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

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
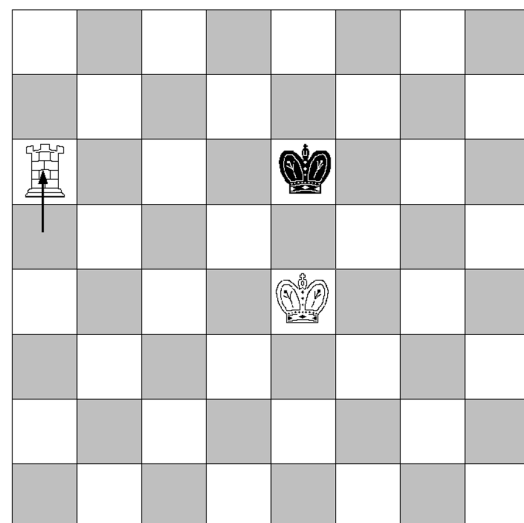
Related Content

Kinematics and Dynamics Modeling of a New 4-DOF Cable-Driven Parallel Manipulator
Hamoon Hadian, Yasser Amooshahiand Abbas Fattah (2011). *International Journal of Intelligent Mechatronics and Robotics (pp. 44-60).*
www.irma-international.org/article/kinematics-dynamics-modeling-new-dof/61156

The Compilation and Validation of a Collection of Emotional Expression Images Communicated by Synthetic and Human Faces
Louise Lawrenceand Deborah Abdel Nabi (2013). *International Journal of Synthetic Emotions (pp. 34-62).*
www.irma-international.org/article/the-compilation-and-validation-of-a-collection-of-emotional-expression-images-communicated-by-synthetic-and-human-faces/97677

Adaptive Control of Bilateral Teleoperation with Time Delay
Ali Shahdiand Shahin Sirouspour (2012). *International Journal of Intelligent Mechatronics and Robotics (pp. 1-27).*
www.irma-international.org/article/adaptive-control-bilateral-teleoperation-time/64216

Multi-Robot Swarm for Cooperative Scalar Field Mapping
Hung Manh La (2016). *Handbook of Research on Design, Control, and Modeling of Swarm Robotics (pp. 383-395).*
www.irma-international.org/chapter/multi-robot-swarm-for-cooperative-scalar-field-mapping/142009

Digital Control Theory Application and Signal Processing in a Laser Scanning System Applied for Mobile Robotics
Miguel Reyes-Garcia, Cesar Sepulveda-Valdez, Oleg Sergiyenko, Moisés Rivas-López, Julio Cesar Rodríguez-Quiñonez, Wendy Flores-Fuentes, Daniel Hernandez-Balbuena, Juan-Ivan Nieto-Hipolito, Fabian N. Murrieta-Rico, Lars Lindnerand Mykhailo Ivanov (2020). *Control and Signal Processing Applications for Mobile and Aerial Robotic Systems (pp. 215-265).*
www.irma-international.org/chapter/digital-control-theory-application-and-signal-processing-in-a-laser-scanning-system-applied-for-mobile-robotics/243767