Coursework Guide

Version History

0.1	02/10/23	First version.
0.2	12/10/23	Changed the paper range to 2022-2023. Part affected: bullet point starting "Select 1-3 good papers from …".

Summary

Title: Reinforcement Learning using Gymnasium environments
Hand-in: Programs AND a written report will need to be submitted online via Moodle. Check the module's Moodle page for the precise deadline.
Late policy: The coursework deadlines are absolute. Late submissions are subject to a 5% deduction of the overall coursework mark per day.

Informal Description

The coursework consists of two tasks as described below. Your aim is to build several reinforcement learning agents and to design, implement and run several basic research-based experiments. You will hand-in software and a report that discusses your work on these tasks. Briefly, task 1 is about implementing some basic RL prototypes for target environments and identification of key literature, gaps, and research questions, whereas task 2 is about designing, developing and running experiments based on the research questions identified in task 1. Please consult the table below for details.

Aims and Outcomes

- If you take the labs seriously, at the end of the semester you should be:
    - comfortable with implementing and modifying reinforcement learning agents,
    - capable of adapting your RL solutions to different kinds of robotic problems with well-defined states, actions and rewards,
    - comfortable with neural network approaches for the mapping of complex high-dimensional states to actions (if you choose to use neural network based RL solutions),
    - capable of scanning the literature in order to understand modern RL techniques, and incorporating/extending these in your own solutions,
    - capable of identifying gaps, and/or weaknesses/limitations in state-of-the-art research, and using this to define research questions for guiding your research,
    - capable of studying and evaluating algorithm performance objectively,
    - capable of designing innovative algorithms and experiments, and reporting the results of these in a clear and well-structured manner.

Rough Timetable

Week	Main Lab	Main activities
1	03/10/23	Getting started. Familiarization with Gym.
2	10/10/23	Task 1
3	17/10/23	Task 1
4	24/10/23	Task 1
5	31/10/23	Task 1. Demos for task 1.
6	07/11/23	Task 2
7	14/11/23	Task 2.
8	21/11/23	Task 2.
9	28/11/23	Task 2
10	05/12/23	Task 2. Demos for task 2.

Laboratory notes

- You will work individually.
- We need to start working hard from the very first day to make the most of the lab sessions. In the first week you will learn the basics of Gymnasium, will experiment with several environments, and will even try some small heuristics on simple control problems (e.g. cartpole).

Getting Started

Preliminary steps

- Check the following three main Gymnasium resources:
  - Farama's general documentation page for Gymnasium.
  - Basic usage page in the above documentation.
  - Gymnasium GitHub page – includes installation instructions.
- Install Gymnasium.
- For the purpose of the coursework it is sufficient to work with the "classic control" set of environments, however do feel free to install and use other categories of environments (e.g. MuJoCo and Atari), if you wish.
- Go through the Basic Usage page.
- You can install Gym on your own machines, or in your local directory in UNM's HPC, or you can also use Google Colaboratory. Please note that in the past there were ways to render environments properly in Colab (e.g. have a look at this tutorial) however this may change from time to time. For an example of a Jupyter notebook for the cart pole example, refer to the module's Moodle page.
- If you still have time on in your first lab, experiment with different environments.
- As mentioned, if you want to use any of the MuJoCo environments you can. Deep Mind recently bought MuJoCo and made it open source, which means there are no more licensing issues. You are not required to use MuJoCo, but if you really want to, you are free to install it, and get the environments setup.
- To see what environments are available use:

```
import gymnasium as gym
print(gym.envs.registry.keys())
```

- To better understand some Gym environments consult this Wiki or scroll to "environments" in the Gymnasium's GitHub page, and search for your environment. For example for the cart pole environment have a look at this page.

Try to come up with some heuristic solutions for Cart Pole

- Try to come up with some simple heuristics to keep the pole up based on your understanding of the environment. You can start from and modify the (failing) heuristic example provided in the Moodle page (i.e. sol-H1-cart-pole-v0).
- Difficult? Let's see whether reinforcement learning helps.

