

A Theory of State Abstraction for Reinforcement Learning

David Abel

Department of Computer Science
Brown University
david_abel@brown.edu

Abstract

Reinforcement learning presents a challenging problem: agents must generalize experiences, efficiently explore the world, and learn from feedback that is delayed and often sparse, all while making use of a limited computational budget. Abstraction is essential to all of these endeavors. Through abstraction, agents can form concise models of both their surroundings and behavior, supporting effective decision making in diverse and complex environments. To this end, the goal of my doctoral research is to characterize the role abstraction plays in reinforcement learning, with a focus on state abstraction. I offer three desiderata articulating what it means for a state abstraction to be useful, and introduce classes of state abstractions that provide a partial path toward satisfying these desiderata. Collectively, I develop theory for state abstractions that can 1) preserve near-optimal behavior, 2) be learned and computed efficiently, and 3) can lower the time or data needed to make effective decisions. I close by discussing extensions of these results to an information theoretic paradigm of abstraction, and an extension to hierarchical abstraction that enjoys the same desirable properties.

1 Introduction

The focus of my doctoral research is on clarifying the *representational* practices that underlie effective Reinforcement Learning (RL), drawing on Information Theory, Computational Complexity, and Computational Learning Theory. The guiding question of my research is: “*How do intelligent agents come up with the right abstract understanding of the worlds they inhabit?*”, as pictured in Figure 1. I study this question by isolating and addressing its simplest unanswered forms through a mixture of theoretical analysis and experimentation.

My interest in this question stems from its foundational role in many aspects of learning and decision making: agents can’t model *everything* in their environment, but must necessarily pick up on *something* about their surroundings in order to explore, plan far into the future, generalize, solve credit assignment, communicate, and efficiently solve problems. Abstraction is essential to all of these endeavors: through abstraction, agents can construct models of both their surroundings and behavior that are compressed and useful. The

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

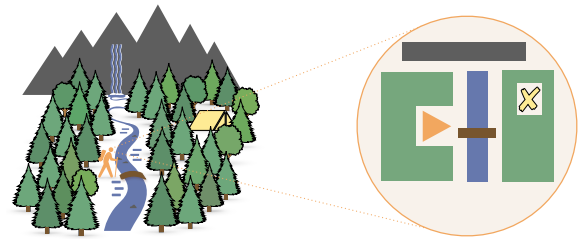


Figure 1: The process of abstraction.

goal of my research is to explain the nature of effective abstraction, hopefully giving rise to a theoretical understanding of what underlies effective RL more generally.

To realize this goal, I propose three desiderata that characterize what it means for an abstraction to be useful for RL:

- (D1) **SUPPORT EFFICIENT DECISION MAKING:** The abstraction enables fast planning and efficient RL.
- (D2) **PRESERVE SOLUTION QUALITY:** Solutions produced from the abstracted model should be useful enough for solving the desired problems.
- (D3) **EASY TO CONSTRUCT:** Creating the abstractions should not require an unrealistic statistical or computational budget.

To date, no unifying theory of abstraction adheres to all three desiderata, or clarifies how to flexibly trade-off between the three prescribed properties.

2 Completed Research

My recent published work focuses on a theory of *state abstraction* that endeavors to satisfy the desiderata.

2.1 State Abstraction

The goal of state abstraction is to reduce the size of a given problem’s state space by grouping together similar states in a way that doesn’t change the essence of the underlying problem being solved (Bertsekas and Castanon 1989; Singh, Jaakkola, and Jordan 1995; McCallum 1996; Dietterich 2000; Andre and Russell 2002; Givan, Dean, and Greig 2003; Li, Walsh, and Littman 2006), dating back to work in approximating dynamic programs (Whitt 1978).

Concretely, a state abstraction maps each true environmental state into an *abstract* state:

Definition 1 (State Abstraction): A **state abstraction**, $\phi : \mathcal{S} \rightarrow \mathcal{S}_\phi$, maps each environmental state $s \in \mathcal{S}$ to an abstract state $s_\phi \in \mathcal{S}_\phi$, where typically $|\mathcal{S}_\phi| \ll |\mathcal{S}|$.

In previous work, we introduce new theory for approximate state abstraction in RL and planning (Abel, Hershkowitz, and Littman 2016). Our main result clarifies when *approximate* abstractions can still preserve near-optimal behavior. At a high level:

Theorem 1 Consider an approximate abstraction $\phi_{f,\varepsilon}$, where, for $\varepsilon \in \mathbb{R}_{\geq 0}$ and for any two states $s_1, s_2 \in \mathcal{S}$:

$$\phi_{f,\varepsilon}(s_1) = \phi_{f,\varepsilon}(s_2) \implies |f(s_1) - f(s_2)| \leq \varepsilon, \quad (1)$$

with $f : \mathcal{S} \rightarrow \mathbb{R}$ some function on state. There exist choices of f such that value is preserved under the abstraction:

$$\forall s \in \mathcal{S} : V^*(s) - V^{\pi_\phi^*}(s) \leq \varepsilon \eta_f, \quad (2)$$

for some constant $\eta_f \in \mathbb{R}_{\geq 0}$ that depends on choice of f .

The takeaway here is that approximate forms of abstraction (those where $\varepsilon > 0$ on the right hand side of Equation 1) can still retain near-optimal behavior (depending on choice of function f). So, state abstraction can satisfy the D2.

Our follow up work extends the above result to the lifelong learning setting in which an agent must interact with a collection of problems all sampled from the same distribution (Abel et al. 2018). Our main results state that: (1) There exist families of state abstractions that can be computed efficiently, and (2) These abstractions can also be used in the lifelong setting in a way that preserves near-optimal behavior with high probability. We conduct simple experiments to corroborate these results, suggesting that state abstractions can lower sample complexity in lifelong RL. The takeaway is that state abstractions can satisfy D2 and D3.

2.2 Current Work

Our recently completed work extends the above state abstraction theory to an information theoretic framework (Abel et al. 2019). We draw a parallel between compression, as understood in Information Theory, and state abstraction, as studied in RL, to offer the first formalism and analysis of the trade-off between compression and performance made by state abstraction. We build on Rate-Distortion theory (Shannon 1948) and the Information Bottleneck method (Tishby, Pereira, and Bialek 1999) to develop an algorithm for computing state abstractions that approximates the optimal trade-off between compression and performance. In the future, we hope to show how the appropriately compressed model can lower the computational and sample complexity of planning and learning. With the right kind of additional analysis, we hope to show the sense in which abstractions can trade-off between D1 and D2.

3 Future Work

There are three specific questions I hope to address before finishing my dissertation. First, as described in Section 2,

we have established theory that suggests there exist state abstractions that can trade-off between D1 and D2. However, whether there is a single theory can satisfy (or trade-off between) all three desiderata remains an open question. My first goal is to answer this question by extending the information theoretic work described in the previous section. Second, all of my state abstraction work so far typically focuses on discrete and often finite state spaces. I am interested in extending our analysis to the case where the underlying state space is continuous. Third, I plan to investigate the relationship between state and action abstraction. Specifically, I am working with a simple scheme for clarifying the conditions under which state-action abstraction pairs can inform hierarchical abstractions that satisfy the desiderata.

References

- Abel, D.; Arumugam, D.; Lehnert, L.; and Littman, M. L. 2018. State Abstractions for Lifelong Reinforcement Learning. In *Proceedings of the International Conference on Machine Learning*.
- Abel, D.; Arumugam, D.; Asadi, K.; Jinnai, Y.; Littman, M. L.; and Wong, L. L. 2019. State Abstraction as Compression in Apprenticeship Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press.
- Abel, D.; Hershkowitz, D. E.; and Littman, M. L. 2016. Near Optimal Behavior via Approximate State Abstraction. In *Proceedings of the International Conference on Machine Learning*.
- Andre, D., and Russell, S. J. 2002. State Abstraction for Programmable Reinforcement Learning Agents. In *In Proceedings of the AAAI Conference on Artificial Intelligence*, 119–125.
- Bertsekas, D. P., and Castanon, D. A. 1989. Adaptive Aggregation Methods for Infinite Horizon Dynamic Programming. *IEEE transactions on Automatic Control* 34(6):589–598.
- Dietterich, T. G. 2000. State abstraction in MAXQ hierarchical reinforcement learning. In *Advances in Neural Information Processing Systems*, 994–1000.
- Givan, R.; Dean, T.; and Greig, M. 2003. Equivalence notions and model minimization in Markov decision processes. *Artificial Intelligence* 147(1-2):163–223.
- Li, L.; Walsh, T. J.; and Littman, M. L. 2006. Towards a Unified Theory of State Abstraction for MDPs. In *Proceedings of the International Symposium on Artificial Intelligence and Mathematics*.
- McCallum, A. K. 1996. *Reinforcement Learning with Selective Perception and Hidden State*. Ph.D. Dissertation, University of Rochester. Dept. of Computer Science.
- Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3):379–423.
- Singh, S. P.; Jaakkola, T.; and Jordan, M. I. 1995. Reinforcement Learning with Soft State Aggregation. In *Advances in Neural Information Processing Systems*, 361–368.
- Tishby, N.; Pereira, F. C.; and Bialek, W. 1999. The Information Bottleneck Method. *Allerton Conference on Communication, Control, and Computing* 368–377.
- Whitt, W. 1978. Approximations of dynamic programs, I. *Mathematics of Operations Research* 3(3):231–243.