Information Theoretic Approaches for Predictive Models: Results and Analysis

Monica Dinculescu

Supervised by Doina Precup

Abstract

Learning the internal representation of partially observable environments has proven to be a difficult problem. State representations which rely on prior models, such as partially observable Markov decision processes (POMDPs) are computation expensive and sensitive to the accuracy of the underlying model dynamics. Recent work by Still and Bialek offers an information theoretic approach that compresses the available history into an internal representation and maximizes its predictive power. We propose an alternative algorithm that ensures a more accurate internal representation, and faster optimal policy convergence times. In addition, in order to validate the asymptotic nature of the theoretical algorithm, we present the first empirical results in the field, comparing the internal representations returned by both approaches, their accuracy, and their predictive powers.

1 Introduction

Increasing attention has been given to modelling partially observable systems, where the agent is not allowed to observe the state directly; instead, observations are received which allow the agent to infer information about the state.

Consider the problem of an agent who interacts with a dynamical system. In this paper, we focus on an active scenario, in which the agent has the capacity to act on the world, and influence the observations produced. Active scenarios can be compared to passive ones, where the agent simply observes the data produced by the world and constructs an internal representation of the system. Thus, in the active scenario, the challenge becomes 1) deriving the internal representation, given the available data as well as 2) deciding on an action strategy.

Having the ability to predict the outcome of an action is central to all tasks that a human or artificial agent may encounter. Thus, rather than trying to solve the learning problem, in which the agent tries to discover the system structure by updating model parameters from observation data, we are interested in constructing an internal representation that has maximal predictive power.

That is, the internal representation of the system should be able to make good predictions about future observations, based on available data.

There have been two predominant approaches in predicting a sequence of observations: partially observable Markov decision processes (POMDPs), and the recently proposed predictive state representations (PSRs)[Littman et al., 2002]. We discuss the advantages and disadvantages of each of the methods, and propose an information theoretic approach based on that developed by [Still and Bialek,2004]. Our approach is different from the above in that the internal representation considers the information that both the internal state, as well as the action performed by the agent have about the future.

Finally, to demonstrate the effectiveness of our approach we consider the example of the float-reset problem, created by [Littman et al., 2002], and compare the predictive powers of the two algorithms.

2 Background

Before discussing the specific details of each of the above approaches, we introduce the basic terminology used in the paper. We define a *history*, x_t^{past} as a finite length sequence of action-observation pairs, $x_t^{past} = [a_{t-k}o_{t-k}, ..., a_{t-1}o_{t-1}]$, where k is the length of the history and a_io_i the action-observation pair observed at time step i in that history. Similarly, a *test* is an ordered sequence of action-observation pairs $t = [a_oo_o, ..., a_{t-1}o_{t-1}]$, from the current time step, into the future. Finally, a *future* y_t^{fut} is an observation $o_t$that will be observed after taking action a_t at time step t.

POMDPs

A partially observable Markov decision process (POMDP) is a general framework for decision making under uncertainty. Formally, a POMDP is defined as a tuple$\mathcal{M} = (S, \mathcal{A}, \mathcal{O}, b_o, T, O)$ where the state set S is the set of states that the system can be in, $\mathcal{A}$ is the discrete set of actions, and $\mathcal{O}$ is the discrete set of observations. The set T consists of $(n \times n)$ transition matrices T^a, where T^a_{ij} is the probability of reaching state j , by taking action a in state i. The set O consists of diagonal $(n \times n)$ observation matrices $O^{a,o}$, where $O^{a,o}_{ii}$is the probability of an observation o , given that action a was selected in state i. Finally, the $(1 \times n)$ vector b_o is an initial probability distribution over the system states [Singh et al., 2004].

Intuitively, the transition function T determines the distribution over next states, given that a certain action was selected in a state, while the observation function O reflects the partially observable nature of the system (the agent cannot determine, with certainty, the true state of the system).

