Global Reinforcement Learning

Milen Pavlov, Pascal Poupart
{mpavlov, ppoupart}@uwaterloo.ca
David R. Cheriton School of Computer Science
University of Waterloo
200 University Ave. W, Waterloo, ON N2L 3G1

When applying reinforcement learning to real world problems it is desirable to make use of any prior information to speed up the learning process. Unfortunately, there does not exist any framework for encoding prior knowledge in RL nor developing general algorithms that make use of it. The state of the art consists of designing algorithms that employ ad-hoc shortcuts, which often rely on problem-specific assumptions. As a result, it is hard to generalize these approaches to many problems. In addition, prior information is often only used implicitly through those assumptions, which makes it difficult to validate the correctness of the prior information.

We propose a new framework, called Global Reinforcement Learning, to facilitate the encoding and exploitation of prior knowledge in a general and principled way. In many RL problems, the learner often has some prior belief about what actions are good in some states, as well as the states that are likely to be reached as a result of the execution of some actions. We employ a Bayesian approach, whereby a wide range of beliefs can be encoded as distributions over policies and environments (e.g., transition dynamics and reward functions). More precisely, we explain how to come up with a joint distribution over policies and environments by specifying constraints reflecting prior knowledge and taking into account the intuition that a policy-environment pair should have high probability when the policy achieves high reward in the associated environment. We also explain how to resolve inconsistencies in the prior knowledge used to specify this joint distribution. Once a prior distribution is obtained, it can be updated based on the trajectories of states and actions observed, using Bayesian RL techniques similar to those proposed in [1, 2]. The overall approach allows us to specify prior information about the environment and learn about the policy, or to specify prior information about the policy and learn about the environment, or to specify partial information about both and learn about both. Hence the name Global Reinforcement Learning.

This is preliminary work, which still needs to be verified empirically.

References

[1] P. Poupart, N. Vlassis, J. Hoey and K. Regan. An analytic solution to discrete Bayesian reinforcement learning. *In Proc. 23rd Int. Conf. on Machine Learning*, Pittsburgh, USA, 2006.

[2] R. Dearden, N. Friedman and S. Russell. Bayesian Q-Learning. *In Proc. 15th Nat. Conf. on Artificial Intelligence*, 1998.