Reinforcement Learning 1

CMPUT 655

Fall 2020

Martha White

Goals for this course

- **Primary goal 1:** Give you dedicated time to learn the fundamentals of RL
  - Otherwise, to take RL courses, you need to do the MOOC on your own time, and do not have the opportunity to discuss and truly understand the fundamentals

- **Primary goal 2**: Complete a research paper in reinforcement learning, intended for a specific venue
  - workshop, conference or journal
  - implicitly, the goal also towards helping you become better at research in this area, including assessing a few papers in RL

Course Structure

- Complete RL Mooc by October 21 (including the midterm)
- Identify your project before October 21
- Focus on your project after October 21

Grading

- Assignments (on Coursera): 30%
- Midterm Exam: 15%
- Mentorship: 10%
- Initial draft: 10% (November 16)
- Final draft: 35% (December 11)

RL Mooc Weekly Structure

- Each week you will complete two modules
- This involves
  - watching about 2 hours of videos (in total for both modules)
  - some readings from the book
  - typically 1 practice quiz and 1 graded notebook per module
- On slido, you will submit a discussion question or upvote existing discussion questions about either/both modules
- On Monday, we will discuss these in class, with some lecturing from me

Birds-eye view

- Mini-Course 1: Fundamentals of RL
  - Bandits (A miniature version of the entire course)
  - The Model-based setting, where someone gives you how the world works
- Mini-Course 2: Sampled-based Learning Methods
  - Learning only through trial-and-error interaction (TD, Monte Carlo, Dyna)
- Mini-Course 3: Prediction and Control with Function Approximation
  - Extending all the stuff before to the setting where we have to approximate the functions/models (e.g., using neural networks)
- Mini-Course 4 is Optional (Putting it all together)

Why is there a midterm?

- (Aka, a final for the Mooc content)
- The unfortunate law of the universe is that exams encourage you to really learn the material
- (And this material matters)

Mentorship sessions

- You will lead a discussion group in the undergraduate RL course
  - Plan to do this three times

- As a grad student, you will be in a leadership role

- The goal of each session is to help each member in your group identify points of confusion and summarize those to give to me

- A big part of mentorship is helping an individual express themselves and help them address confusion

- These occur from 1-2 pm on Wednesday. Does that work for you?

Let’s poll with slido

- Go to: https://app.sli.do/
- Use event code: Y956
- Answer the poll, about Wednesday discussions
- You can also put any questions/discussion topics you want there right now, for class today

Primary Goal 2

- Become exposed to research in this area
  - complete a literature survey for your project
  - read a few joint papers in the group
  - discuss research projects in the group
- Be part of a larger research group, pushing ideas and understanding forward together, in a collaborative way
- Complete a concrete research project, from start-to-finish with a paper that can actually be submitted by the end of the course

Completing a project to the end is important and non-trivial

- The last 10% can be the most difficult
- You should get experience polishing a paper, in the right format for a venue, of the right length
  - …we all become latex experts. vspace is your friend
- Once you write the paper to be a concise paper with a clear message, it can significantly change what you thought you set out to do
  - For other research projects in your past, it was likely not a strong requirement that the paper would have to pass peer review
  - When you know it's going to be reviewed, you have to ask yourself: "What is the scientific community getting from this paper? What am I trying to show? Is it meaningful and did I actually show it?"

Working as a team is good

- Projects must be done with others: groups of 2-4
- There will be related topics across projects (e.g., exploration); you should draw on the insights from your peers to make progress on your project
- We have a slack channel to facilitate discussions about papers, projects, etc.; let's be active there, it will be way more fun!

