Maxim Lapan　著

深度强化学习实践

（影印版）

Deep Reinforcement Learning Hands-On

东南大学出版社
SOUTHEAST UNIVERSITY PRESS

Packt

深度强化学习实践(影印版)

Deep Reinforcement Learning Hands-On

Maxim Lapan 著

南京　东南大学出版社

图书在版编目(CIP)数据

深度强化学习实践:英文/(俄罗斯)马克西姆·拉潘(Maxim Lapan)著. —影印本. —南京:东南大学出版社,2019.5

书名原文:Deep Reinforcement Learning Hands-On
ISBN 978-7-5641-8321-9

Ⅰ.①深… Ⅱ.①马… Ⅲ.①机器学习-算法-英文
Ⅳ.①TP181

中国版本图书馆 CIP 数据核字(2019)第 046195 号
图字:10-2018-496 号

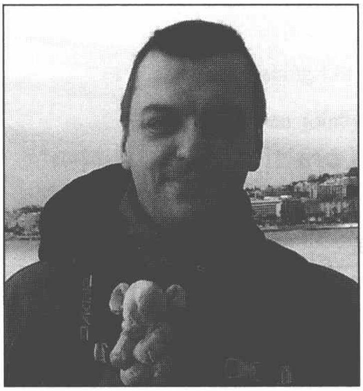
深度强化学习实践(影印版)

出版发行:东南大学出版社
地　　址:南京四牌楼 2 号　　邮编:210096
出 版 人:江建中
网　　址:http://www.seupress.com
电子邮件:press@seupress.com
印　　刷:常州市武进第三印刷有限公司
开　　本:787 毫米×980 毫米　16 开本
印　　张:34.25
字　　数:670 千字
版　　次:2019 年 5 月第 1 版
印　　次:2019 年 5 月第 1 次印刷
书　　号:ISBN 978-7-5641-8321-9
定　　价:109.00 元

本社图书若有印装质量问题,请直接与营销部联系。电话(传真):025-83791830

Contributors

About the author



Maxim Lapan is a deep learning enthusiast and independent researcher. His background and 15 years' work expertise as a software developer and a systems architect lays from low-level Linux kernel driver development to performance optimization and design of distributed applications working on thousands of servers. With vast work experiences in big data, Machine Learning, and large parallel distributed HPC and nonHPC systems, he has a talent to explain a gist of complicated things in simple words and vivid examples. His current areas of interest lie in practical applications of Deep Learning, such as Deep Natural Language Processing and Deep Reinforcement Learning.

Maxim lives in Moscow, Russian Federation, with his family, and he works for an Israeli start-up as a Senior NLP developer.

I'd like to thank my family: wife Olga and kids Ksenia, Julia, and Fedor, for patience and support. It was a challenging time, writing this book and it wouldn't be possible without you, thanks! Julia and Fedor did a great job gathering samples for MiniWoB (Chapter 13, *Web Navigation*) and testing ConnectFour agent's playing skills (Chapter 18, *AlphaGo Zero*).

I also want to thank the technical reviewers, Oleg Vasilev and Mikhail Yurushkin, for their valuable comments and suggestions about the book contents.

About the reviewers

Basem O. F. Alijla received his Ph.D. degree in intelligent systems from USM, Malaysia, in 2015. He is currently an assistant professor with Software Development Department, IUG in Palestine. He has authored number of technical papers published in journals and international conferences. His current research interest include, Optimization, Machine Learning, and Data mining.

Oleg Vasilev is a professional with a background in Computer Science and Data Engineering. His university program is Applied Mathematics and Informatics in NRU HSE, Moscow, with a major in Distributed Systems. He is a staff member on a Git-course, Practical_RL and Practical_DL, taught on-campus in HSE and YSDA. Oleg's previous work experience includes working in Dialog Systems Group, Yandex, as Data Scientist. He currently holds a position of Vice President of Infrastructure Management in GoTo Lab, an educational corporation, and he works for Digital Contact as a software engineer.

> I'd like to thank Alexander Panin (`@justheuristic`), my mentor, for opening the world of Machine Learning to me. I am deeply grateful to other Russian researchers who helped me in mastering Computer Science: *Pavel Shvechikov, Alexander Grishin, Valery Kharitonov, Alexander Fritzler, Pavel Ostyakov, Michail Konobeev, Dmitrii Vetrov*, and *Alena Ilyna*. I also want to thank my friends and family for their kind support.

Mikhail Yurushkin holds a PhD in Applied Mathematics. His areas of research are high performance computing and optimizing compilers development. He was involved in the development of a state-of-the-art optimizing parallelizing compiler system. Mikhail is a senior lecturer at SFEDU university, Rostov on Don, Russia. He teaches advanced DL courses, namely Computer Vision and NLP. Mikhail has worked for over 7 years in cross-platform native C++ development, machine learning, and deep learning. Now he works as an individual consultant in ML/DL fields.

Packt is Searching for Authors Like You

If you're interested in becoming an author for Packt, please visit `authors.packtpub.com` and apply today. We have worked with thousands of developers and tech professionals, just like you, to help them share their insight with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Preface

The topic of this book is Reinforcement Learning—which is a subfield of Machine Learning—focusing on the general and challenging problem of learning optimal behavior in complex environment. The learning process is driven only by reward value and observations obtained from the environment. This model is very general and can be applied to many practical situations from playing games to optimizing complex manufacture processes.

Due to flexibility and generality, the field of Reinforcement Learning is developing very quickly and attracts lots of attention both from researchers trying to improve existing or create new methods, as well as from practitioners interested in solving their problems in the most efficient way.