The state representation in a POMDP is a $(1 \times n)$ belief vector $b(h)$, where $b_i(h)$ is probability of the system being in the hidden state i , given that the history h has been observed. The belief state is updated by computing:

$$b(h,a,o) = \frac{b(h)T^aO^{a,o}}{b(h)T^aO^{a,o}e_n^T},$$

where e_n^T is the $1 \times n$ vector of all 1's.
Thus, predictions are given by

$$P(t|h) = P(a_1o_1...a_ko_k|h) = b(h)T^{a_1}O^{a_1,o_1}...T^{a_k}O^{a_k,o_k}e_n^T$$

The main disadvantage of the POMDP approach is that, as seen above, estimating the state assumes perfect knowledge of the underlying model. Algorithms that try to learn the dynamics require, at minimum, an assumption about the topology of the model. In addition, belief state maintenance has, in the worst case, complexity equal to the size of the state space, and exponential in the number of variables. Taken together with the fact that there exist dynamic systems with a finite state space that cannot be modeled by any finite POMDPs [Jaeger, 1998], this shows that the capabilities of the POMDP framework are rather limited.

PSRs

Predictive state representations are a recent approach that try to represent the state of a system as a set of predictions of observable outcomes of experiments that the agent can perform on the system.

A set of tests $\mathcal{Q} = \{q_1...q_m\}$ is called a linear *Predictive State Representation* (PSR) if, for all histories h , the probability of any test t occurring can be computed as a linear combination of the predictions for the tests in $\mathcal{Q}$. More generally, $\mathcal{Q}$ is a linear PSR if, for every test t, there exists a $1 \times q$ projection vector m_t , such that, for all histories h, $P(t|h) = P(h)m_t^T$. We call the tests $q_1...q_m$ *in* $\mathcal{Q}$, the *core tests* of the PSR [Littman et al., 2002].

In addition, [Singh et al., 2004] define a *System Dynamics Matrix* $\mathcal{D}$ as an ordering over all possible tests and histories, of all lengths, where each of $\mathcal{D}$'s entries is the probability of a test given a history, $P(t|h)$. The finitely many linearly independent columns of $\mathcal{D}$ are the core tests of the PSR. The matrix is generated by computing the weight vectors m_t from the model parameters, for each test t.

A reason for interest in PSRs is that the state representation is constructed only from actions and observations seen, thus resulting in a less restrictive model. However, empirical results show that learning the PSR weight vectors requires significant amounts of data, making learning algorithms both data and computationally expensive [Singh, Littman et. al].

Active Learning and Optimal Predictions

[Still and Bialek,2004] recently proposed a solution to the specifed problem. Using principles from information theory, they propose a state representation with the objective of extracting predictive information from a time series.

The state representation is constructed as a finite set of states s_t that should 1) compress the information contained by the past time series and 2) retain maximal predictive powers. Intuitively, the better the state representation, the closer its future predictions should be to those given by the available data alone.

We want to find a mapping from $x_t^{past} \rightarrow s_t$, such the s_t is a lossy compression of the past histories with maximal predictive power. Using information theory, this can be formalized as the optimization principle :

$$F = max_{p(s_t|x_t^{past})}[I(s_t|y_t^{fut}) - \lambda I(s_t, x_t^{past})]$$

where $I(X,Y)$is the mutual information of X relative to Y, given by$I(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$.

Taken together with a Markovian assumption, the solution to the above principle is the optimal $x_t^{past} \rightarrow s_t$ mapping. This is the Gibbs probability distribution:

$$p(s_t|x_t^{past}) \sim p(s_t) e^{-\frac{1}{\lambda} D_{KL}[p(y_t^{fut}|x_t^{past})||p(y_t^{fut}|s_t)]},$$

where in place of the energy, we find the Kullbach-Leibler distance:

$$D_{KL}[p(y_t^{fut}|x_t^{past})||p(y_t^{fut}|s_t)] = p(y_t^{fut}|x_t^{past}) \log \frac{p(y_t^{fut}|x_t^{past})}{p(y_t^{fut}|s_t)}$$

