ECE 493 Topic 42 - Probabilistic Reasoning and Reinforcement Learning

Note: *Also Known As ECE 457C Reinforcement Learning*

Offered: Spring 2021	Instructor: Prof. Mark Crowley
Website	Updated List of Resources and Links
Piazza	YouTube

Course Description

Introduction to Reinforcement Learning (RL) theory and algorithms for learning decision-making policies in situations with uncertainty and limited information. Topics include Markov decision processes, classic exact/approximate RL algorithms such as value/policy iteration, Q-learning, State-action-reward-state-action (SARSA), Temporal Difference (TD) methods, policy gradients, actor-critic, and Deep RL such as Deep Q-Learning (DQN), Asynchronous Advantage Actor Critic (A3C), and Deep Deterministic Policy Gradient (DDPG).

Required Background

The course will use concepts from ECE 203 and ECE 307 on Bayesian Probability and Statistics, these will be reviewed but familiarity will help significantly. All other concepts needed for the course will be introduced directly. Examples, assignments and projects will depend on programming ability in Python.

Course Staff

Instructor: Prof. Mark Crowley	TA: Nham Van Le	TBD
Contact: mcrowley@uwaterloo.ca (but for better results, use piazza or book an appointment via doodle)	**Contact:** nv3le@uwaterloo.ca	**Contact:** tbd
Office Hour: Bookable Meetings	**Office Hour:** TBD	

Weekly Schedule

- Most of the **lectures are available already on Youtube**, *each week* there are associated videos to watch along with readings from the Sutton and Barto textbook.
- A Detailed Course Calendar in the Course Gingko List will let you know which materials to study each week before Monday's Live Session.
    - **Live Lecture/Review/Discussion Sesssion -** Every **Monday 4pm - 5:30pm (ET)** there will be a live session with the Prof which will be:

        - *Either* a live lecture with new/revised content from what is on YouTube. Lectures will be recorded and available later on YouTube.
        - *Or* a live discussion session to review the content from last week's self-study, working through problems, answering questions, breaking out for discussions. These sessions will be recorded and remain available on LEARN during the term.
    - **Tutorials:** TBD - we'll see how, and if, we do tutorials this year with the new online arrangement.
- ***Bookable 1-on-1 Meetings* - Mondays 11:15am-12:00pm and Fridays 10:00am - 11:am**
    - These are regular, one-on-one (or a small group) meeting slots with the professor, bookable via doodle a few days ahead of time.
    - Doodle Bookable Calendar
- **TA office hours** - to be determined (some times on Tuesday-Thursdays)

Grade Breakdown

- Assignment 1: **15%** - Implement fundamental, exact algorithms on simple domain such as grid world, Value Iteration, Policy Iteration. *Evaluation*: Some automated grading as well as code review.
- Assignment 2: **15%** - Implement RL algorithms for simple domain, SARSA, Q-Learning, Eligibility Traces. *Evaluation:* Short report with graphs, automated grading and code review.
- Midterm: **20%**
- Assignment 3: **20%** - Implement Policy Gradient, Actor Critic, Value Function Approximations for larger RL domains. Implement RL algorithm for more complex domain using simple Deep Learning representation of value function.

  Evaluation: Short report with graphs and code review. We might set-up at kaggle in-class competition as well.
- **Final Exam**: **30%** - all topics, leaning towards latter half of course

Course Details

Learning Objectives

This course complements other AI courses in ECE by focussing on the methods for representation and reasoning about uncertain knowledge for the purposes of analysis and decision making. At each stage of the course we will look at relevant applications of the methods being discussed.

For example, in 2016 the AI program "AlphaGO" defeated human world class players of the game Go for the first time. This system requires many different methods to enable reasoning, probabilistic inference, planning and decision optimization. In this course we will build up the fundamental knowledge about these components and how they combine together to make such systems possible.

1. Identify and Explain the component theoretical concepts of Reinforcement Learning systems.
2. Implement or instantiate using a library any of the core Reinforcement Learning algorithms on a variety of domains.
3. Evaluate the performance of a particular RL system on a given domain through proper experimental design, statistical analysis and visualization.

