Reinforcement Learning and Markov Decision Processes
A Few Pointers to the Field

Martijn van Otterlo (otterlo@cs.kuleuven.be)

Compiled for the SIKS course on "Learning and Reasoning" – May 2009

Abstract

Reinforcement learning (RL) has developed into a large research field. The current state-of-the-art is comprised of several subfields dealing with, for example, hierarchical abstraction and relational representations. This overview is targeted at researchers interested in RL who want to know where to start when studying RL in general, and where to start *within* the field of RL when faced with specific problem domains. This overview is by no means complete, nor does it describe all relevant texts. In fact, there are many more. The main function of this overview is to provide a reasonable amount of good entry points into the rich field of RL. All texts are widely available and most of them are online.

General and Introductory Texts

There are many texts that introduce the exciting field of RL *and Markov decision processes (see for example the mentioned PhD theses at the end of this overview). Furthermore, many recent* AI *and machine learning textbooks cover basic* RL. *Some of the core texts in the field are the following.*

- M. L. Puterman. *Markov Decision Processes—Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, 1994
- D. P. Bertsekas and J. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA, 1996
- L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996
- S. S. Keerthi and B. Ravindran. Reinforcement learning. In E. Fiesler and R. Beale, editors, *Handbook of Neural Computation*, chapter C3. Institute of Physics and Oxford University Press, New York, New York, 1997
- R. S. Sutton and A. G. Barto. *Reinforcement Learning: an Introduction*. The MIT Press, Cambridge, 1998
- C. Boutilier, T. Dean, and S. Hanks. Decision theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94, 1999
- M. van Otterlo. *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for Adaptive Sequential Decision Making under Uncertainty in First-Order and Relational Domains*. IOS Press, Amsterdam, The Netherlands, 2009

The book by Sutton and Barto is available online, for free.
You can find it at http://www.cs.ualberta.ca/∼ sutton/book/the-book.html

Function Approximation, Generalization and Abstraction

Because most problems are too large to represent explicitly, the majority of techniques in current RL *research employs some form of generalization, abstraction or function approximation. Ergo, there are innumerable texts that deal with these matters. Some interesting starting points are the following.*

▶ J. N. Tsitsiklis and B. van Roy. An analysis of temporal-difference learning with function approximation. *IEEE Transactions on Automatic Control*, 42:674–690, 1997
▶ R. Fitch, B. Hengst, D. Suc, G. Calbert, and J. Scholz. Structural abstraction experiments in reinforcement learning. In *Proceedings of AI-2005*, 2005
▶ L. Li, T. Walsh, and M. L. Littman. Towards a unified theory of state abstraction for MDPs. In *Proceedings of AI-MATH*, 2006

Factored Representations

Factored representations have their roots in graphical models such as Bayesian networks. Such structured representations allow for powerful abstractions and efficient algorithms to solve large problems. The following texts describe ways to deal with factored MDPs*, using both model-based and model-free solution algorithms. Furthermore, the final reference addresses the important topic of automatic generation of factorizations.*

▶ C. Boutilier, R. Dearden, and M. Goldszmidt. Stochastic dynamic programming with factored representations. *Artificial Intelligence*, 121(1–2):49–107, 2000
▶ C. Guestrin, D. Koller, R. Parr, and S. Venkataraman. Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research (JAIR)*, 19:399–468, 2003
▶ B. Sallans and G. E. Hinton. Reinforcement learning with factored states and actions. *Journal of Artificial Intelligence Research (JAIR)*, 5:1063–1088, 2004
▶ T. Degris, O. Sigaud, and P. H. Wuillemin. Learning the structure of factored markov decision processes in reinforcement learning problems. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2006

Hierarchical Reinforcement Learning

Hierarchical RL *is concerned with temporal structure and hierarchical task structures in the context of sequential decision making. Some problems can be decomposed into subproblems that can be solved relatively independently. Some of the first and well-known systems are described in the following texts.*

▶ P. Dayan and G. E. Hinton. Feudal reinforcement learning. In *Proceedings of the Neural Information Processing Conference (NIPS)*, pages 271–278, 1993
▶ R. S. Sutton, D. Precup, and S. Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1–2):181–211, 1999
▶ T. G. Dietterich. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research (JAIR)*, 13:227–303, 2000

Two recent overviews are the following. The first is quite technical and detailed about subtle concepts, but limited in scope. The second is more conceptual in nature, and describes a wider scope of methods.

▶ A. Barto and S. Mahadevan. Recent advances in hierarchical reinforcement learning. *Discrete event systems*, 13(4):341–379, 2003
▶ M. R. K. Ryan. Hierarchical decision making. In J. Si, A. G. Barto, W. B. Powell, and D. Wunsch, editors, *Handbook of Learning and Approximate Dynamic Programming*. Wiley-IEEE Press, Piscataway, NJ, 2004

Partially Observable Problems

Partially observable MDP*s model problems where the full environment's state cannot be observed. Solutions to these problems can sometimes be obtained by ignoring partial observability, by using memory, by analyzing history, by policy-based methods or by building probabilistic belief states. Some of the first algorithms for* POMDP*s focus on exact solutions, whereas much of the recent research has focused on so-called point-based approximation algorithms.*

▶ L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stoachastic domains. *Artificial Intelligence*, 101:99–134, 1998
▶ J. Pineau, G. Gordon, and S. Thrun. Point-based approximations for fast POMDP solving. *Journal of Artificial Intelligence Research (JAIR)*, 2005

Lately, research has focused on predictive models as well. They come in the form of predictive representations of state and temporal-difference networks.

