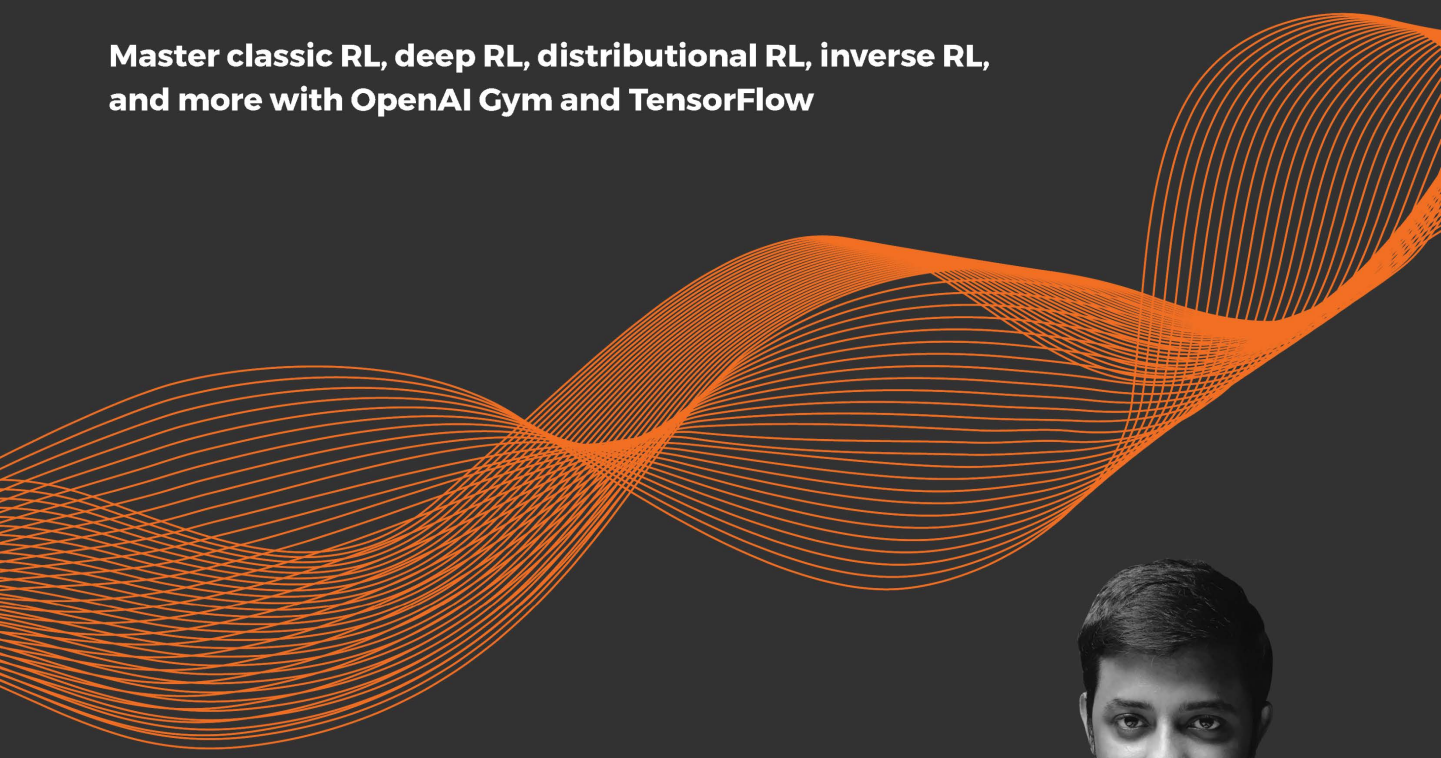
EXPERT INSIGHT

Deep Reinforcement Learning with Python

Master classic RL, deep RL, distributional RL, inverse RL, and more with OpenAI Gym and TensorFlow



Second Edition

Sudharsan Ravichandiran

Packt

Deep Reinforcement Learning with Python

Second Edition

Master classic RL, deep RL, distributional RL, inverse RL, and more with OpenAI Gym and TensorFlow

Sudharsan Ravichandiran

Packt

BIRMINGHAM - MUMBAI

Deep Reinforcement Learning with Python

Second Edition

Copyright © 2020 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing or its dealers and distributors, will be held liable for any damages caused or alleged to have been caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

Producers: Ben Renow-Clarke and Aarthi Kumaraswamy
Acquisition Editor – Peer Reviews: Divya Mudaliar
Content Development Editor: Bhavesh Amin
Technical Editor: Aniket Shetty
Project Editor: Janice Gonsalves
Copy Editor: Safis Editing
Proofreader: Safis Editing
Indexer: Pratik Shirodkar
Presentation Designer: Pranit Padwal

First published: June 2018

Second edition: September 2020

Production reference: 1300920

Published by Packt Publishing Ltd.