It has been proven that, in the limit as $\lambda \rightarrow 0$, the assignment becomes deterministic. Thus, the optimal internal state that a specific history is mapped to is the one that gives a future prediction closest to that given by the history alone (i.e. minimizes the D_{KL} distance between $p(y_t^{fut}|x_t^{past})$and $p(y_t^{fut}|s_t)$):

$$s_t^*(x_t^{past}) = argmin_{s_t} D_{KL}[p(y_t^{fut}|x_t^{past})||p(y_t^{fut}|s_t)]$$

The bane of the above representation is that the actions taken by the agent are independent of the internal state. Since future observations are the result of an action being taken, the intuition is that the internal state taken together with the action should in fact be predictive.

3 Proposed Solution

The fundamental idea behind our approach is that actions, taken together with the internal state are predictive, rather than just the state itself.

The motivation behind this approach is that the future produced is solely caused by the action choice; therefore, it is reasonable to expect that predictions that account for the action strategy should be more accurate than those that depend on the state alone.

Thus, the above optimization principle can be modified to incorporate the information that the state and action jointly hold about the future. Formally,

$$F = max_{p(s_t|x_t^{past})}[I(\{s_t, a_t\}|y_t^{fut}) - \lambda I(s_t, x_t^{past})]$$

The first term in the equation ensures that the internal representation maximizes the information held about the future, while the second represents the compression of histories into states.

The $x_t^{past} \rightarrow s_t$ assignment in this case becomes stochastic. Since the action choices are probabilistic and depend on the optimal action policy (given by $p(a_t|x_t^{past})$), the state-history mapping will be as well, and depend on the aforementioned policy. By solving the optimization principle, we obtain another Gibbs distribution:

$$p(s_t|x_t^{past}) \sim e^{-\frac{1}{\lambda}\sum_a p(a_t|x_t^{past}) \cdot D_{KL}[p(y_t^{fut}|x_t^{past}, a_t)||p(y_t^{fut}|s_t, a_t)]},$$

where

$$p(y_t^{fut}|a_t, s_t) \sim \frac{\sum_{x_t^{past}} p(y_t^{fut}|x_t^{past}, a_t)p(a_t|x_t^{past})p(s_t|x_t^{past})p(x_t^{past})}{\sum_{x_t^{past}} p(a_t|x_t^{past})p(s_t|x_t^{past})p(x_t^{past})}.$$

Thus, the optimal internal state that a specific history is mapped to is:

$$s_t^*(x_t^{past}) = argmin_{s_t} \sum_a p(a_t|x_t^{past}) \cdot D_{KL}[p(y_t^{fut}|x_t^{past}, a_t)||p(y_t^{fut}|s_t, a_t)$$

Algorithmic Implementation

Based on the theoretical analysis presented above, we have constructed an iterative algorithm that performs well in any partially observable setting. The algorithm is composed of three parts. First, we simulate the initial estimates of $p(y_t^{fut}|x_t^{past}, a_t)$ and $p(x_t^{past})$, by taking random actions and creating a finite length initial trajectory. Secondly, we iteratively solve the above two equations, until $p(s_t|x_t^{past})$ converges. This result is the optimal $x_t^{past} \rightarrow s_t$ assignment that maps the specific history being observed to an internal state. Finally, an action is taken according to an action policy, producing a new history and observation, which are used to update the $p(y_t^{fut}|x_t^{past}, a_t)$ and $p(x_t^{past})$ distributions.

The iterative algorithm converges to a local optimum after every iteration, according to the same arguments that were previously used by [Still and Bialek,2004].

4 Example: Interactively learning a float-reset system

Description

We describe a very simple partially observable system, created by [Littman et al., 2002]. The system is composed of 5 hidden states, 2 actions and 2 observations. The

float action moves to the state on the left/right with uniform probability, and always produces an observation of 0. The reset action moves to the state on the far right. If the system was already in that state, the observation produced is a 1; otherwise, a 0 is observed. Each observed history is of length 3 (i.e. extends 3 time steps into the past).