Topics

1. Motivation and Context
    - Importance of reasoning and decision making about uncertainty.
    - Connection to Artificial Intelligence and Machine Learning.
    - Probability review.
2. Decision making under uncertainty:
    - Multi-Armed Bandit (MAB) problems, Thompson Sampling.
    - Markov Decision Processes (MDPs), Influence Diagram representation.
3. Solving MDPs
    - Theory, Bellman equations
    - Relation to Control Theory
    - Value Iteration, Policy Iteration
4. The Reinforcement Learning Problem
    - Approximately solving MDPs by interacting with the environment
    - SARSA algorithm
    - Q-learning algorithm
5. Temporal Difference Learning
    - Eligibility Traces
    - TD(λ)
6. Direct Policy Search
    - Policy Gradients methods
    - Actor-Critic methods
7. State Representation
    - Value Function Approximation
    - Stochastic Gradient Descent
8. Basics of Neural Networks (review or refer to ECE657A content)
    - fully connected, multi-layer perceptrons
    - supervised training, back-propagation
    - regularization methods
9. Deep Reinforcement Learning
    - Deep Q- Networks (DQN)
    - Experience replay buffers and mini-batch training
    - A2C, DDPG, PPO
10. Other Challenges (brief)
    - Partially Observable MDPs (POMDPs)
    - Multi-Agent RL (MARL)
11. Other ways to solve (PO)MDPs (if time permits)
    - Monte-Carlo Tree Search, Explaining AlphaGo
    - Curiousity based learning
    - Soft-Actor Critic

12. Wrap-up and Review

Getting Help:

- **Discussion board:**
    - *Piazza* will be the main place for detailed discussion and questions. Students can post anonymously (from students only), post a collaborative answer and course staff can confirm these, post their own or run Live Q&A events.
    - Go there there and sign up with your UWaterloo email now!
- **Pre-recorded Video Lectures:** These will be made available on the course youtube channel, and links from whithin Learn
- **LEARN Website:** The main course content, announcements, grade stracking and materials will be made availble on Learn. All registered students should see this in their LEARN courses.
- **Email the Teaching Assistant and Instructor:** Office Hours will be arranged once term starts as needed.
- **AccessAbility Services :** http://uwaterloo.ca/accessability-services
    - If you need any accommodation, assistance with exams, learning environment, assignments, talk to this office and they can help you set it up as securely and anonymously as possible.

Discussion Group Protocols

- Posts on Piazza can be public or anonymous to your classmates, but they will *never* be anonymous to the TAs and Instructor.
- Be kind. Assume the best, not the worst. Think before you hit enter.
- Posts which are considered offensive, abusive, bullying, discriminatory to any group or person, will be made private or deleted and followed up with private discussion.
- If you feel there is innapropriate, hurtful behaviour occuring on the dicussion forum, please notify the professor, TAs or department staff as you feel appropriate.
- If you really can't get in touch with anyone and it is an emergency you can contact Prof. Crowley directory via Microsoft Teams messaging (please don't abuse this though :|)

General University of Waterloo Guidelines:

Academic Integrity: In order to maintain a culture of academic integrity, members of the University of Waterloo community are expected to promote honesty, trust, fairness, respect and responsibility. Check http://www.uwaterloo.ca/academicintegrity/ for more information.

Grievance: A student who believes that a decision affecting some aspect of his/her university life has been unfair or unreasonable may have grounds for initiating a grievance. Read Policy 70, Student Petitions and Grievances, Section 4, http://www.adm.uwaterloo.ca/infosec/ Policies/policy70.htm. When in doubt please be certain to contact the departments administrative assistant who will provide further assistance.

Discipline: A student is expected to know what constitutes academic integrity---check http: //www.uwaterloo.ca/academicintegrity/ to avoid committing an academic offence, and to take responsibility for his/her actions. A student who is unsure whether an action constitutes an offence, or who needs help in learning how to avoid offences (e.g., plagiarism, cheating) or about rules for group work/collaboration should seek guidance from the

course instructor, academic advisor, or the undergraduate Associate Dean. For information on categories of offences and types of penalties, students should refer to Policy 71, Student Discipline, http://www.adm.uwaterloo.ca/infosec/Policies/policy71.htm. For typical penalties check Guidelines for the Assessment of Penalties, http://www.adm.uwaterloo.ca/infosec/guidelines/penaltyguidelines.htm.

Appeals: A decision made or penalty imposed under Policy 70 (Student Petitions and Grievances) (other than a petition) or Policy 71 (Student Discipline) may be appealed if there is a ground. A student who believes he/she has a ground for an appeal should refer to Policy 72 (Student Appeals) http://www.adm.uwaterloo.ca/infosec/Policies/policy72.htm.

Note for Students with Disabilities: The Office for Persons with Disabilities (OPD), located in Needles Hall, Room 1132, collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations to lessen the impact of your disability, please register with the OPD at the beginning of each academic term.