Have a look at a Q-learning solution

- Example: s1cart-pole-v0-sol1.
- Try to run the code.
- Read the code. Try to understand it as much as possible, although note, it will only fully make sense once we have done Q-Learning in the lectures.

Task Description

- **Requirements for Task 1:**
  - **Title.** Prototypes, literature, gaps, and research questions.
  - **Prototypes:**
    - **Environment selection [compulsory].** You are free to select any **2 environments** from within the control category. Example of a suitable environment: CartPole-v1.
    - **Environment selection [optional].** You can select **zero, one, or two** other environment from any of the environmental categories, if you find this interesting or useful for your experiments. Please recall that different environments may impose significant changes to your reinforcement learning algorithm since, for example, they may involve continual action spaces, or other representational differences.
    - **Methods allowed**: only reinforcement learning.
    - **Aim:** for each environment develop at least one viable proof of concept based on RL.
  - **Literature:**
    - **Steps**:
      - Explore the recent RL literature (in general or in relation of a topic of interest (e.g. noise)).
      - Select 1-3 good papers from the date range 2022-2023 and highlight their gaps (i.e. limitations and/or open

questions/problems). Note that although these 1-3 papers will be your "core/seed" papers, you should still study the literature more broadly (i.e. your report should cite other papers apart from the core papers).

- Select your gaps for further investigation. Justify your choices.
- Design at least 2 research questions based on your selected gaps.

▪ **Aim**: clearly outline 1-3 selected papers, overall gaps, selected gaps, and research questions. Note that it is crucial for the papers, gaps and research questions to be 100% credible, i.e.: (1) the papers must be recent and good, (2) the gaps must be genuine open problems, and (3) the research questions must sit squarely in the gaps and must point in useful directions.

▪ **Important constraint**: Every student must have a different set of core papers and/or a different set of gaps and/or a different set of research questions (RQs). Once a student has defined their selected papers, gaps, and RQs, they must email them to me, in order for me to check and approve them. Please note that this process will operate on a "first come first served" basis. Please also note that if two students share the same papers, they can still be different in terms of the chosen gaps or RQs, however, it is preferable if all elements are distinct.

- **Requirements for Task 2:**
  - **Title.** Research questions and experiments.
  - **Environment selection.** It is recommended that you use the same environments you used in task 1.
  - **Methods allowed**: reinforcement learning.
  - **Goals**. Keywords: novel experiments and insights. The aim of this task is for you to design, develop, run, and analyze, experiments that address the research questions your listed in task 1. The mains tasks would be: (1) design experiments that address the research questions, (2) implement the experiments, (3) adjust the RL solutions according to the experiments, (4) run the experiments and collect results, (5) analyze the results and assess whether they answered the research questions, (6) either proceed back to step 1 with adjustments to the experiments/solutions, or proceed with additional experiments (depending on time and completion status). Document your findings.
- **Requirements for all tasks (i.e. tasks 1 and 2):**
  - **Performance**. Define one or more valid performance measures, apart from the default/compulsory one, i.e.: the average number of episodes needed before learning a problem (see below for more information).
  - **Evaluation**. Run your experiments and report your results for all of the chosen environments consistently.
  - **Four "Is"**. Try to make sure your work is: innovative (i.e. novel), inventive (i.e. not technically trivial), impactful (e.g. generate new knowledge), and informed (i.e. it is based on a solid understanding of the literature).

- **Demo.** Show and explain the performance of your solutions, and the results of your experiments.

Performance Evaluation

- Compared to previous years, this year will allow more flexibility regarding performance evaluation. You can use external leaderboards for fun (e.g.: https://github.com/openai/gym/wiki/Leaderboard), however a greater emphasis will be put on internal comparisons (i.e. your own experimental conditions) and innovation.
- One key performance measure that you should recall is the number of episodes required before solving the problem. In other words, here you are interested in the speed of learning. Care must be taken in being explicit and consistent regarding what constitutes having solved the problem.

Remarks on Hardware Integrations

- The hardware component (Lego EV3 robots) is completely optional, and only provides a small amount of bonus marks (i.e. 3 marks). As discussed in the introductory lecture, this is to preempt any potential issues with a future pandemic, and in order to maximize the opportunities to learn reinforcement learning via Gymnasium.