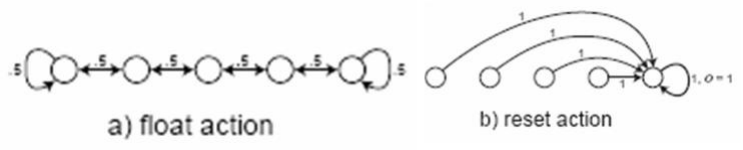

Figure 1. Float-reset problem

We have evaluated both the initial algorithm proposed by Still and Bialek, as well as our proposed algorithm, in 4 different scenarios:

- The agent has sufficient knowledge of the past (i.e. the initial distributions are sufficient to capture all the structure in the environment), and the internal representation is an uncompressed mapping of the available histories (the number of internal states s_t is equal to that of hidden states in the system, namely $s_t = 5$).
- The agent has sufficient knowledge of the past, and the internal representation is a compression of the available histories (the number of internal states s_t is smaller than that of hidden states in the system, namely $s_t = 3$).
- The agent has insufficient knowledge of the past (i.e. the initial distributions are too sparse and do not capture all the structure in the environment), and the internal representation is an uncompressed mapping of the available histories ($s_t = 5$).
- The agent has insufficient knowledge of the past and the internal representation is an compression of the available histories ($s_t = 3$).

Empirical results

Due to space constraints, we will only present the results for the fourth scenario, where the agent has insufficient knowledge of the past, and the internal representation is a compression of the available histories. The results for the other three scenarios are consistent with the presented one. We are concerned with determining which algorithm produces the internal representation with the greatest predictive power, and that best compresses the available data.

Optimization Function To begin, we compare the values of each of the optimization functions over time. This ensures that the trade off between the predictive power and the compression capacities of the internal representation is maintained.

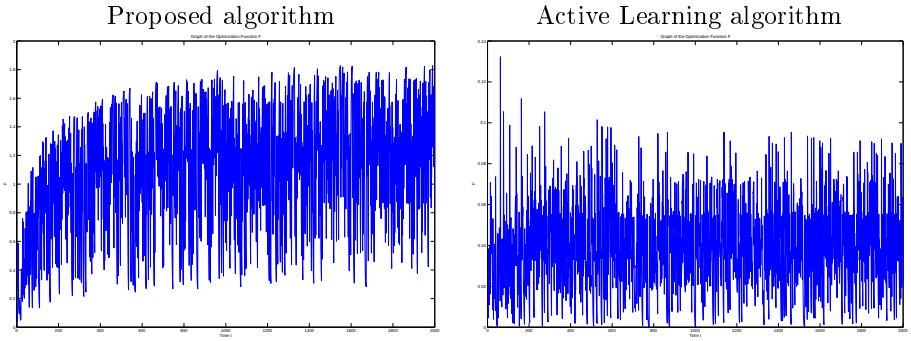

Figure 2: Optimization function F. The horizontal axis in each plot is the number of iterations of the algorithm.

Notice that the value of the optimization function in our approach increases with the number of iterations. This implies that the accuracy of the internal representation increases as well. Conversely, the internal representation as constructed by the original Active Learning algorithm does not vary much. In other words, its predictive powers improves very little past the initial ones. This is problematic, as a noisy initial distribution will result in a very inaccurate state-history mapping.

Lossy Compression The internal representation has to be a lossy compression of the past histories. In other words, the predictions given by a state should be very similar to those given by the histories that become assigned to it.

