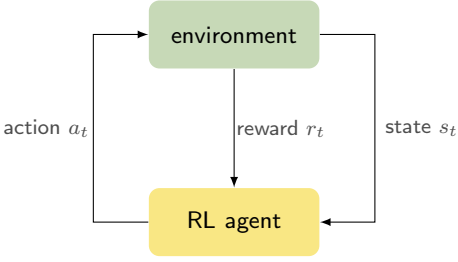
facebook
Artificial Intelligence Research

Exploration-Exploitation in Reinforcement Learning

Introduction

Mohammad Ghavamzadeh, Alessandro Lazaric and Matteo Pirotta
Facebook AI Research

Reinforcement Learning



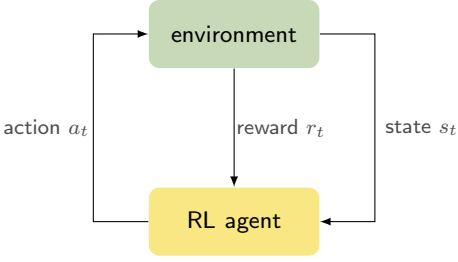
"**Reinforcement learning** is learning how to map states to actions so as to **maximize** a numerical **reward** signal in an unknown and **uncertain** environment.

In the most interesting and challenging cases, **actions** affect not only the immediate reward but also the **next situation** and all subsequent rewards (**delayed reward**).

The agent is not told which actions to take but it must discover which actions yield the most reward by trying them (**trial-and-error**)."

— Sutton and Barto [1998]

Reinforcement Learning



"**Reinforcement learning** is learning how to map states to actions so as to **maximize** a numerical **reward** signal in an unknown and **uncertain** environment.

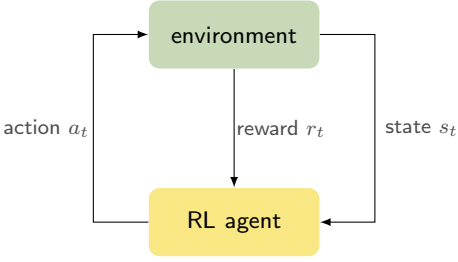
In the most interesting and challenging cases, **actions** affect not only the immediate reward but also the **next situation** and all subsequent rewards (**delayed reward**).

Exploration

The agent is not told which actions to take but it must discover which actions yield the most reward by trying them (**trial-and-error**)."

— Sutton and Barto [1998]

Reinforcement Learning



Exploitation

"**Reinforcement learning** is learning how to map states to actions so as to **maximize** a numerical **reward** signal in an unknown and **uncertain** environment.

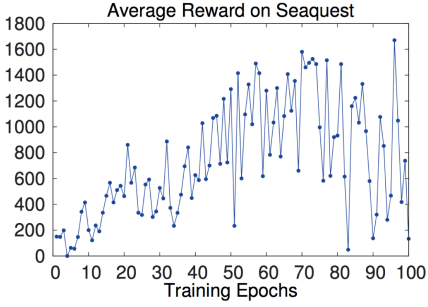
In the most interesting and challenging cases, **actions** affect not only the immediate reward but also the **next situation** and all subsequent rewards (**delayed reward**).

Exploration

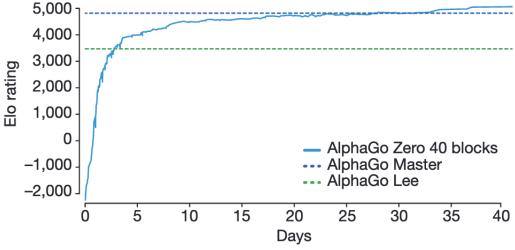
The agent is not told which actions to take but it must discover which actions yield the most reward by trying them (**trial-and-error**)."

— Sutton and Barto [1998]

Why This Tutorial?



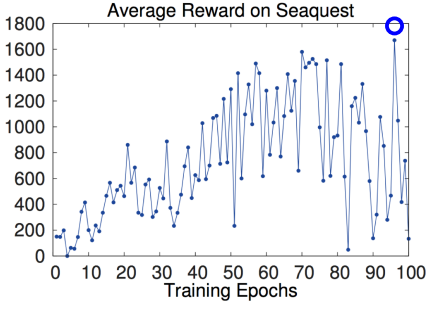
Mnih et al. [2015]



Silver et al. [2016]

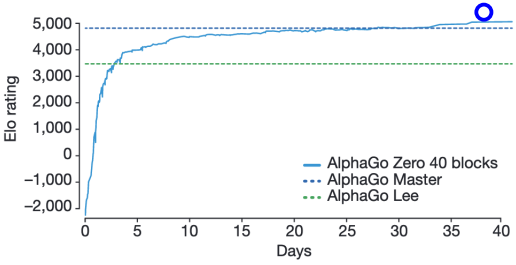
Why This Tutorial?