- If students opt to include a hardware component there will be no restrictions placed on the design of the robot, the task chosen, the sensors chosen, the firmware chosen, and the language chosen. Students are free to choose any of these aspects in a manner that facilitates the demonstration of what they have learnt via Gymnasium. Although this freedom brings advantages in terms of flexibility regarding how to implement the integration, it comes at a cost in terms of potential unsupported issues. If students decide to go for these optional bonus marks, they need to be prepared to face time sinks and limited robot access time.

- I will collect information on who might potentially want to go for this option, and then I will decide on how best to allocate robots, since the cohort number is larger than the number of robots available, and assignments are individual.

Assessment – Overall

Component	Marks (100)	Description	Main Criteria
Demo	10	One demo for each task.	Evidence of understanding of the base code. Evidence of innovation. Strong justifications and arguments. Clear communication. [Demo 1: 5 marks; Demo 2: 5 marks.]
Software	20	Multiple files organized with a clear structure.	Is the code complete? Is the code well-designed, clean, elegant, and well commented? Is the code complex/challenging enough?
Task 1 - report	20	1-2 page report summarizing task 1	Are the core papers (1-3) well explained? Are the overall gaps well identified and explained? Are the selected gaps justified properly? Are the research questions grounded in the gaps, and are they clear, concrete, and heading in the right direction?
Task 2 - paper	50	Mini-conference paper summarizing all of the work done on both tasks.	Are the structure, grammar and argumentation of the paper/report good? Are the introduction, background, methods, results and analyses, clear, comprehensive and insightful? Does the paper show critical and creative thinking?
Hardware	3	Bonus marks for hardware integration	As discussed before, integration of your algorithms with hardware is optional. Marks are bonus marks, and the total overall marks are naturally capped at 100.

Assessment Criteria for the Report (task 1) and Paper (task 2)

- 1st an excellent, well-written report/paper demonstrating extensive understanding and good insight.
- 2:1 a comprehensive, well-written report/paper demonstrating thorough understanding and some insight.
- 2:2 a competent report/paper demonstrating good understanding of the implementation.
- 3rd an adequate report/paper covering all specified topics at a basic level of understanding.
- F an inadequate report/paper failing to cover the specified topics.

Report guide (task 1)

- Note that the report for task 1 will be submitted for marking at the end, together with the paper for task 2, however, you can and should always seek feedback from me, earlier, during the labs.
- The report for task 1 has no fixed format, as long as it is well structured and well organized. The only constraint is that it should be 1-2 pages long.

- This report will exclusively focus on: (1) a very brief summary of your prototypes, (2) brief summaries of your selected core papers, and why they were chosen, (3) lengthier explanations on the weaknesses/gaps of the papers, (4) an explanation and justification of your selected gaps, and (5) an explanation and justification of your research questions, and how they are grounded in the gaps.

Paper Guide (task 2)

You should design your final report as a conference paper. The paper should contain:

- [8 marks] Introduction (about 1 page). An introduction containing an introduction to the main concepts, problem statement, overview of background, research questions, and your main contributions.

- [8 marks] Background (about 0.5 pages). A brief overview of the field and the key papers closely related to your work (this will include the core 1-3 papers and other relevant papers). The core selected papers with their gaps, and why there were chosen selected, must be clearly explained.

- [8 marks] Methods (about 1 page). A detailed and concise description of how you implemented task 2 (e.g. algorithms and experimental design).

- [10 marks] Results (about 1 page). An overview of your key results encompassing performance measures and other results leading to insights about the problem and/or your solutions.

- [10 marks] Discussion (about 0.5 pages). Your interpretation of the results, your conclusions, and proposed future work.

- [6 marks] References & Appendices (not included in the word count).

Note: Writing a concise report/paper is a core part of the assignment. The total number of pages for your paper (i.e. main sections, excluding references and Appendices) cannot exceed 4 pages (with a minimum page margin of 2.5cm on each side), using single line spacing, a two-column format, and a minimum font size of 12 in Times New Roman font).