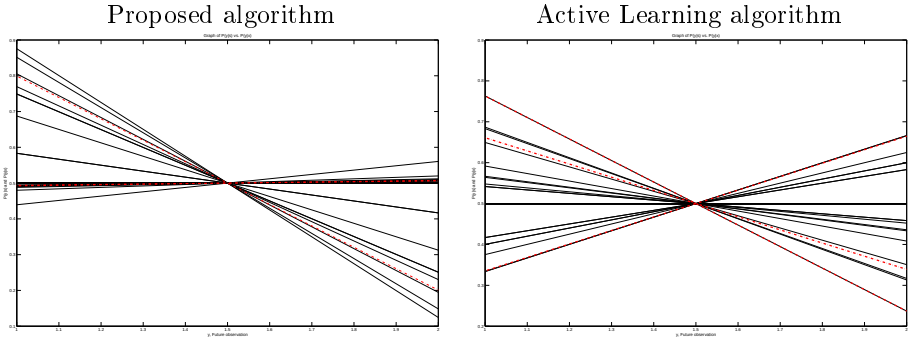

Figure 3: History-state mapping. The horizontal axis in each plot contains the possible future observations, namely y_t^{fut}. The vertical axes are $p(y_t^{fut}|x_t^{past})$, in black, and $p(y_t^{fut}|s_t)$, in red.

Notice that the number of states used in the internal representation differs between algorithms. This is because, while the original Active Learning algorithm is deterministic, and thus employs all 3 available internal states, our

proposed algorithm is probabilistic. Since there are two possible observations, there are two different possible predictions, and thus only 2 internal states are needed to represent the system.

If the internal representation is to be a compression of the available histories, then the future predictions as given by it (the red lines) should best approximate the future predictions given by the entire data (the black lines). The future information contained in the available history is better approximated by the 2 internal states as constructed by the proposed algorithm, than the 3 states used in the Active Learning algorithm, as can be seen from the above graph.

Predictive Powers From the problem definition, we know with certainty that a float action will produce an observation of 0. Consequently, a good internal representation should make the same prediction regardless of previous observations. On the other hand, the observation produced by a reset action depends on the action taken on the previous time step. Specifically, if the previous action was a reset, we know with certainty that the system is in the end state. Thus, taking another reset action is expected to produce an observation of 0. If the previous action was a float, then the probability of observing a 1 is identical to the probability of the system being in the end state.

With this in mind, we begin by considering the case when the last action taken is a reset.

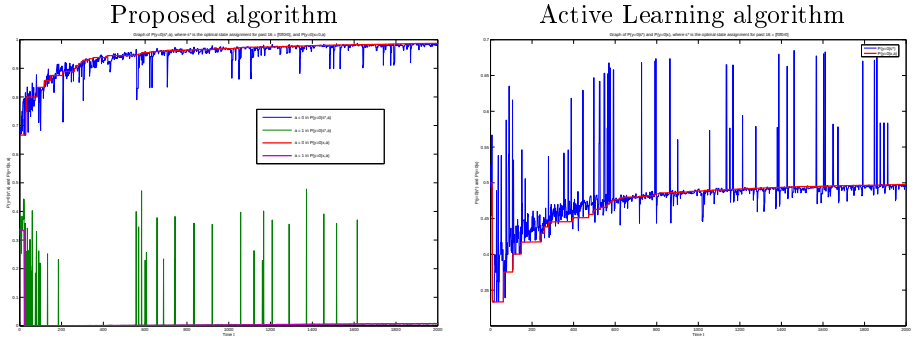

Figure 4: Future predictions. The horizontal axis in each plot is the number of iterations of the algorithm. The vertical axis differs between the plots. In the case of the proposed algorithm, it is $p(y_t^{fut} = 0 | x_t^{past} = [f0f0r0], a = float)$, in red, vs. $p(y_t^{fut} = 0 | s_t^*, a = float)$, in blue, and $p(y_t^{fut} = 0 | x_t^{past} = [f0f0r0], a = reset)$, in purple, vs. $p(y_t^{fut} = 0 | s_t^*, a = reset)$, in green. In the case of the original Active Learning algorithm the vertical axis is $p(y_t^{fut} = 0 | x_t^{past} = [f0f0r0])$, in red vs. $p(y_t^{fut} = 0 | s_t^*)$, in blue. Here, s_t^* is the optimal state to which the history is mapped.

