A Brief Survey of Reinforcement Learning

A modern beamer theme

Giancarlo Frison

November 26, 2018

Table of contents

Introduction

What is Reinforcement Learning

[1]

Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

What it is Not

Although RL can induce to an optimization, there are major differences within:

- Supervised learning

What it is Not

Although RL can induce to an optimization, there are major differences within:

- Supervised learning
- Mathematical optimization

What it is Not

Although RL can induce to an optimization, there are major differences within:

- Supervised learning
- Mathematical optimization
- Genetic programming

What it is Not

Although RL can induce to an optimization, there are major differences within:

- Supervised learning
- Mathematical optimization
- Genetic programming

Metropolis

The **METROPOLIS** theme is a Beamer theme with minimal visual noise inspired by the HSRM Beamer Theme by Benjamin Weiss.

Enable the theme by loading

```
\documentclass{beamer}
\usetheme{metropolis}
```

Note, that you have to have Mozilla's *Fira Sans* font and XeTeX installed to enjoy this wonderful typography.

Sections

Sections group slides of the same topic

```
\section{Elements}
```

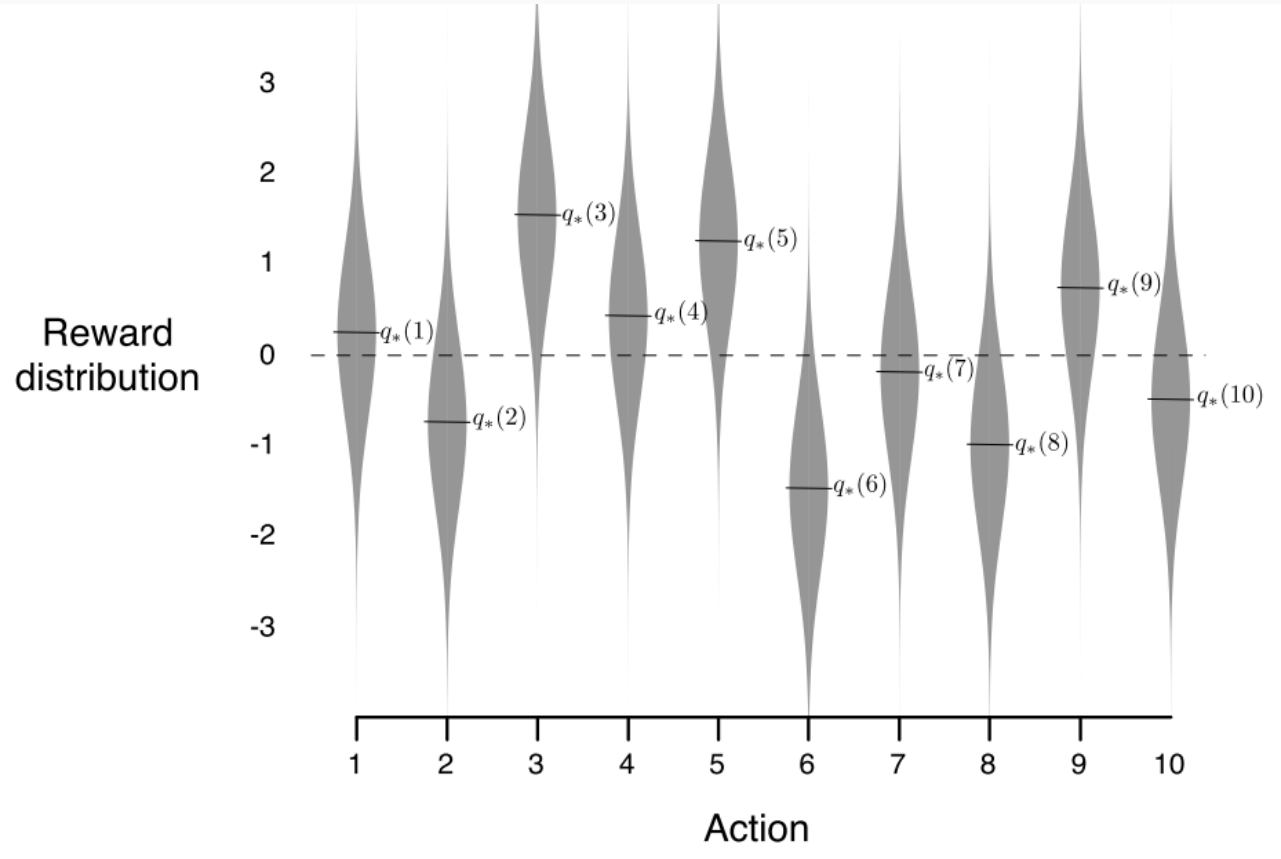
for which METROPOLIS provides a nice progress indicator ...