Livery Place

35 Livery Street

Birmingham B3 2PB, UK.

ISBN 978-1-83921-068-6

www.packt.com

To my adorable mom, Kasthuri, and to my beloved dad, Ravichandiran.

- Sudharsan Ravichandiran

packt.com

Subscribe to our online digital library for full access to over 7,000 books and videos, as well as industry leading tools to help you plan your personal development and advance your career. For more information, please visit our website.

Why subscribe?

- Spend less time learning and more time coding with practical eBooks and Videos from over 4,000 industry professionals
- Learn better with Skill Plans built especially for you
- Get a free eBook or video every month
- Fully searchable for easy access to vital information
- Copy and paste, print, and bookmark content

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.Packt.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at customercare@packtpub.com for more details.

At www.Packt.com, you can also read a collection of free technical articles, sign up for a range of free newsletters, and receive exclusive discounts and offers on Packt books and eBooks.

Contributors

About the author

Sudharsan Ravichandiran is a data scientist, researcher, best-selling author, and YouTuber (search for *Sudharsan reinforcement learning*). He completed his bachelor's in information technology at Anna University. His area of research focuses on practical implementations of deep learning and reinforcement learning, which includes natural language processing and computer vision. He is an open source contributor and loves answering questions on Stack Overflow. He also authored a best-seller, *Hands-On Reinforcement Learning with Python, 1st edition*, published by Packt Publishing.

I would like to thank my most amazing parents and my brother, Karthikeyan, for inspiring and motivating me. My huge thanks to the producer of the book, Aarthi, and the editors, Bhavesh, Aniket, and Janice. Special thanks to the reviewers, Sujit Pal and Valerii Babushkin, for providing their valuable insights and feedback. Without all their support, it would have been impossible to complete this book.

About the reviewers

Sujit Pal is a Technology Research Director at Elsevier Labs, an advanced technology group within the Reed-Elsevier Group of companies. His areas of interests include semantic search, natural language processing, machine learning, and deep learning. At Elsevier, he has worked on several initiatives involving search quality measurement and improvement, image classification and duplicate detection, and annotation and ontology development for medical and scientific corpora. He has co-authored a book on deep learning and writes about technology on his blog *Salmon Run*.

Valerii Babushkin is the senior director of data science at X5 Retail Group, where he leads a team of 100+ people in the area of natural language processing, machine learning, computer vision, data analysis, and A/B testing. Valerii is a Kaggle competitions Grand Master, ranking globally in the top 30. He studied cybernetics at Moscow Polytechnical University and mechatronics at Karlsruhe University of Applied Sciences and has worked with Packt as an author of the *Python Machine Learning Tips, Tricks, and Techniques* course and a technical reviewer for some books on reinforcement learning.

Table of Contents

Preface

With significant enhancement in the quality and quantity of algorithms in recent years, this second edition of *Hands-On Reinforcement Learning with Python* has been revamped into an example-rich guide to learning state-of-the-art **reinforcement learning** (**RL**) and deep RL algorithms with TensorFlow 2 and the OpenAI Gym toolkit.

In addition to exploring RL basics and foundational concepts such as the Bellman equation, Markov decision processes, and dynamic programming, this second edition dives deep into the full spectrum of value-based, policy-based, and actor-critic RL methods. It explores state-of-the-art algorithms such as DQN, TRPO, PPO and ACKTR, DDPG, TD3, and SAC in depth, demystifying the underlying math and demonstrating implementations through simple code examples.

The book has several new chapters dedicated to new RL techniques including distributional RL, imitation learning, inverse RL, and meta RL. You will learn to leverage Stable Baselines, an improvement of OpenAI's baseline library, to implement popular RL algorithms effortlessly. The book concludes with an overview of promising approaches such as meta-learning and imagination augmented agents in research.

