
Model-based Bayesian Reinforcement Learning with Tree-based State Aggregation

Cosmin Paduraru, Doina Precup, Stephane Ross and Joelle Pineau
McGill University
Montreal, Quebec, Canada

Model-based Bayesian RL provides an elegant way of incorporating model uncertainty for trading off between exploration and exploitation. Yet, significant work remains to be done for extending model-based Bayesian RL to continuous state spaces; in this paper, we present one such extension. The key feature of our approach is its search through the space of model structures, thus adapting not only the model parameters but also the structure itself to the problem at hand.

In order to do model-based Bayesian RL in continuous state spaces we need to define a class of models that can handle continuous states. In this work we focus on the simple state aggregation approach. In this approach, the model is composed of a partitioning (also referred to as an aggregation or discretization) of the state space, and a transition matrix which defines the probabilities of transitioning between partitions. For simplicity, we assume a small set A of discrete actions (although an extension to the continuous action case is also in the works), and we maintain a separate transition matrix for each action. We denote the partitioning by Ω , and the transition matrix for action a by Θ_a . We also assume that the transition distribution is uniform within the next state partition. Thus, we approximate the transition probabilities from state s to state s' by

$$P(s'|s, a) = \frac{1}{\mu(\omega(s'))} \Theta_a(\omega(s), \omega(s'))$$

where $\mu(\omega(s'))$ is the volume of $\omega(s')$, the partition containing s' . The partitions in Ω are created by splitting existing partitions, which makes a tree structure the most appropriate way of describing Ω .

State aggregation models come with several caveats. First, better discrimination between states always comes at the cost of decreased generalization, and vice versa. Second, this class of models will in most cases not contain the true transition model. Third, the probabilities of transitioning between partitions are not in fact multinomial. Indeed, these probabilities depend on the distribution of states inside a partition, which makes them have a non-tractable dependency on the history and the action selection mechanism.

Bayesian learning actually enables us to get around some of these problems. The main reason for this is that our model-based Bayesian RL algorithm causes the distribution over partitionings to change over time in response to the observed data (this is also done heuristically in other state aggregation work, such as Munos and Moore (1999)). Thus, the first issue could be handled by having good generalization (large partitions) in the beginning, and then discretizing more as sufficient data is available. The second and third issue should also be less problematic as the discretization becomes finer.

For the technical implementation of these ideas, we draw inspiration from two existing papers: Ross and Pineau (2008) and Chipman, George and McCulloch (1998). Ross and Pineau proposed a Bayesian RL algorithm for factored MDPs, which allows them to handle finite MDPs with large state spaces. Chipman, George and McCulloch describe a Bayesian approach to supervised learning that uses tree-based state aggregation. In this paper, we only present an overview of our implementation, and leave most of the details for a future lengthier publication.

Similarly to previous work, we break up the probability of a model as $P(\Omega, \Theta) = P(\Omega)P(\Theta|\Omega)$, where $\Theta = \{\Theta_a|a \in A\}$. For the posterior, we use the same factorization, but additionally condition on the history.

Since there is an infinite number of possible discretizations, the distribution $P(\Omega)$ over the discretizing trees cannot be maintained explicitly, so an approximate, particle filter style approach is taken. A set of trees is initially sampled from the prior (which gives more weight to smaller trees), and the probabilities of these structures is maintained. When the likelihood of the maintained set of trees falls below some threshold, a new set of trees is sampled using the well-known Metropolis-Hastings algorithm, as described by Chipman et al. (1998). This requires that previous transitions be stored, because we need to compute the likelihood of the data under different models.

The distribution over partition-to-partition transition models $P(\Theta|\Omega)$ is represented as a Dirichlet and maintained by updating counts. At several points in the algorithm (updating the probabilities of the structures, computing the Metropolis-Hastings ratio) we need to marginalize over Θ . Fortunately, this can be done in closed form.

For selecting the optimal action, we have to find an approximation of the optimal policy in the resulting Bayes-adaptive MDP. Our current approach for finding such a policy is sample-based online planning with uniform action sampling (as used for instance by Ross and Pineau (2008)), although we are also considering more informed action selection methods (e.g. Castro and Precup (2007)).

There are several existing Bayesian RL algorithms that work in continuous state spaces. Ross et al. (2008) present a model-based Bayesian RL method for continuous-state, continuous-action POMDPs; however, it requires the transition model to be Gaussian. The Gaussian process temporal difference work of Engel et al. (2005) can also handle continuous state spaces by representing the value function as a kernel-based Gaussian.

Note that a more general class of models could have been produced by having different aggregations over s and s' . Even more general would be to have the aggregation over s' be conditioned on the partition containing s . Also, one could have different partitionings for different actions.

We are currently evaluating our method on standard continuous-state reinforcement learning problems, such as mountain car and puddle world.

References

- Castro, P. S., & Precup, D. (2007). Using linear programming for Bayesian exploration in Markov Decision Processes. *Proceeding of the 20th International Joint Conference on Artificial Intelligence*.
- Chipman, H., George, E., & McCulloch, R. (1998). Bayesian CART Model Search. *Journal of the American Statistical Association*, 935–960.
- Engel, Y., Mannor, S., & Meir, R. (2005). Reinforcement learning with gaussian processes. *Proceedings of the International Conference on Machine Learning (ICML)*.
- Munos, R., & Moore, A. (1999). Variable resolution discretization for high-accuracy solutions of optimal control problems. *International Joint Conference on Artificial Intelligence*.
- Ross, S., Chaib-draa, B., & Pineau, J. (2008). Bayesian reinforcement learning in continuous pomdps with application to robot navigation. *Proceedings of the IEEE International Conference on Robotics and Automation*.
- Ross, S., & Pineau, J. (2008). Model-based bayesian reinforcement learning in large structured domains. *Proceedings of the 24th UAI*.