In the case of the original Active Learning algorithm, the internal representation is independent of the action choice, and thus the future observations

depend on the state alone. Notice that the predictions given by the proposed internal representation converge very quickly to the value of the predictions given by the history alone, thus proving its accuracy. This convergence is faster than that of the original algorithm.

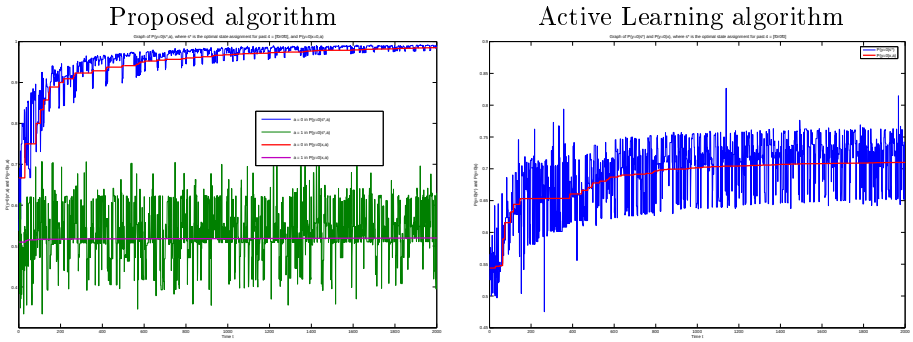

Figure 5: Future predictions. The horizontal axis in each plot is the number of iterations of the algorithm. The vertical axis differs for both plots. In the case of the proposed algorithm, it is $p(y_t^{fut} = 0|x_t^{past} = [f0r0f0], a = float)$, in red, vs. $p(y_t^{fut} = 0|s_t^*, a = float)$, in blue, and $p(y_t^{fut} = 0|x_t^{past} = [f0r0f0], a = reset)$, in purple, vs. $p(y_t^{fut} = 0|s_t^*, a = reset$, in green. In the case of the original Active Learning algorithm the vertical axis is $p(y_t^{fut} = 0|x_t^{past} = [f0r0f0])$, in red vs. $p(y_t^{fut} = 0|s_t^*)$, in blue. Here, s_t^* is the optimal state that the history gets mapped to.

Since the previous action was a float, the system can be found in either the end state, or the one to the left of it, each with a probability of 0.5. Thus, the probability of seeing an observation of 0 after taking another reset action is also 0.5. Again, the internal representation constructed by the proposed algorithm predicts future observations that are closer to those predicted from all the available data alone.

Clustering Finally, we are interested in whether the predictions given by the internal representation are consistent. In other words, whether the same type of histories consistently get mapped together, over time.

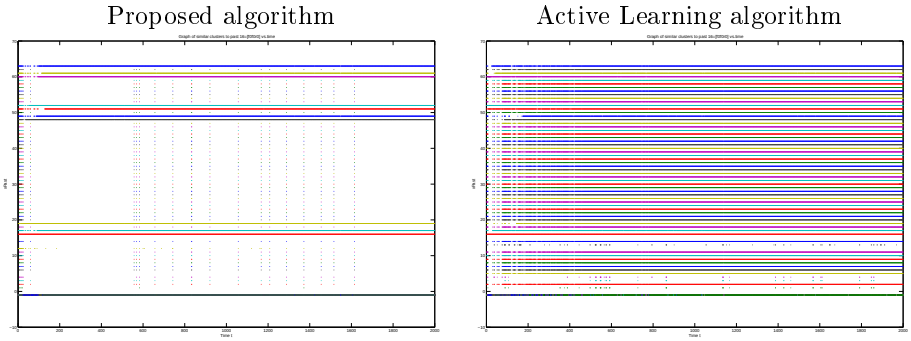

Figure 6: History clusters. The horizontal axis in each plot is the number of iterations of the algorithm. The vertical axis represents all the histories of length 3 that get mapped to the same state as $x_t^{past} = [f0f0r0]$.