Who this book is for

If you're a machine learning developer with little or no experience with neural networks interested in artificial intelligence and want to learn about reinforcement learning from scratch, this book is for you. Basic familiarity with linear algebra, calculus, and Python is required. Some experience with TensorFlow would be a plus.

What this book covers

Chapter 1, *Fundamentals of Reinforcement Learning*, helps you build a strong foundation on RL concepts. We will learn about the key elements of RL, the Markov decision process, and several important fundamental concepts such as action spaces, policies, episodes, the value function, and the Q function. At the end of the chapter, we will learn about some of the interesting applications of RL and we will also look into the key terms and terminologies frequently used in RL.

Chapter 2, *A Guide to the Gym Toolkit*, provides a complete guide to OpenAI's Gym toolkit. We will understand several interesting environments provided by Gym in detail by implementing them. We will begin our hands-on RL journey from this chapter by implementing several fundamental RL concepts using Gym.

Chapter 3, *The Bellman Equation and Dynamic Programming*, will help us understand the Bellman equation in detail with extensive math. Next, we will learn two interesting classic RL algorithms called the value and policy iteration methods, which we can use to find the optimal policy. We will also see how to implement value and policy iteration methods for solving the Frozen Lake problem.

Chapter 4, *Monte Carlo Methods*, explains the model-free method, Monte Carlo. We will learn what prediction and control tasks are, and then we will look into Monte Carlo prediction and Monte Carlo control methods in detail. Next, we will implement the Monte Carlo method to solve the blackjack game using the Gym toolkit.

Chapter 5, *Understanding Temporal Difference Learning*, deals with one of the most popular and widely used model-free methods called **Temporal Difference** (**TD**) learning. First, we will learn how the TD prediction method works in detail, and then we will explore the on-policy TD control method called SARSA and the off-policy TD control method called Q learning in detail. We will also implement TD control methods to solve the Frozen Lake problem using Gym.

Chapter 6, *Case Study – The MAB Problem*, explains one of the classic problems in RL called the **multi-armed bandit** (**MAB**) problem. We will start the chapter by understanding what the MAB problem is and then we will learn about several exploration strategies such as epsilon-greedy, softmax exploration, upper confidence bound, and Thompson sampling methods for solving the MAB problem in detail.

Chapter 7, *Deep Learning Foundations*, helps us to build a strong foundation on deep learning. We will start the chapter by understanding how artificial neural networks work. Then we will learn several interesting deep learning algorithms, such as recurrent neural networks, LSTM networks, convolutional neural networks, and generative adversarial networks.

Chapter 8, *A Primer on TensorFlow*, deals with one of the most popular deep learning libraries called TensorFlow. We will understand how to use TensorFlow by implementing a neural network to recognize handwritten digits. Next, we will learn to perform several math operations using TensorFlow. Later, we will learn about TensorFlow 2.0 and see how it differs from the previous TensorFlow versions.

Chapter 9, *Deep Q Network and Its Variants*, enables us to kick-start our deep RL journey. We will learn about one of the most popular deep RL algorithms called the **Deep Q Network** (**DQN**). We will understand how DQN works step by step along with the extensive math. We will also implement a DQN to play Atari games. Next, we will explore several interesting variants of DQN, called Double DQN, Dueling DQN, DQN with prioritized experience replay, and DRQN.

Chapter 10, *Policy Gradient Method*, covers policy gradient methods. We will understand how the policy gradient method works along with the detailed derivation. Next, we will learn several variance reduction methods such as policy gradient with reward-to-go and policy gradient with baseline. We will also understand how to train an agent for the Cart Pole balancing task using policy gradient.

Chapter 11, *Actor-Critic Methods – A2C and A3C*, deals with several interesting actor-critic methods such as advantage actor-critic and asynchronous advantage actor-critic. We will learn how these actor-critic methods work in detail, and then we will implement them for a mountain car climbing task using OpenAI Gym.

Chapter 12, *Learning DDPG, TD3, and SAC*, covers state-of-the-art deep RL algorithms such as deep deterministic policy gradient, twin delayed DDPG, and soft actor, along with step by step derivation. We will also learn how to implement the DDPG algorithm for performing the inverted pendulum swing-up task using Gym.