Communicating is Key

- Once we are done the MOOC (by October), you will describe progress weekly to your classmates

- We will be trying a **weekly stand-up**, where we break-up into subgroups and update on progress for the past week and goals for the coming week
  - The update should be relatively short, a few minutes

- The goal of the stand-up is
  - to hear about each others projects
  - make you more organized on your weekly progress
  - get feedback and suggestions to get unstuck (e.g., my NN is blowing up)

Learning to do Good RL research

- I will spend some in-class time explaining how to scope a project, some pitfalls to avoid in RL research and how to write an RL paper

- We will read a few papers as a group, and critically evaluate them
  - You will submit a paper you would like to read, likely related to your project
  - We will pick amongst your requests

- This should help you more critically evaluate your own work, and also help identify what should not be imitated in existing work

Disclaimer about Goals

- This course is like a Research in RL 101
- You are unlikely to learn advanced ideas in RL
- If your primary goal is to complete a concrete research project in RL, then **take this course**
- **Do not take this course** if you are expecting to go much beyond the material in the Mooc
  - or at least we will not go beyond the Mooc material in a carefully crafted way

Let’s break-out into discussion groups

- Goals: to meet your fellow classmates, help identify your goals for the course, and the delta in your learning

- In your group, answer the following questions:

- **What are you hoping to get out of the course?**

- **What topics do you think you already understand well?**

- **What topics are you hoping to get more insight into?**

- **What topics do you wish we could cover in supplementary lectures?**

Identifying a Project

- Even before you have the RL background, you will need to
- 1. Find a team for the project
- 2. Identify the project question
- This is hard to do before having all the background, but a necessity to give enough time to complete the project
- You must submit a short project proposal by October 12
  - 1-2 paragraphs clearly identifying the problem being tackled

Options for Identifying a Project

- **Option 1:** You already have a project in mind or are already working on a research project. If want to find someone to work with, feel free to post it on slack to find a partner.

- **Option 2:** We have created a list of potential projects, in collaboration with RLAI. Each project has a name beside it, to consult for more details and potentially to collaborate with on the project (even if they are not in the course)

- **Option 3:** I encourage talking to professors in the department, to work on a project with them or on a problem that they specify.

Criteria for selecting a project

- Pick a concrete, feasible topic
  - Too broad: “I am going to investigate exploration”
  - More specific: “I am going to investigate a simple idea I have for estimating counts for a state, under function approximation, for use in count-based exploration approaches”
- Err on the side of small —> the goal of this course is to get you to do a simple (small) idea well, not a big idea poorly
  - In the beginning, many projects start smaller as you become more knowledgeable in the topic
  - Once you have a good background, you can more easily know if your ideas are novel and/or write that bigger journal paper
- Can relate to an on-going research project outside of this course

But I’ve done projects before…

- Projects in most courses are not expected to actually be submitted to a (reinforcement learning) venue
- The ultimate goal of this course is to get you a big step of the way to the ambitious goal of completing a publishable paper
  - the goal is not about getting a course project completed, which can be meandering and full of a bunch of stuff
- The big step towards the final full paper is to complete a smaller portion of that paper well

An example successful strategy

- You want to conduct an empirical study comparing two representative algorithms for policy search in RL

- Your ultimate goal is to test two hypotheses about when one might be more beneficial than the other

- For the project, you can thoroughly answer just the first question, with only one environment

- This gets 1) the paper written, 2) a thorough literature survey in place and 3) the empirical methodology well worked out

An example successful strategy (cont…)

- To actually get accepted, you need to do a bit more
- But you have laid a solid foundation to complete this outside the course
- In your paper itself, you can write up a Next steps/Limitations section, to highlight what still needs to be done

An example of a failing strategy

- You want to see if you can extend an interesting idea, with a simple strategy
- You implement it and find it is not better
- Then you think: Maybe I just didn’t make my network complex enough, or use a new fancy algorithm?
- 2 months later you are still fiddling, and do not have a concrete answer to if they idea is promising, and why or why not

An alternative response to failure in the idea

- **Step back and simplify**

- Understand why it is not working, likely on a simpler more carefully crafted toy environment

- Set-up an experiment where you are *sure* you know what the outcome will be (you might surprise yourself when you run it)

- **Strive for understanding**, not for fixing/performance

Potential concerns

- What if the idea does not work out?
  - Insight gained from a concrete project almost always leads to a reasonable direction, that can be at least published in a workshop
  - For this course, I will allow negative results, if you try an idea that many would agree should be effective
- What if I do not complete the final draft on time?
  - For example, I specified a broad topic and didn't end up completing it in time
  - This is a project course (not an easy course); you have to complete the project
  - You will hopefully learn an important skill: scaling back a project to be the right size, even it seems to be getting out-of-hand
  - I recommend ensuring that progress is being made regularly, and that the scope of the project remains feasible; if not, talk to me early

