CMPUT 365

Reinforcement Learning

Fall 2021

Instructor: Adam White
University of Alberta

Covid, Campus & You

- These are strange and sometimes scary times
    - Things will evolve throughout the term & this will be a continuing discussion
- The university expects everyone of you to:
    - Wear a mask if within 2m of someone else
    - Be vaccinated or frequently show negative covid test
- Let’s use common sense and respect other people’s wants:
    - Respect physical distances!!!
    - None of us are in a position to judge, be rude, or shame others!
- Any questions?

Covid chat

- What are your concerns?

Some background

- This course **used** to be taught as CMPUT 366 & 397
- It was time to make a Reinforcement Learning course
  - The UofA is a world-leader in RL
  - The approaches in RL will be useful in science & industry
- We made a MOOC, to make the topic more accessible to the world

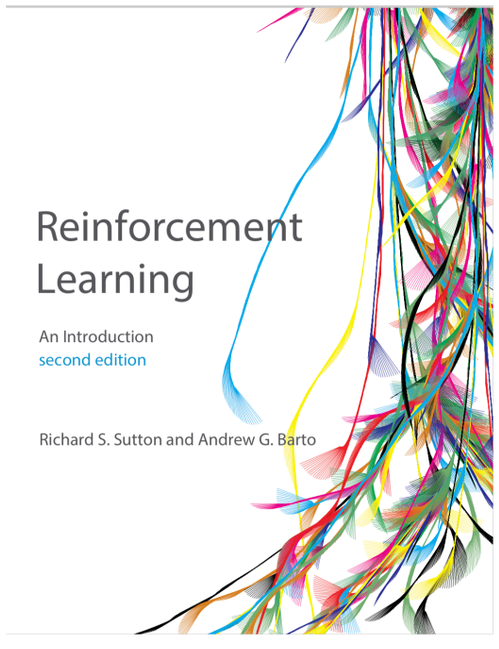
This Course

- **We will use the RL MOOC for this course (all lectures)**
- In-class time will be spent on
  - Summarizing the week's videos main ideas
  - Practice quiz review
  - Labs to help you on your notebook assignments
  - Worksheets and short answer questions
  - Q&A every lecture (Discord and in-class)

Course Information

- Github pages with updated schedule and syllabus
- Coursera RL Mooc
- Course eClass page
  - Zoom link, announcements, and place to submit work
- Getting Started and FAQ on Github page

Textbook



- Readings will be from: Reinforcement Learning: An Introduction, by R Sutton and A Barto, MIT Press.
    - available freely online
    - you can order online - be wary of fake sellers

Registering for RL on Coursera

- **We have our own private session, on Coursera**
- Please register today —> the first item is due next Tuesday!
- See the announcement and the Getting Started document linked
- **If you don’t use the private session you won’t get credit for submitted work! This is very important!!**

Flipped class

- The idea is that you actually do the readings and watch the videos on your own!
  - The course material strongly builds on itself
  - The assignments and quizzes build
  - You need to do this work out of class
- **We will not re-lecture the content of the video!!**

Flipped class

- Lecture time will be used to:
  - clarify things that you report are unclear or difficult (we have a good sense of those things from previous years)
  - Supplement the readings and videos with extra exercises, background, and extra content
  - Help you a bit understanding the quizzes and with your notebooks

Evaluation

- Assignments/Quizzes (completed in Coursera) – 30%
- Project - 5%
- Participation - 5% (quizes on Coursera)
- Mini essays (x2): 5%
- In-class quiz: 5%
- Midterm – 20%
- Final – 30%

Weekly Quizzes and Assignments

- Each week is a different module, with an associated Practice Item and Graded Item
    - usually a **Practice Quiz and a Graded Notebook** (programming)
- In preparation for class, on your own you need to:
    - Watch the lectures online (at most 1 hour of time)
    - Do the assigned reading for that module (at most 1 hour)
    - Complete the quizzes/assignments (about 3-4 hours)

Weekly Quizzes and Assignments

- **You must complete the practice component by EOD Sunday and the graded component by Noon Friday**
    - All due dates are in coursera and in schedule

Project

- Most will complete the capstone project (Course 4 of the RL Mooc)
- OR it might be possible to join a project with a graduate student, in the RL graduate course
    - less clear-cut and more difficult (research is hard)
    - please email me if you want to do this, but do know that it could be time consuming and risky

