
Relational Reinforcement Learning for Classical Planning (Extended Abstract)

Alan Fern
SungWook Yoon
Robert Givan

AFERN@PURDUE.EDU
SY@PURDUE.EDU
GIVAN@PURDUE.EDU

School of Electrical and Computer Engineering, Purdue University

AI researchers have long studied algorithms for planning and learning-to-plan within highly structured, relational domains. A traditional criticism of these “classical” domains is that they typically assume an idealized, deterministic model. This, in part, has led AI researchers to studying planning and learning within a decision-theoretic framework, which explicitly handles stochastic environments and generalized notions of reward/goals. However, most decision-theoretic algorithms are based on explicit or propositional domain models, and so far have not shown the ability to scale up to the kind of domains studied in classical planning.

Our recent work (Fern et al., 2003; Fern et al., 2004) addresses this gap between classical and decision-theoretic planning—developing algorithms for very large, relationally structured, decision-theoretic planning problems. In this talk, we will provide an overview of this work. In addition, we will discuss some of our current limitations and directions for future work.

Our work can be viewed as a form of relational reinforcement learning (RRL) where we assume a strong simulation model of the environment. That is, we assume access to a black-box simulator, which we can provide any (relationally represented) state/action pair and receive a sample from the appropriate next-state and reward distributions. The goal is to interact with the simulator in order to learn a good policy, i.e. one that achieves high expected total reward.

Here we are particularly interested in RRL for large classical planning domains, and their stochastic variants, such as the blocks world and logistics problems. There are at least two primary challenges in designing an RRL technique for these problems. First, value functions can be extremely difficult to represent and

learn in these domains. However, most RRL techniques are based, in part, on representing and learning (approximate) value function representations. Second, non-zero reward is typically only received at goal states and consequently is extremely sparse and unlikely to be achieved by random wandering. Thus, bootstrapping the learning process is a critical issue for RRL.

To address the first challenge above, we present a form of approximate policy iteration (API) that completely avoids representing value functions. Existing forms of API typically represent policies via value functions and produce a sequence of value-function regression problems. Instead, our approach represents policies directly as state-action mappings and learns improved policies via a sequence of cost-sensitive classification problems. We argue that reducing to classification rather than regression can have practical benefits as it is often significantly easier to specify an effective policy space via a direct policy-language bias rather than via a cost-function bias, particularly in relational domains.

To address the second challenge, i.e. learning with sparse reward, we utilize a combination of two ideas. First, we leverage heuristics, developed for state-of-the-art classical planners, in order to provide an approximate reward signal to guide learning. Second, we introduce the idea of learning from random walks, which can be viewed as a form of reward shaping.

Empirically we use our techniques to learn good policies for a number of benchmark classical planning domains (both deterministic and stochastic variants). The resulting policies can be viewed as automatically learned, high quality domain-specific planners.

References

- Fern, A., Yoon, S., & Givan, R. (2003). Approximate policy iteration with a policy language bias. *NIPS*.
- Fern, A., Yoon, S., & Givan, R. (2004). Learning domain-specific control knowledge from random walks. *ICAPS*.