Lecture 0 - Course Introduction

Guiliang Liu

The Chinese University of Hong Kong, Shenzhen

DDA4230: Reinforcement Learning
Course Page: [Click]

DDA 4230 Resources

Check our course page.

Please post your question on the discussion board in the BlackBoard (BB) system.

- Step 1: Search for existing questions.
- Step 2: Create a thread.
- Step 3: Post your question.

Course Page Link (all the course relevant materials will be posted here):

https://guiliang.github.io/courses/cuhk-dda-4230/dda_4230.html

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Course Stuff

Instructor: Guiliang Liu

- **Office Hours**: Monday 2:50 PM -3:50 PM.
- **Office**: Room 302, Teaching Complex C.
- **Email**: liuguiliang@cuhk.edu.cn

Lectures:

- **Room**: Room 302, Teaching Complex C.
- **Hours**: Monday & Wednesday, 1:30 - 2:50 PM.
- Onsite only, No online lectures.

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Course Stuff

Teaching Assistant: Bo Yue / Hengming Zhang

- **Office Hours**: Friday 5:00 PM -6:00 PM.
- **Office**: Room 611, Teaching Complex B (TXB, the new building).
- **Email**: 224040353@link.cuhk.edu.cn/225045014@link.cuhk.edu.cn

Tutorial:

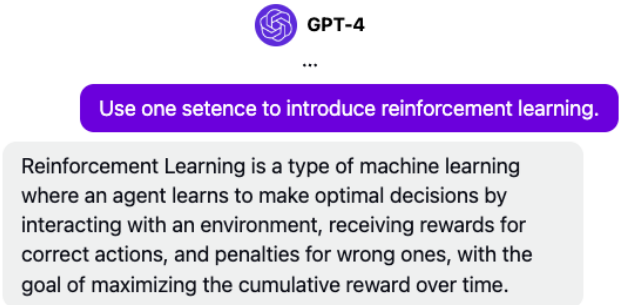
- **Room**: Bldg 206, Teaching C (TC) Building.
- **Hours**: Tuesday, 20:00 PM -20:50 PM.

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Course Content

What is Reinforcement Learning?

Course Content



Keywords: Make Optimal Decisions, Interact with the Environment, Maximize Cumulative Rewards.

Course Content

Part I

- **Markov decision process**: formulation and definition, properties, Markov chain, Markov process and its solution and covering time, MDP examples, policy function and value function.
- **Bandit algorithms**: greedy algorithms, explore then commit, upper confidence bound, Thompson sampling, lower bounds.

Course Content

Part II

- **Dynamic programming:** dynamic programming, policy iteration, value iteration, search algorithms.
- **Monte Carlo methods**: MC policy evaluation, policy gradient, off-policy learning
- **Temporal-difference methods**: SARSA, eligibility traces, Q-learning.
- **Deep variants of the aforementioned algorithms**: deep Q-learning, policy gradient methods, and actor-critic methods.

Part III

Course Content

- **Applications and recent advancements of reinforcement learning.**

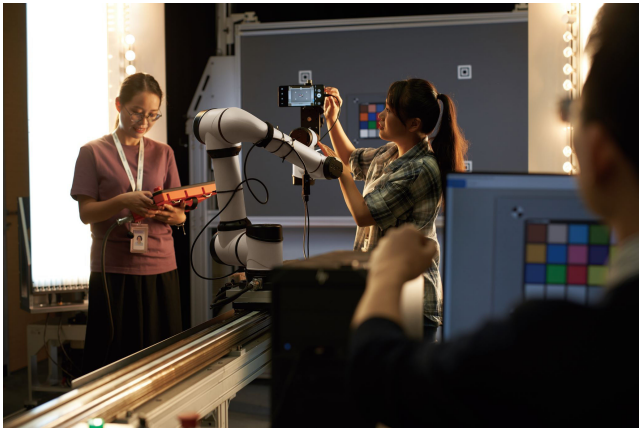
  Monta Carlo Tree Search (Bandit algorithms and Monte Carlo methods)

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Course Content

- **Applications and recent advancements of reinforcement learning.**

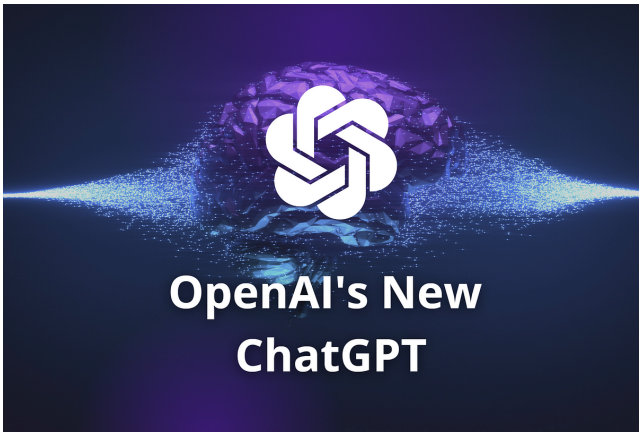
  Embodied AI (Temporal-difference methods and Deep RL)