Multi-armed Bandit

Multi-armed Bandit

The multi-armed bandit problem has been the subject of decades of intense study in statistics, operations research, electrical engineering, computer science, and economics [**?**]

Q Value's Action



An example bandit problem [**?**]. Obtained measures after repeated pullings with 10 arms.

Q Value's Action

$\mathbf{Q_n}$ is the estimated value of its action after n selections.

$$Q_{n+1} = \frac{R_1 + R_2 + ... + R_n}{n}$$

A more scalable formula, updates the average with incremental and small constant:

$$Q_{n+1} = Q_n + \frac{1}{n}(R_n - Q_n)$$

General expression of the badint algorithm at the fundation of RL. *Target* could be considered the reward R by now.

$$\textsf{NewEstimate} = \textsf{OldEstimate} + \textsf{StepSize}(\textsf{Target} - \textsf{OldEstimate})$$

Gambler's Dilemma

When pulled, an arm produces a random payout drawn independently of the past. Because the distribution of payouts corresponding to each arm is not listed, the player can learn it only by experimenting.

Exploitation

Earn more money by exploiting arms that yielded high payouts in the past.

Exploration

Exploring alternative arms may return higher payouts in the future.

Markov Decision Process

Definition of MDP

Figure 1: 2015 DARPA Robotics Challange [**?**]

Despite as in bandits, MDP formalizes the decision making (Policy π) in sequential steps, aggregated in Episodes.

Actions and States

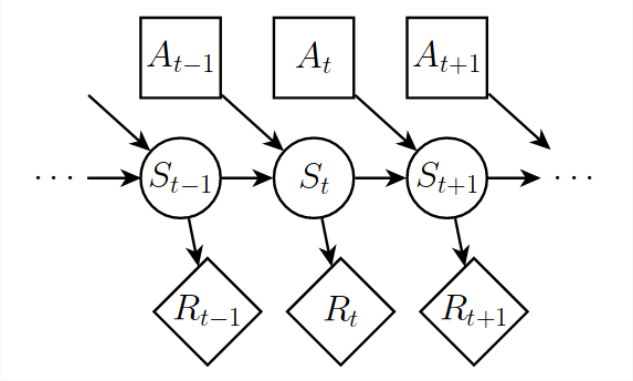


Figure 2: Model representation of MDP [**?**]

MDP strives to find the best π to all possible states. In Markov processes, the selected action depends **only** on current state.

How to evaluate an Agent?

Given that:

- Agent maximizes future rewards.
- γ is the rewards' discount factor.
- Policy π defines a particular way of acting.
- $v_\pi(s)$ is the expected return from s following π *thereafter*.
- $q_\pi(s, a)$ is the $v_\pi(s)$ taking action a

A recursive algorithm could be indentified, known as the Bellman equation. Iteratevely computes the value Q from the terminal state:

$$Q_t(s, a) = R_{t+1} + \gamma max[v(S_{t+1})]$$

Grid World

🏁	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	🏁

- $A = up, down, right, left$
- Terminal states are the flagged boxes.
- $R_t = 0$ for terminal states.
- $R_t = -1$ for other states.

The problem is to define the best π. Value function is computed by iterative policy evaluation.

Grid World

Iteration

k = 1

Calculated V_k

0	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	0

Policy π_k

🏁	←	?	?
↑	?	?	?
?	?	?	↓
?	?	→	🏁

Titleformats

Metropolis titleformats

METROPOLIS supports 4 different titleformats:

- Regular
- SMALLCAPS
- ALLSMALLCAPS
- ALLCAPS

They can either be set at once for every title type or individually.

Small caps

This frame uses the `smallcaps` titleformat.

Potential Problems

Be aware, that not every font supports small caps. If for example you typeset your presentation with pdfTeX and the Computer Modern Sans Serif font, every text in smallcaps will be typeset with the Computer Modern Serif font instead.

References i


Omanomanoman.
CC BY-SA 4.0.
CC BY-SA 4.0, https://commons.wikimedia.org/.

List of Figures i