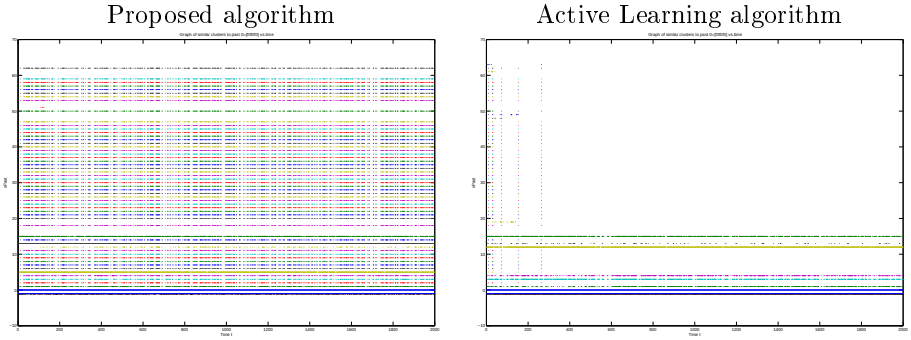

Figure 7: History clusters. The horizontal axis in each plot is the number of iterations of the algorithm. The vertical axis represents all the histories of length 3 that get mapped to the same state as $x_t^{past} = [f0r0f0]$.

Using the same argument used in the lossy compression case, we expect to see two different clusters being formed by the proposed algorithm. All the histories that are mapped to the same state as $x_t^{past} = [f0f0r0]$ are, in fact, histories that end in either an $r0$ or $r1$ action-observation pair. This is because the future prediction based on each of these histories is known with certainty. Consequently, the remaining histories will be mapped to the second state, as the future prediction in each of these cases is probabilistic. If these histories were to be mapped to two different states, rather than one, it may be possible to tell small probability differences in the case of the reset action. However, it is unrealistic to assume that this could happen, as the agent is learning from noisy and insufficient data. This however, is not the case when considering the original Active Learning algorithm. Since the state-history mapping is deterministic, in order to retain good predictive powers, the states need to be more spread out among the possible histories.

5 Conclusion

We have presented an information theoretic approach to the problem of predicting future observations from limited available data, by proposing that the internal representation, together with the action strategy employed by the agent, should retain maximal predictive powers, and compress the past histories seen by an agent. We have compared our proposed algorithm to that of [Still and Bialek,2004] in four different tasks. The experimental results illustrate an increase in the predictive power of the internal representation, as well as a faster convergence to the optimal state-history mapping. These experiments suggest that our approach of considering actions in creating the internal state representations is a viable solution for predicting partially observable systems, using a limited amount of data. Future work will analyze further problems that can be addressed by our approach, as well as compare our internal representation of the world to that constructed by a PSR.

References

[Still and Bialek,2004] Susanne Still and William Bialek. Active Learning and optimal predictions. 2004.

[Littman et al., 2002] Michael Littman, Richard S. Sutton and Satinder Singh. Predictive representations of state. In *NIPS 14*, pages 1555-1561, 2002.

[Singh et al., 2004] Satinder Singh, Michael R. James, and Matthew R. Ruddary. Predictive state representations: A new theory for modelling dynamical systems. In *Proceedings of UAI*, pages 512-519, 2004.

[Jaeger, 1998] Herbert Jaeger. Discrete-time, discrete-valued observable operator models: a tutorial. In *GMD Report 42*, 1998.

[Singh, Littman et. al] Satinder Singh, Michael L. Littman, Nicholas K. Jong, David Pardoe, Peter Stone. Learning Predictive State Representations. In *The Twentieth International Conference on Machine Learning (ICML-2003)*, 2003.

[Lovejoy, 1991] William S. Lovejoy. A survey of algorithmic methods for partially observable Markov decision processes. In *Annals of Operations Research, 28*, pages 47-65, 1991.

[Tishby et al., 1999] Naftali Tishby, Fernando Pereira, William Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference*, pages 368-377, 1999.