This book was written as an attempt to fill the obvious lack of practical and structured information about Reinforcement Learning methods and approaches. On one hand, there are lots of research activity all around the world, new research papers are being published almost every day, and a large portion of Deep Learning conferences such as NIPS or ICLR is dedicated to RL methods. There are several large research groups focusing on RL methods application in Robotics, Medicine, multi-agent systems, and others. The information about the recent research is widely available, but is too specialized and abstract to be understandable without serious efforts. Even worse is the situation with the practical aspect of RL application, as it is not always obvious how to make a step from the abstract method described in the mathematical-heavy form in a research paper to a working implementation solving actual problem. This makes it hard for somebody interested in the field to get an intuitive understanding of methods and ideas behind papers and conference talks. There are some very good blog posts about various RL aspects illustrated with working examples, but the limited format of a blog post allows the author to describe only one or two methods without building a complete structured picture and showing how different methods are related to each other. This book is my attempt to address this issue.

Another aspect of the book is its orientation to practice. Every method is implemented for various environments, from very trivial to quite complex. I've tried to make examples clean and easy to understand, which was made possible by the expressiveness and power of PyTorch. On the other hand, complexity and requirements of examples are oriented to RL hobbyists without access to very large computational resources, such as clusters of GPUs or very powerful workstations. This, I believe, will make the fun-filled and exciting RL domain accessible for a much wider audience than just research groups or large AI companies. However, it is still **Deep** Reinforcement Learning, so, having access to a GPU is highly recommended. Approximately, half of the examples in the book will benefit from running them on GPU. In addition to traditional medium-sized examples of environments used in RL, such as Atari games or continuous control problems, the book contains three chapters (8, 12, and 13) that contain larger projects, illustrating how RL methods could be applied to more complicated environments and tasks. These examples are still not full-sized real-life projects (otherwise they'll occupy a separate book on their own), but just larger problems illustrating how the RL paradigm can be applied to domains beyond well-established benchmarks.

Another thing to note about examples in the first three parts of the book is that I've tried to make examples self-contained and the source code was shown in full. Sometimes this led to repetition of code pieces (for example, training loop is very similar in most of the methods), but I believe that giving you the freedom to jump directly into the method you want to learn is more important than avoiding few repetitions. All examples in the book is available on Github: `https://github.com/PacktPublishing/Deep-Reinforcement-Learning-Hands-On`, and you're welcome to fork them, experiment, and contribute.

Who this book is for

The main target audience are people who have some knowledge in Machine Learning, but interested to get a practical understanding of the Reinforcement Learning domain. A reader should be familiar with Python and the basics of deep learning and machine learning. Understanding of statistics and probability will be a plus, but is not absolutely essential for understanding most of the book's material.

What this book covers

Chapter 1, What is Reinforcement Learning?, contains introduction to RL ideas and main formal models.

Chapter 2, OpenAI Gym, introduces the reader to the practical aspect of RL, using open-source library gym.

Chapter 3, Deep Learning with PyTorch, gives a quick overview of the PyTorch library.

Chapter 4, The Cross-Entropy Method, introduces you to one of the simplest methods of RL to give you the feeling of RL methods and problems.

Chapter 5, Tabular Learning and the Bellman Equation, gives an introduction to the Value-based family of RL methods.

Chapter 6, Deep Q-Networks, describes DQN, the extension of basic Value-based methods, allowing to solve complicated environment.

Chapter 7, DQN Extensions, gives a detailed overview of modern extension to the DQN method, to improve its stability and convergence in complex environments.

Chapter 8, Stocks Trading Using RL, is the first practical project, applying the DQN method to stock trading.

Chapter 9, Policy Gradients – An Alternative, introduces another family of RL methods, based on policy learning.

Chapter 10, The Actor-Critic Method, describes one of the most widely used method in RL.

Chapter 11, Asynchronous Advantage Actor-Critic, extends Actor-Critic with parallel environment communication, to improve stability and convergence.

Chapter 12, Chatbots Training with RL, is the second project, showing how to apply RL methods to NLP problems.

Chapter 13, Web Navigation, is another long project, applying RL to web page navigation, using MiniWoB set of tasks.

Chapter 14, Continuous Action Space, describes the specifics of environments, using continuous action spaces and various methods.

Chapter 15, Trust Regions – TRPO, PPO, and ACKTR, is yet another chapter about continuous action spaces describing "Trust region" set of methods.

Chapter 16, Black-Box Optimization in RL, shows another set of methods that don't use gradients in explicit form.

Chapter 17, Beyond Model-Free – Imagination, introduces model-based approach to RL, using recent research results about imagination in RL.

Chapter 18, AlphaGo Zero, describes the AlphaGo Zero method applied to game Connect Four.

To get the most out of this book

All chapters in the book describing RL methods have the same structure: in the beginning we discuss the motivation of the method, its theoretical foundation, and intuition behind it. Then, we follow several examples of the method applied to different environment with full source code. So, you can use the book in different ways:

1. To quickly become familiar with some method of methods you can read only introductory part of the relevant chapter or chapter's section.
2. To get deeper understanding of the way method is implemented you can read the code and the comments around.
3. To gain deep familiarity with the method (the best way to learn, I believe) you should try to reimplement the method and make it working, using provided source code as a reference point.

In any case, I hope the book will be useful for you!

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-Hands-On`. 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://www.packtpub.com/sites/default/files/downloads/DeepReinforcementLearningHandsOn_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; "The method `get_observation()` is supposed to return to the agent the current environment's observation."

A block of code is set as follows:

```
def get_actions(self):
    return [0, 1]
```

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

```
def get_actions(self):
    return [0, 1]
```

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

```
$ xvfb-run -s "-screen 0 640x480x24" python 04_cartpole_random_monitor.py
```

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: "In practice it's some piece of code, which implements some **policy**."

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`.

Table of Contents