &\quad -\|\widehat{O}(\cdot|\hat{b}) - O(\cdot|\hat{b})\|_1 \\ &\geq \gamma\|b - \hat{b}\|_1 - \|\widehat{O}(\cdot|\hat{b}) - O(\cdot|\hat{b})\|_1 \\ &\geq \gamma\|b - \hat{b}\|_1 - \sqrt{2\epsilon_O}\end{aligned}$$

where we used $\|\widehat{O}(\cdot|\hat{b}) - O(\cdot|\hat{b})\|_1 \leq \max_s \|O(\cdot|s) - O(\cdot|s)\|_1$ and Pinsker's inequality to bound $\|\widehat{O}(\cdot|\hat{b}) - O(\cdot|\hat{b})\|_1 \leq \sqrt{2KL(\widehat{O}(\cdot|\hat{b})||O(\cdot|\hat{b}))}$.

Combing with Pinsker's inequality, we have

$$\begin{aligned}KL(O(\cdot|b)||\widehat{O}(\cdot|\hat{b})) &\geq \frac{1}{2}(|O(\cdot|b) - \widehat{O}(\cdot|\hat{b})|_1)^2 \\ &\geq \frac{1}{2}(\gamma\|b - \hat{b}\|_1 - \sqrt{2\epsilon_O})^2 \\ &\geq \frac{1}{2}(\gamma\|b - \hat{b}\|_1)^2 - \gamma\|b - \hat{b}\|_1\sqrt{2\epsilon_O} \\ &\quad +\epsilon_O \\ &\geq \frac{1}{2}(\gamma\|b - \hat{b}\|_1)^2 - 2\gamma\sqrt{2\epsilon_O} + \epsilon_O \\ &\geq \frac{1}{2}\left(\frac{\gamma KL(b||\hat{b})}{\log\frac{1}{\mu}}\right)^2 - 3\gamma\sqrt{\epsilon_O} + \epsilon_O\end{aligned}$$

where the before last inequality follows from $\|b - \hat{b}\|_1 \leq 2$ and the last inequality from Lemma 3.10. □

4 Extension to DBNs

In many application one of the main limitations is the representation of the state which can grow exponentially in the number of state variables. Dynamic Bayesian Networks [Dean & Kanazawa, 1989] have compact representation of the environment. However, the compact representation of the network does not guarantee that one can represent compactly the belief state.

The factored representation can be thought as having an update of the following form

$$\tilde{b}_{t+1} = H\hat{T}\hat{U}_o\tilde{b}_t, \tag{3}$$

where H projects the belief state into the closest point in the factored representation space. Next we adopt the following definition from [Boyen & Koller, 1998] regarding the quality of the factored representation.

Definition 4.1 *An approximation $\tilde{b}$ of $\hat{b}$, incurs error ϵ if relative to true belief state b we have*

$$KL(b||\hat{b})) - KL(b||\tilde{b}) \leq \epsilon_p$$

Armed with these definitions one can prove an equivalent Theorem to Theorem 3.2 with respect to $KL(\tilde{b}_t||b_t)$. Due to lack of space and the similarity to the previous results we omit them.

5 Conclusions and Open Problems

In this paper we presented a new parameter in HMM, which governs how much information the observation convey. We showed how one can do fairly good monitoring in absence of an accurate model/unknown starting state/compcat representation as long as the HMM' observations are valuable. An open question that remains is whether the characterization can be made weaker and still an agent would be able to track, for instance if in most states the observations are valuable but there are few in which they are not, can we still monitor the state? Another very important research direction is that of planning in POMDPs. Our results show that one can monitor the belief state of the agent, however this is not enough for the more challenging problem of planning, where one should also decide which actions to take. It is not clear whether our characterization can yield approximate planning as well. The major problem with planning is that of taking the best action w.r.t to a distribution over states can lead to disastrous state and one should look into the long term implications of her actions due to the uncertainty. We leave these interesting problems to future work.

References

Boyen, X., & Koller, D. (1998). Tractable inference for complex stochastic processes. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence(UAI)* (pp. 33–42).

Dean, T., & Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence, 5(3)*, 142–150.

Doucet, A. (1998). *On sequential simulation-based methods for bayesian filtering* (Technical Report CUED/F-INFENG/TR. 310). Cambridge University Department of Engineering.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, *82*, 35–45.

Kearns, M., Mansour, Y., & Ng, A. (2002). A sparse sampling algorithm for near-optimal planning in large markov decision. *Machine Learning*, *49(2-3)*, 193–208.

McAllester, D., & Singh, S. (1999). Approximate planning for factored pomdps using belief state simplification. *In Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI)* (pp. 409–416).

Moral, P. D. (2004). *Feynman-kac formula. genealogical and interacting particle systems*. Springer.

Ng, B., Peshkin, L., & Pfeffer, A. (2002). Factored particles for scalable monitoring. *Proceedings of the Eighteenth Conf. on Uncertainty in Artificial Intelligence (UAI)*.

Puterman, M. (1994). *Markov decision processes*. Wiley-Interscience.

Rabiner, L. R., & Juang, B. H. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 4–15.