Course Content

- **Applications and recent advancements of reinforcement learning.**

  Reinforcement Learning from Human Feedback (Policy Gradient Methods)



Scoring Scheme

Component	Weight (%)
Assignments (written and coding homework)	30%
Midterm exam	20%
Final project	50%

- **4 Assignments:** Release an assignment per 3 weeks. Due in 3 weeks. (Please submit the assignments on Blackboard.)
- In-Class **Midterm Exam:** November **3rd (Monday)** 1:30 PM to 2:50 PM. This takes a lecture session. (The midterm is onsite.)
- **Final Project:** Due on December 20th (Friday).

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Assignment

The assignment will take the format of:

- **Question and Answer.** Answer the questions related to the course content. The questions include proof and calculations.
- **Code implementation.** Implement the algorithm, run the program, and report the answer.

Each student should finish the assignment on his or her own(No group work).

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Midterm Example

The midterm exam will be composed of 3-5 regular questions. It will test definitions, formulations, examples, and analyses. No code and implementations.

- **Cheat sheets are allowed.** Two pages of A4 paper can be taken into the exam room. You can write anything or print anything on it.
- **No Electronic Equipment.** No cell phone, laptop, or IPad.
- **No Chatting.** Do not chat with other people. Students should finish the exam by themselves.

Final Project

The final project (50%) will be the final showcase of your exploration in reinforcement learning. Choose one out of:

1. Write a publishable academic manuscript of your research work.
2. Write a literature review of your topic of interest.

The report should be in a single column and be scripted in LaTeX. You should include the main content in the first 10 pages. Content after 10 pages are treated as appendix.

For 1, collaboration (inter and intra-course) is allowed and each student will be marked by their part of the contribution.

For 2, each student should submit their own review.

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Final Project

Please submit a final project proposal (scripted in LaTeX, limited to 1 pages and preferably shorter) to tell us what you plan to do in your final project.

- Few proposals will be refused and the student should find a new topic.
- The proposal will not be marked (pass/refuse only)
- Due on November 15th (Friday).

Final Project

I will provide examples of each of these options (check our homepage or Slack Channel).

Dos

- Use Google Scholar https://scholar.google.com/.
- Use dblp https://dblp.org/.
- Use LLMs for learning or searching .
- Read online blogs https://lilianweng.github.io/posts/2023-06-23-agent/.

Don'ts

- Copy and paste from LLMs like GPT https://poe.com/GPT-4.

Final Project

Final Project Examples:

Yihan Pan, Zhenghang Xu, Jin Guang, Jingjing Sun, Chengwenjian Wang, Xuanming Zhang, Xinyun Chen, J.G. Dai, Yichuan Ding, Pengyi Shi, Hongxin Pan, Kai Yang, Song Wu. "A high-fidelity, machine-learning enhanced queueing network simulation model for hospital ultrasound operations." 2021 Winter Simulation Conference (WSC). IEEE, 2021.

Jing Dong, Ke Li, Shuai Li, Baoxiang Wang. "Combinatorial bandits under strategic manipulations." Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 2022.

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Drop and Late Policy

- **Late Drop.** A late drop from the course is not encouraged. Under special circumstances, students may apply for a late drop, but there is no guarantee that the request can be approved by the school office.
- **Late Submission.** A late submission should receive a 10% penalty for each date after the due. Note that the penalty can accumulate until it reaches 100% (late for 10 days). If you need special care (e.g., for surgery and other health problem), PLEASE let me know in advance (see my contact in the previous page).

Honesty in Academic Work

The Chinese University of Hong Kong, Shenzhen places very high importance on honesty in academic work submitted by students, and adopts a policy of zero tolerance on academic dishonesty. Sub-categories include (please refer to this page):

- Plagiarism.
- Undeclared multiple submission.
- Cheating in tests and examinations.
- Distribution/ Sharing/ Copying of teaching materials without the consent of the course teachers to gain unfair academic advantage in the courses.

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Course Resource

- Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018. main book
- Lattimore, Tor, and Csaba SzepesvÃ¡ri. Bandit algorithms. Cambridge University Press, 2020.
- Bertsekas, Dimitri. Dynamic programming and optimal control: Volume I. Vol. 4. Athena scientific, 2012.
- SzepesvÃ¡ri, Csaba. Algorithms for reinforcement learning. Springer Nature, 2022.

Question and Answering (Q&A)

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen