An Introduction to Deep Reinforcement Learning
Other titles in Foundations and Trends® in Machine Learning

*Non-convex Optimization for Machine Learning*
Prateek Jain and Purushottam Ka

*Kernel Mean Embedding of Distributions: A Review and Beyond*
Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur and Bernhard Scholkopf
ISBN: 978-1-68083-288-4

*Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 1 Low-Rank Tensor Decompositions*
Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee, Ivan Oseledets, Masashi Sugiyama and Danilo P. Mandic
ISBN: 978-1-68083-222-8

*Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 2 Applications and Future Perspectives*
Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee, Ivan Oseledets, Masashi Sugiyama and Danilo P. Mandic
ISBN: 978-1-68083-276-1

*Patterns of Scalable Bayesian Inference*
Elaine Angelino, Matthew James Johnson and Ryan P. Adams
ISBN: 978-1-68083-218-1

*Generalized Low Rank Models*
Madeleine Udell, Corinne Horn, Reza Zadeh and Stephen Boyd
An Introduction to Deep Reinforcement Learning

Vincent François-Lavet
McGill University
vincent.francois-lavet@mcgill.ca

Peter Henderson
McGill University
peter.henderson@mail.mcgill.ca

Riashat Islam
McGill University
riashat.islam@mail.mcgill.ca

Marc G. Bellemare
Google Brain
bellemare@google.com

Joelle Pineau
Facebook, McGill University
jpineau@cs.mcgill.ca

Full text available at: http://dx.doi.org/10.1561/2200000071
Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2018, Volume 11, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.
# Contents

1 Introduction 
   1.1 Motivation ................................................. 2 
   1.2 Outline .................................................. 3 

2 Machine learning and deep learning 
   2.1 Supervised learning and the concepts of bias and overfitting 7 
   2.2 Unsupervised learning ....................................... 9 
   2.3 The deep learning approach ................................. 10 

3 Introduction to reinforcement learning 
   3.1 Formal framework .......................................... 16 
   3.2 Different components to learn a policy .................... 20 
   3.3 Different settings to learn a policy from data ............ 21 

4 Value-based methods for deep RL 
   4.1 Q-learning ............................................... 24 
   4.2 Fitted Q-learning ......................................... 25 
   4.3 Deep Q-networks ............................................ 27 
   4.4 Double DQN ................................................ 28 
   4.5 Dueling network architecture ............................... 29 
   4.6 Distributional DQN ......................................... 31 
   4.7 Multi-step learning ......................................... 32
4.8 Combination of all DQN improvements and variants of DQN 34

5 Policy gradient methods for deep RL 36
  5.1 Stochastic Policy Gradient ...................... 37
  5.2 Deterministic Policy Gradient .................. 39
  5.3 Actor-Critic Methods .......................... 40
  5.4 Natural Policy Gradients ....................... 42
  5.5 Trust Region Optimization ...................... 43
  5.6 Combining policy gradient and Q-learning .......... 44

6 Model-based methods for deep RL 46
  6.1 Pure model-based methods ...................... 46
  6.2 Integrating model-free and model-based methods ... 49

7 The concept of generalization 53
  7.1 Feature selection ............................... 58
  7.2 Choice of the learning algorithm and function approximator selection ....................... 59
  7.3 Modifying the objective function ................ 61
  7.4 Hierarchical learning ........................... 62
  7.5 How to obtain the best bias-overfitting tradeoff .. 63

8 Particular challenges in the online setting 66
  8.1 Exploration/Exploitation dilemma ............... 66
  8.2 Managing experience replay .................... 71

9 Benchmarking Deep RL 73
  9.1 Benchmark Environments ....................... 73
  9.2 Best practices to benchmark deep RL ............ 78
  9.3 Open-source software for Deep RL ............... 80

10 Deep reinforcement learning beyond MDPs 81
  10.1 Partial observability and the distribution of (related) MDPs 81
  10.2 Transfer learning ............................. 86
  10.3 Learning without explicit reward function ....... 89
  10.4 Multi-agent systems ........................... 91
11 Perspectives on deep reinforcement learning
11.1 Successes of deep reinforcement learning
11.2 Challenges of applying reinforcement learning to real-world problems
11.3 Relations between deep RL and neuroscience

12 Conclusion
12.1 Future development of deep RL
12.2 Applications and societal impact of deep RL

Appendices
A Appendix
A.1 Deep RL frameworks

References
An Introduction to Deep Reinforcement Learning

Vincent François-Lavet\textsuperscript{1}, Peter Henderson\textsuperscript{2}, Riashat Islam\textsuperscript{3}, Marc G. Bellemare\textsuperscript{4} and Joelle Pineau\textsuperscript{5}

\textsuperscript{1}McGill University; vincent.francois-lavet@mcgill.ca
\textsuperscript{2}McGill University; peter.henderson@mail.mcgill.ca
\textsuperscript{3}McGill University; riashat.islam@mail.mcgill.ca
\textsuperscript{4}Google Brain; bellemare@google.com
\textsuperscript{5}Facebook, McGill University; jpineau@cs.mcgill.ca

ABSTRACT

Deep reinforcement learning is the combination of reinforcement learning (RL) and deep learning. This field of research has been able to solve a wide range of complex decision-making tasks that were previously out of reach for a machine. Thus, deep RL opens up many new applications in domains such as healthcare, robotics, smart grids, finance, and many more. This manuscript provides an introduction to deep reinforcement learning models, algorithms and techniques. Particular focus is on the aspects related to generalization and how deep RL can be used for practical applications. We assume the reader is familiar with basic machine learning concepts.

1

Introduction

1.1 Motivation

A core topic in machine learning is that of sequential decision-making. This is the task of deciding, from experience, the sequence of actions to perform in an uncertain environment in order to achieve some goals. Sequential decision-making tasks cover a wide range of possible applications with the potential to impact many domains, such as robotics, healthcare, smart grids, finance, self-driving cars, and many more.

Inspired by behavioral psychology (see e.g., Sutton, 1984), reinforcement learning (RL) proposes a formal framework to this problem. The main idea is that an artificial agent may learn by interacting with its environment, similarly to a biological agent. Using the experience gathered, the artificial agent should be able to optimize some objectives given in the form of cumulative rewards. This approach applies in principle to any type of sequential decision-making problem relying on past experience. The environment may be stochastic, the agent may only observe partial information about the current state, the observations may be high-dimensional (e.g., frames and time series), the agent may freely gather experience in the environment or, on the contrary, the data
may be may be constrained (e.g., not access to an accurate simulator or limited data).

Over the past few years, RL has become increasingly popular due to its success in addressing challenging sequential decision-making problems. Several of these achievements are due to the combination of RL with deep learning techniques (LeCun et al., 2015; Schmidhuber, 2015; Goodfellow et al., 2016). This combination, called deep RL, is most useful in problems with high dimensional state-space. Previous RL approaches had a difficult design issue in the choice of features (Munos and Moore, 2002; Bellemare et al., 2013). However, deep RL has been successful in complicated tasks with lower prior knowledge thanks to its ability to learn different levels of abstractions from data. For instance, a deep RL agent can successfully learn from visual perceptual inputs made up of thousands of pixels (Mnih et al., 2015). This opens up the possibility to mimic some human problem solving capabilities, even in high-dimensional space — which, only a few years ago, was difficult to conceive.

Several notable works using deep RL in games have stood out for attaining super-human level in playing Atari games from the pixels (Mnih et al., 2015), mastering Go (Silver et al., 2016b) or beating the world’s top professionals at the game of Poker (Brown and Sandholm, 2017; Moravčík et al., 2017). Deep RL also has potential for real-world applications such as robotics (Levine et al., 2016; Gandhi et al., 2017; Pinto et al., 2017), self-driving cars (You et al., 2017), finance (Deng et al., 2017) and smart grids (François-Lavet, 2017), to name a few. Nonetheless, several challenges arise in applying deep RL algorithms. Among others, exploring the environment efficiently or being able to generalize a good behavior in a slightly different context are not straightforward. Thus, a large array of algorithms have been proposed for the deep RL framework, depending on a variety of settings of the sequential decision-making tasks.

### 1.2 Outline

The goal of this introduction to deep RL is to guide the reader towards effective use and understanding of core methods, as well as provide
references for further reading. After reading this introduction, the reader should be able to understand the key different deep RL approaches and algorithms and should be able to apply them. The reader should also have enough background to investigate the scientific literature further and pursue research on deep RL.

In Chapter 2, we introduce the field of machine learning and the deep learning approach. The goal is to provide the general technical context and explain briefly where deep learning is situated in the broader field of machine learning. We assume the reader is familiar with basic notions of supervised and unsupervised learning; however, we briefly review the essentials.

In Chapter 3, we provide the general RL framework along with the case of a Markov Decision Process (MDP). In that context, we examine the different methodologies that can be used to train a deep RL agent. On the one hand, learning a value function (Chapter 4) and/or a direct representation of the policy (Chapter 5) belong to the so-called model-free approaches. On the other hand, planning algorithms that can make use of a learned model of the environment belong to the so-called model-based approaches (Chapter 6).

We dedicate Chapter 7 to the notion of generalization in RL. Within either a model-based or a model-free approach, we discuss the importance of different elements: (i) feature selection, (ii) function approximator selection, (iii) modifying the objective function and (iv) hierarchical learning. In Chapter 8, we present the main challenges of using RL in the online setting. In particular, we discuss the exploration-exploitation dilemma and the use of a replay memory.

In Chapter 9, we provide an overview of different existing benchmarks for evaluation of RL algorithms. Furthermore, we present a set of best practices to ensure consistency and reproducibility of the results obtained on the different benchmarks.

In Chapter 10, we discuss more general settings than MDPs: (i) the Partially Observable Markov Decision Process (POMDP), (ii) the distribution of MDPs (instead of a given MDP) along with the notion of transfer learning, (iii) learning without explicit reward function and (iv) multi-agent systems. We provide descriptions of how deep RL can be used in these settings.
1.2. Outline

In Chapter 11, we present broader perspectives on deep RL. This includes a discussion on applications of deep RL in various domains, along with the successes achieved and remaining challenges (e.g. robotics, self driving cars, smart grids, healthcare, etc.). This also includes a brief discussion on the relationship between deep RL and neuroscience.

Finally, we provide a conclusion in Chapter 12 with an outlook on the future development of deep RL techniques, their future applications, as well as the societal impact of deep RL and artificial intelligence.


References


References


References


References


References


Walsh, T. 2017. It’s Alive!: Artificial Intelligence from the Logic Piano to Killer Robots. La Trobe University Press.