Superhuman performance



Mnih et al. [2015]

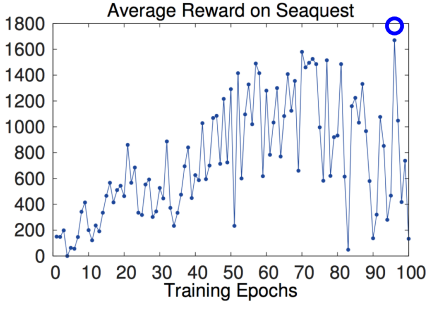
Beating world champion



Silver et al. [2016]

Why This Tutorial?

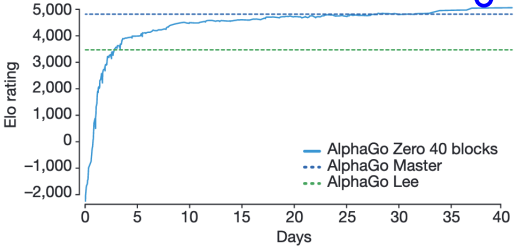
Superhuman performance



Mnih et al. [2015]

10 million frames

Beating world champion

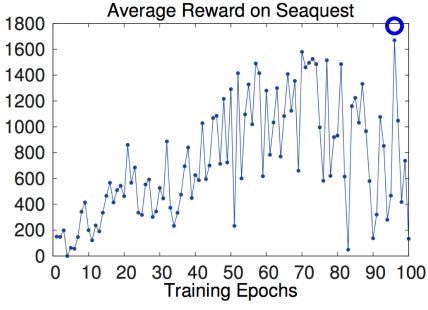


Silver et al. [2016]

4.9 million games

Why This Tutorial?

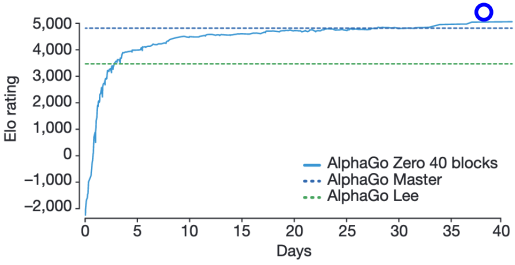
Superhuman performance



Mnih et al. [2015]

10 million frames

Beating world champion



Silver et al. [2016]

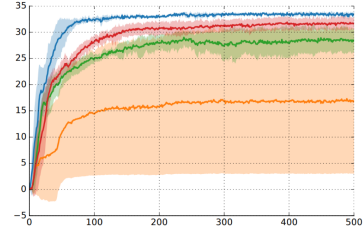
4.9 million games

Even best RL algorithms are very **sample inefficient**

Why This Tutorial?

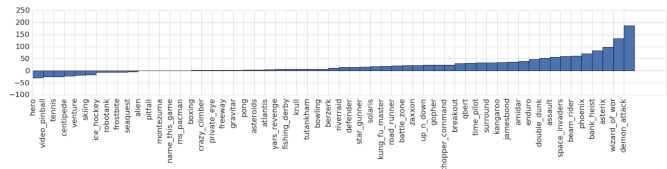
Better exploration may significantly **improve the sample efficiency**

**Optimism in face of uncertainty*



Tang et al. [2017]

**Thompson sampling*



Fortunato et al. [2018]

**inspired by*

Objective of the Tutorial

- Formalize the exploration-exploitation dilemma
- Review exploration design principles (optimism and randomness) and illustrate their theoretical guarantees
- Review how design principles can be scaled up into DeepRL

Organization

- Part 1. The Exploration-Exploitation Dilemma in Finite-Horizon MDPs
  ET (8:45am - 9:00am)
- Part 2. Regret Minimization Algorithms in Tabular MDPs
  ET (9:00am - 10:00am)
- Part 3. Effective and Scalable Exploration in DeepRL
  ET (10:00am - 11:45am with coffee)
- Part 4. Regret Minimization Algorithms in Continuous MDPs
  ET (11:45am - 12:20pm)

Website

`https://rlgammazero.github.io`

Meire Fortunato, Mohammad Gheshlaghi Azar, Bilal Piot, Jacob Menick, Matteo Hessel, Ian Osband, Alex Graves, Volodymyr Mnih, Rémi Munos, Demis Hassabis, Olivier Pietquin, Charles Blundell, and Shane Legg. Noisy networks for exploration. In *ICLR (Poster)*. OpenReview.net, 2018.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

Haoran Tang, Rein Houthooft, Davis Foote, Adam Stooke, Xi Chen, Yan Duan, John Schulman, Filip De Turck, and Pieter Abbeel. #exploration: A study of count-based exploration for deep reinforcement learning. In *NIPS*, pages 2753–2762, 2017.