

Chapter 13

Relational Representations and Traces for Efficient Reinforcement Learning

Eduardo F. Morales

National Institute of Astrophysics, Optics and Electronics, México

Julio H. Zaragoza

National Institute of Astrophysics, Optics and Electronics, México

ABSTRACT

This chapter introduces an approach for reinforcement learning based on a relational representation that: (i) can be applied over large search spaces, (ii) can incorporate domain knowledge, and (iii) can use previously learned policies on different, but similar, problems. The underlying idea is to represent states as sets of first order relations, actions in terms of those relations, and to learn policies over such generalized representation. It is shown how this representation can produce powerful abstractions and that policies learned over this generalized representation can be directly applied, without any further learning, to other problems that can be characterized by the same set of relations. To accelerate the learning process, we present an extension where traces of the tasks to be learned are provided by the user. These traces are used to select only a small subset of possible actions increasing the convergence of the learning algorithms. The effectiveness of the approach is tested on a flight simulator and on a mobile robot.

INTRODUCTION

Sequential decision making under uncertainty has been studied in fields such as decision-theoretic planning (Puterman, 1994), reinforcement learning (Sutton & Barto, 1989) and economics. The idea is to decide on each state, which is the best

action to perform, given uncertainty on the outcomes of the actions and trying to obtain the maximum benefit in the long run. Optimal sequence decision making can be formalized as a Markov decision process or MDP, where several dynamic programming techniques have been developed (Puterman, 1994). These techniques, obtain optimal policies, i.e., the best action to perform on each state, however, they require knowing a

DOI: 10.4018/978-1-4666-4607-0.ch013

transition model. Reinforcement learning (RL), on the other hand, can learn optimal or near optimal policies while interacting with an external environment, without a transition model (Sutton & Barto, 1989). In either case, the representation used in traditional models for MDPs require all the possible states to be explicitly represented. This restricts its applicability to only very simple domains as the number of possible states grows exponentially with the number of relevant variables. In order to cope with this curse of dimensionality several approaches have been suggested in recent years. Some of these methods include, function approximation (e.g., (Chapman & Kaelbling, 1991)), hierarchical (e.g., (Dietterich, 2000)) and temporal abstraction (e.g., (Sutton, Precup, & Singh, 1999a)), and factored representations (e.g., (Kaelbling, Littman, & Cassandra, 1998)). Despite recent advances, there is still on-going research into trying to deal more effectively with large search spaces, to incorporate domain knowledge, and to transfer previously learned policies to other related problems. Most work, however, uses a propositional framework and still do not scale well as many domains are more clearly defined in terms of objects and relations.

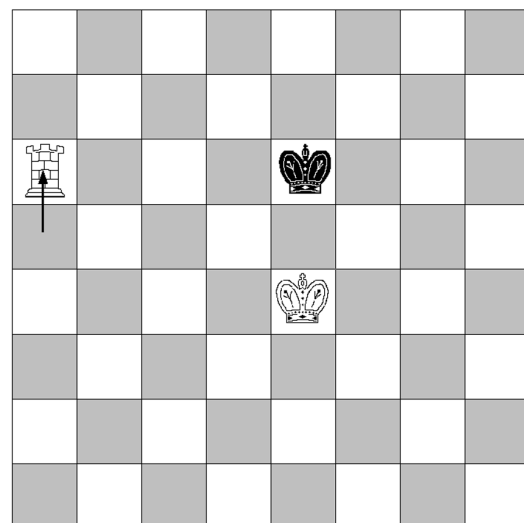
Suppose we want to learn a policy to play a simple chess endgame from a particular side. Even for learning how to win in a simple and deterministic endgame like king-rook vs. king (KRK), there are more than 175,000 not-in-check legal positions. The number of possible actions for the king-rook side is in general 22 (8 for the king and 14 for the rook), which sum up to nearly 4 million possible state-action pairs. Even with modern computers, learning directly in this representation is just too slow. In this domain, however, there are many states that are essentially the same, in the sense that they all share the same set of relations. For a chess player, the exact location of each piece is not as important as the relations that hold between the pieces to decide which movement to perform (see also (Charness, 1977, Groot, 1965)). For instance, in the KRK domain, whenever the two

kings are *in_opposition* (in the same rank or file with one square in between), the rook is in the file/rank that divides both kings, and the rook is at least two squares apart from the opposite king (i.e., it is safe to check), a good move is to check the opposite king to force it to move towards a border (see Figure 1). This action is applicable to all the chess board positions where the previously mentioned relations hold between the pieces, and already captures more than 1,000 chess positions with a white rook. In fact, it is applicable to chess boards of different sizes.

In this chapter, a relational representation for reinforcement learning is used to represent a small set of abstract states with a set of properties (e.g., *in_opposition*, *rook_divides*, etc.), represent actions in terms of such properties (e.g., If *in_opposition* And *rook_divides* Then *check*) and learn which action to apply on each state.

There has been a growing interest in using first-order logic for modeling MDPs (e.g., (Driessens & Ramon, 2003, Dzeroski, Raedt, & Driessens, 2001, Fern, Yoon, & Givan, 2003, Morales, 2003, Otterlo, 2009)). The use of a relational representation has many advantages.

Figure 1. Forcing the opponent king to move towards a border



24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/relational-representations-and-traces-for-efficient-reinforcement-learning/84898

Related Content

Structural Condition Monitoring with the Use of the Derivative-Free Nonlinear Kalman Filter

Gerasimos Rigatos and Argyris Soldatos (2015). *Handbook of Research on Advancements in Robotics and Mechatronics* (pp. 377-407).

www.irma-international.org/chapter/structural-condition-monitoring-with-the-use-of-the-derivative-free-nonlinear-kalman-filter/126022

The Semantic Dominance of Emotional Templates in Cognitive Structures

Tom Adi (2015). *International Journal of Synthetic Emotions* (pp. 1-13).

www.irma-international.org/article/the-semantic-dominance-of-emotional-templates-in-cognitive-structures/160800

Modeling, Simulation and Motion Cues Visualization of a Six-DOF Motion Platform for Micro-Manipulations

Umar Asif and Javaid Iqbal (2013). *Advanced Engineering and Computational Methodologies for Intelligent Mechatronics and Robotics* (pp. 137-151).

www.irma-international.org/chapter/modeling-simulation-motion-cues-visualization/76445

Intelligent Wildlife Tracking Using Ubiquitous Technological Suite

Mayur Sunil Jawalkar, Nayan Desale, Fanil Suratwala, Amol Lamkhade, Parikshit N. Mahalle and Poonam N. Railkar (2017). *International Journal of Synthetic Emotions* (pp. 44-59).

www.irma-international.org/article/intelligent-wildlife-tracking-using-ubiquitous-technological-suite/181640

Hindi Optical Character Recognition and Its Applications

Rashmi Gupta, Dipti Gupta, Megha Dua and Manju Khari (2017). *Detecting and Mitigating Robotic Cyber Security Risks* (pp. 28-39).

www.irma-international.org/chapter/hindi-optical-character-recognition-and-its-applications/180059