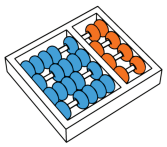
MC934B/MO436A - Reinforcement Learning
Institute of Computing - Unicamp
Second Semester 2021



Prof. Esther Colombini
esther@ic.unicamp.br

Project 2 - Deadline: 12/12/2021

1 Goal

The goal of this assignment is to apply deep reinforcement learning control methods to a problem modeled and specified by the group. You must clearly define:

- The problem addressed
- The MDP formulation
- The characteristics of the selected algorithms (policy-based, value-based, actor-critic)

The work consists of finding an adequate solution to the chosen problem, evaluating it according to: computational cost, optimality, influence of reward function, state and action space representations. You are required to clearly define:

- How the problem was modeled
- Implementation specifics and restrictions

2 Problem

Write an environment that implements your problem. You can adapt environments that are available in the literature, but, in any case, you should fully characterize your environment by defining:

- The nature of your environment (episodic/not episodic, deterministic/stochastic)
- What are your terminal states (when they exist)
- How is your reward function defined
- All parameters employed in your methods (discount factor, learning rate, etc.)

Your group is required to implement the following methods, both for a deterministic and a stochastic environment:

1. **An off-policy method with a neural network as function approximator:** Your team can use DQN, Double-DQN, HER, DDPG, SAC or any of its variations with actor-critic formulations.
2. **An on-policy method with a neural network as function approximator:** Your team can use REINFORCE and its variations, A2C, TRPO, PPO or any variations with actor-critic formulations.

You have to clearly state your action space (discrete or continuous) and how the chosen methods work with the nature of the action-space representation.

The system must be evaluated according to the quality of the solutions found and a critical evaluation is expected on the relationship between adopted parameters x solution performance. Graphs and tables representing the evolution of the solutions are mandatory. Additional comparisons with the literature are welcome, although they are not mandatory. If your team decided to maitain the same theme as project 1, compare the results with the prior algorithms implemented.

3 Programming language

You should use Pyhton as programming language. However, interfacing with other languages and libraries is permitted once provided the reasons for such adoption.

4 Evaluation and Discussion

The system should be evaluated according to the quality of the solutions found, and a critical evaluation is expected on the relationship between adopted parameters x solution quality. Graphs, tables, and images representing the results are mandatory. Further comparisons with the literature are welcome, although not mandatory. A link to a video of up to 5 minutes should be indicated in the report. This video will be used as the presentation of the work. Prepare it like that. You can find good examples of videos in the links below.

- `https://www.youtube.com/watch?v=DryhqYb1YjE`
- `https://www.youtube.com/watch?v=z99CI4YXTZ4`
- `https://www.youtube.com/watch?v=0P0Xeb_JzrI`

Please, discuss in the report:

- The advantages and disadvantages of each method adopted to your problem. How did the on-policy training compare to the off-policy formulation regarding stability, number of samples to convergence, etc.?
- How the reward function influenced the quality of the solution. Was your group able to achieve the expected policy given the reward function defined?
- How non-linear function approximation influenced the results. What were the advantages and disadvantages of using it in your problem? This question should be answered by those groups that kept the same themes as project 1.

5 Groups

The groups must be composed of 4 members.

6 Report

The definition of the problem, the solution, and the results obtained must be presented in a report created as a Jupyter notebook. Please, make sure you put the graphs, tables, comparisons, and critical analysis in the notebook. The report should clearly indicate what the contribution of each team member was.