Project Approval

- I will approve the project, after looking at the proposals
  - And I might give suggestions for scoping
- The main criteria for approval are: is it a sensible question and is it feasible?
  - Is it ok that I'm double dipping, and using a similar project across courses? Yes (but make sure they are still complementary, not exactly the same)
  - Is it ok that I'm working with a senior PhD student, that is doing some of the work? Yes (if this course helps you get research done, awesome)

Sharing code and getting access to computation

- My group (and RLAI) has some code
  - Ask on slack, and Andy and Shivam might point you to some repos
- There are frameworks available out there for experiments
- **Computation**: ask your advisor for access to computation
  - if you do not yet have an advisor, then your temporary one might give you access to their Compute Canada account. You could ask

Some logistics

- Andy and Shivam are (volunteer) TAing this course
- They are amazing and will try to help answer questions, but do recognize that they have limited time
- I will have office hours for additional help from me, but that will be shared across everyone
  - I cannot feasibly manage 30 research projects in this course
  - I am here to support and facilitate
- One of the goals of having RLAI specify some projects, is to have this additional point of contact

Immediate announcements

- A discussion question should be submitted for the first two modules, by Sunday
- You must complete the graded components for the first two modules by next Thursday
- Andy is going to cover background material next week Wednesday, and likely on Monday the following week
  - Due to unfortunate circumstances, I am missing both of these
- These two sessions will also include in-class discussions

Let’s switch now to organizing projects!

Some project types

- Empirical Study — requires a thorough evaluation
- Algorithmic Project — a new idea, where sometimes it is enough to do a demonstration rather than a thorough study
- Theory Project — provide an answer to an open theoretical question
- Applied Project — this one is not listed in the google doc, because an Applied project might take too much time in this course, unless you are bringing a partially completed

Empirical Study

- An empirical study is not easy!
  - Ignore most of the papers in RL; they are not really doing empirical work, typically they are really just showing that their algorithm runs
- It requires 1) designing an experiment to highlight what you would like to see and 2) using a sound, and typically exhaustive methodology
- You need sufficient runs, fair choices for algorithms, etc.
- e.g., https://arxiv.org/abs/1602.08771

Algorithmic Project

- You need to clearly motivate what problem you are solving and why your proposed idea solves this problem
- You need results substantiating these claims
  - Theory, without experiments, is enough
  - Else, your experiments should provide evidence for your claims
- Ask yourself: do I have evidence supporting my claims about this algorithm? Did I design my experiment to highlight what I want to show about this algorithm?

Some topic areas

- Policy evaluation and prediction, especially in the off-policy setting (data gathered under one policy, want to evaluate another)

- Improving (stable) online control, with policy gradient methods and/or action-value methods

- Optimization strategies for RL, and (hyper)parameter selection

- Exploration

- Using learned models in RL

- Batch RL, including learning from a batch for pre-training

An example of a concrete project

- Title: Understanding the role of partial greedification in API

- Motivation: Policy gradient methods can be seen as an instance of API, where the greedification step is not fully greedy. Instead, the policy parameters are only updated a part of the way to the greedy policy. This contrasts methods like Q-learning. There is some reason to believe, however, that avoiding fully greedy steps is beneficial for stability in learning and convergence to better solutions, particularly as the action-values are incorrect during learning. We investigate the benefits, if any, of using partially greedy steps in a simple toy environment. We vary the accuracy of the action-values during learning (both due to estimation and approximation), as well as the amount of greedification per step, within Soft Actor-Critic for discrete action problems.

Let’s break-out into discussion groups

- Goals: to meet your fellow classmates, start finding potential collaborators and start finding a project

- In your group, answer the following questions:

- **Do you already have a project in mind? If yes, what is it?**

- **If no, what topics are you interested in?**

- After discussing with the first group for 15 minutes, we can do one more random shuffle