Summer project in reinforcement learning

This is the rough outline of a 6-8 week extended project on reinforcement learning. The goal of the project is to get a student up to speed with the basics of reinforcement learning and the ability to use standard libraries to solve simple problems.

The project is roughly divided into two components: (i) a dive into the theory and main applications of reinforcement learning, and (ii) a replication of a published study.

Topics covered

- ★ Basic overview of MDPs; value and policy iteration ★
- Monte Carlo prediction and control; temporal-difference learning; tabular methods
- Value-based methods: on-policy prediction and control, off-policy prediction and control
- ★ Policy-based methods: policy gradient algorithms, REINFORCE, actor-critic ★

Project

1. Select a recent (2019 or later) paper from NIPS, ICML, or a similar journal. The paper could (i) propose a new algorithm, (ii) propose new approximation architecture, or (iii) propose a new domain application.
2. Using the OpenAI Gym, SLM Lab, and other free tools, reproduce the paper's result (or try it on a simpler problem).
3. Explore possible generalizations, new applications, improvements, etc.

Reading materials

Books

Sutton, Richard S., and Andrew G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018 (price comparison)

Graesser, Laura, and Wah Loon Keng. *Foundations of Deep Reinforcement Learning: Theory and Practice in Python*. Addison-Wesley Professional, 2019. (price comparison)

Libraries

- PyTorch
- OpenAI Gym
- SLM Lab

Schedule (by weeks)

1. Basic overview of MDPs; value and policy iteration

   a. Sutton and Barto — Chapters 2-4
   b. Graesser and Keng — Chapter 1

2. Monte Carlo prediction and control; temporal difference learning; tabular methods

   a. Sutton and Barto — Chapters 5-6

3. Value-based methods: on-policy prediction and control, off-policy prediction and control

   a. Sutton and Barto — Chapters 9-10 (Chapter 11 optional)
   b. Graesser and Keng — Chapters 3-5

4. Policy-based methods: policy gradient algorithms, REINFORCE, actor-critic

   a. Sutton and Barto — Chapter 13
   b. Graesser and Keng — Chapters 2, 6

5. Project (Reading phase)

   a. Sutton and Barto — Chapter 14
   b. Graesser and Keng — Chapters 14-17

6. Project (Implementation phase)

   a. Sutton and Barto — Chapter 15
   b. Graesser and Keng — Chapters 10-13

7. Project (Implementation phase)

   a. Sutton and Barto — Chapter 16
   b. Graesser and Keng — Chapters 7-9

8. Project (Publish—to GitHub—phase)

   a. Sutton and Barto — Chapter 17