▶ M. L. Littman, R. S. Sutton, and S. Singh. Predictive representations of state. In *Proceedings of the Neural Information Processing Conference (NIPS)*, 2001
▶ R. S. Sutton and B. Tanner. Temporal-difference networks. In *Proceedings of the Neural Information Processing Conference (NIPS)*, 2005

Relational Reinforcement Learning

The use of powerful relational (or: object-based) knowledge representation frameworks is one of the most recent directions in RL. *Employing first-order logic in* RL *contexts poses a number of difficult knowledge representation problems, but at the same time creates many new opportunities for generalization over objects in complex worlds, for transfer of knowledge, and for learning in indefinite (or even infinite) size domains. Some of the first arguments for relational* RL *as well as the first model-based and model-free algorithms can be found in the following texts.*

▶ L. P. Kaelbling, T. Oates, N. Hernandez, and S. Finney. Learning in worlds with objects. In *The AAAI Spring Symposium*, 2001
▶ C. Boutilier, R. Reiter, and B. Price. Symbolic dynamic programming for first-order MDP's. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 690–697, 2001
▶ S. Džeroski, L. De Raedt, and K. Driessens. Relational reinforcement learning. *Machine Learning*, 43:7–52, 2001

More recent descriptions of challenges and motivations, as well as a recent survey of the field so far, can be found in the following texts.

▶ P. Tadepalli, R. Givan, and K. Driessens. Relational reinforcement learning: An overview. In *Proceedings of the Workshop on Relational Reinforcement Learning at ICML'04*, 2004
▶ M. van Otterlo. A survey of reinforcement learning in relational domains. Technical Report TR-CTIT-05-31, CTIT, University of Twente, Enschede, The Netherlands, July 2005
▶ M. van Otterlo. *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for Adaptive Sequential Decision Making under Uncertainty in First-Order and Relational Domains*. IOS Press, Amsterdam, The Netherlands, 2009

Some Recent Reinforcement Learning Theses

Many PhD theses contain excellent treatments of, and introductions to, topics in RL. *Some recent theses that may be useful, are the following ones.*

- M. A. Wiering. *Explorations in Efficient Reinforcement Learning*. PhD thesis, Faculteit der Wiskunde, Informatica, Natuurkunde en Sterrenkunde, Universiteit van Amsterdam, 1999
- S. I. Reynolds. *Reinforcement Learning with Exploration*. PhD thesis, The School of Computer Science, The University of Birmingham, UK, 2002
- B. Ratitch. *On Characteristics of Markov Decision Processes and Reinforcement Learning in Large Domains*. PhD thesis, The School of Computer Science, McGill University, Montreal, 2005
- M. van Otterlo. *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains*. PhD thesis, Department of Computer Science, University of Twente, Enschede, The Netherlands, 2008. may, 512pp

The following ones are full theses about more specific topics such as factored representations, hierarchical RL*, relational* RL*, function approximation,* POMDP*s and multi-agent* RL:

- C. Guestrin. *Planning Under Uncertainty in Complex Structured Environments*. PhD thesis, Computer Science Department, Stanford University, August 2003
- M. R. K. Ryan. *Hierarchical Reinforcement Learning: A Hybrid Approach*. PhD thesis, University of NSW, School of Computer Science and Engineering, Sidney, Australia, 2004
- K. Driessens. *Relational reinforcement learning*. PhD thesis, Department of Computer Science, Catholic University of Leuven, Belgium, May 2004
- B. Bakker. *The State of Mind: Reinforcement Learning with Recurrent Neural Networks*. PhD thesis, University of Leiden, The Netherlands, Faculteit Wiskunde, Natuurwetenschappen en Geneeskunde, 2004
- M. T. J. Spaan. *Approximate Planning under Uncertainty in Partially Observable Environments*. PhD thesis, Universiteit van Amsterdam, 2006
- J. R. Kok. *Coordination and Learning in Cooperative Multiagent Systems*. PhD thesis, Universiteit van Amsterdam, 2006

RL *is a very active research field. Additional topics that we have not mentioned include i) the connection between learning and planning, ii) the use of Markov decision processes in the field of operations research, iii) large-scale applications of* RL*, iv) the transfer of learned knowledge to other, related tasks, and v) theoretical advances, for example concerning convergence, sample complexity and computational complexity of algorithms. Many of these advances are published in journals such as "Journal of Machine Learning Research" (*JMLR*) and "Journal of Artificial Intelligence Research" (*JAIR*), and conferences such as the European and international conferences on machine learning (*ECML *and* ICML*), "Uncertainty in Artificial Intelligence" (*UAI*) and "Neural Information Processing" (*NIPS*).*