Chapter 13, *TRPO, PPO, and ACKTR Methods*, deals with several popular policy gradient methods such as TRPO and PPO. We will dive into the math behind TRPO and PPO step by step and understand how TRPO and PPO helps an agent find the optimal policy. Next, we will learn to implement PPO for performing the inverted pendulum swing-up task. At the end, we will learn about the actor-critic method called actor-critic using Kronecker-Factored trust region in detail.

Chapter 14, *Distributional Reinforcement Learning*, covers distributional RL algorithms. We will begin the chapter by understanding what distributional RL is. Then we will explore several interesting distributional RL algorithms such as categorical DQN, quantile regression DQN, and distributed distributional DDPG.

Chapter 15, *Imitation Learning and Inverse RL*, explains imitation and inverse RL algorithms. First, we will understand how supervised imitation learning, DAgger, and deep Q learning from demonstrations work in detail. Next, we will learn about maximum entropy inverse RL. At the end of the chapter, we will learn about generative adversarial imitation learning.

Chapter 16, *Deep Reinforcement Learning with Stable Baselines*, helps us to understand how to implement deep RL algorithms using a library called Stable Baselines. We will learn what Stable Baselines is and how to use it in detail by implementing several interesting Deep RL algorithms such as DQN, A2C, DDPG TRPO, and PPO.

Chapter 17, *Reinforcement Learning Frontiers*, covers several interesting avenues in RL, such as meta RL, hierarchical RL, and imagination augmented agents in detail.

To get the most out of this book

You need the following software for this book:

- Anaconda
- Python
- Any web browser

Download the example code files

You can download the example code files for this book from your account at `http://www.packtpub.com`. If you purchased this book elsewhere, you can visit `http://www.packtpub.com/support` and register to have the files emailed directly to you.

You can download the code files by following these steps:

1. Log in or register at `http://www.packtpub.com`.
2. Select the **SUPPORT** tab.
3. Click on **Code Downloads & Errata**.
4. Enter the name of the book in the **Search** box and follow the on-screen instructions.

Once the file is downloaded, please make sure that you unzip or extract the folder using the latest version of:

- WinRAR / 7-Zip for Windows

- Zipeg / iZip / UnRarX for Mac
- 7-Zip / PeaZip for Linux

The code bundle for the book is also hosted on GitHub at `https://github.com/PacktPublishing/Deep-Reinforcement-Learning-with-Python`. We also have other code bundles from our rich catalog of books and videos available at `https://github.com/PacktPublishing/`. Check them out!

Download the color images

We also provide a PDF file that has color images of the screenshots/diagrams used in this book. You can download it here: `https://static.packt-cdn.com/downloads/9781839210686_ColorImages.pdf`.

Conventions used

There are a number of text conventions used throughout this book.

`CodeInText`: Indicates code words in text, database table names, folder names, filenames, file extensions, pathnames, dummy URLs, user input, and Twitter handles. For example: "`epsilon_greedy` computes the optimal policy."

A block of code is set as follows:

```
def epsilon_greedy(epsilon):
    if np.random.uniform(0,1) < epsilon:
        return env.action_space.sample()
    else:
        return np.argmax(Q)
```

When we wish to draw your attention to a particular part of a code block, the relevant lines or items are highlighted:

```
def epsilon_greedy(epsilon):
    if np.random.uniform(0,1) < epsilon:
        return env.action_space.sample()
    else:
        return np.argmax(Q)
```

Any command-line input or output is written as follows:

```
source activate universe
```

Bold: Indicates a new term, an important word, or words that you see on the screen, for example, in menus or dialog boxes, also appear in the text like this. For example: "The **Markov Reward Process** (**MRP**) is an extension of the Markov chain with the reward function."

Warnings or important notes appear like this.

Tips and tricks appear like this.

Get in touch

Feedback from our readers is always welcome.

General feedback: Email `feedback@packtpub.com`, and mention the book's title in the subject of your message. If you have questions about any aspect of this book, please email us at `questions@packtpub.com`.

Errata: Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you have found a mistake in this book we would be grateful if you would report this to us. Please visit, `http://www.packtpub.com/submit-errata`, selecting your book, clicking on the Errata Submission Form link, and entering the details.

Piracy: If you come across any illegal copies of our works in any form on the Internet, we would be grateful if you would provide us with the location address or website name. Please contact us at `copyright@packtpub.com` with a link to the material.

If you are interested in becoming an author: If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, please visit `http://authors.packtpub.com`.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions, we at Packt can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about Packt, please visit `packtpub.com`.