Evaluation

- Assignments/Quizzes (Graded Items on Coursera) – 30%
- Project (Capstone on Coursera) - 5%
- Participation (Practice Items) - 5%
- Mini essays (x2): 5%
- In-class quiz: 5%. //questions from weekly practice quizzes
- Midterm – 20%
- Final – 30%

Discord

- This is your way to ask questions as you go through the material on your own.
- Go to Discord
  - post a question
  - During lecture or after
- In-class lecture content will be driven by these questions

In-class Sessions

- Monday: Review module material and review practice quiz questions

- Wednesday help on your notebook with TAs (like a lab) & worksheet questions

- Friday additional lecture material, worksheet questions

- Everyday: Q&A and I address some of your questions

Worksheets

- We will post a worksheet with 1-2 questions to do in class
- help reinforce the material
- get you to answer more open-ended questions than the quizzes on the platform
- give you a better idea what long-form answers might look like on an exam

Disclaimer: We might adjust as we go along

- CMPUT 365, as a flipped class, is still quite new
- The COVID situation might change things too

- Additionally, if you find something about the structure is not working, or wish we could do something else in class, we might be able to do it! So feel free to tell me.

No Lab

- There is No Lab

- In-class time is already hands-on

Grades are not based on a normal curve

- Grades are relative to your fellow classmates, but I do not fail the bottom X% (thus, not a normal curve)

- We will provide letter grade boundaries at the end of the course

- Letter grades are provided by a clustering of percentages

  - This allows for adjusting due to yearly differences

- This is a third year course, so grades are typically skewed a bit higher

Prerequisites

- Some comfort or interest in thinking abstractly and with mathematics
- Elementary statistics, probability theory
    - conditional expectations of random variables
        - there will be two class sessions devoted to a tutorial review of basic probability and statistics
- Basic linear algebra: vectors, vector equations, gradients
- Programming skills (Python)
    - If Python is a problem, we have posted some tutorials
    - **Dont plan to learn how to program in this course—we won't teach that!**

Instruction Team

- Prof: Adam White
- TAs (grad students doing research in RL)
  - Andy
  - Tian
  - Yongchang
  - Subhojeet

Contacting us

- Use Discord
    - Start here if you can. Others have your question too!
- Use course email to email TAs: *coming soon*
- For personal issues (e.g., missing your exam), please email Adam
- See FAQ for more details

Office Hours

- I have office hours on Wednesday, 2-4 (after class)
- TAs will list their office hours on eClass & Github page
- We will try to accommodate different timezones
    - Note: all in-class sessions will be recorded too to accommodate different timezones

Collaboration

- Working together to solve the problems is encouraged
- But you must write-up your answers individually
- You must write your own code; every single character from your fingers
  - **No copy-paste from your friends or the internet!!!**
- You must acknowledge all the people you talked with in solving the problems

What is Plagiarism

- Taking things from others and passing it off as your own work without credit

Test time: are these ok?

- Writing down answers to assignments in a group?
- Getting a tutor to help write your code?
- Letting your friend look at your code or assignment question?
- Searching for and using assignment solutions from the internet?
- Not indicating on your assignment who you talked with?
- Discussing ideas without writing anything down?

Policies on Integrity

- Cheating is reported to university whereupon it is out of our hands

- Possible consequences:
  - A mark of 0 for assignment
  - A mark of 0 for the course
  - A permanent note on student record
  - Suspension / Expulsion from university

Academic Integrity

- The University of Alberta is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behavior (online at www.ualberta.ca/secretariat/appeals.htm) and avoid any behavior which could potentially result in suspicions of cheating, plagiarism, misrepresentation of facts and/or participation in an offence. Academic dishonesty is a serious offence and can result in suspension or expulsion from the University.

Course Overview

- Main Topics:
  - Learning (by trial and error)
  - Planning (search, reason, thought, cognition)
    - Prediction (evaluation functions, knowledge)
    - Control (action selection, decision making)
- Recurring issues:
  - Demystifying the illusion of intelligence

Birds-eye view

- Mini-Course 1: Fundamentals of RL
    - Bandits and 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
- 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)
- Putting it all together (your project) is Mini-Course 4

High-level view

- Bandits and Online learning (Ch2, C1M1):
  - formalizing a problem and discussing solution methods
  - A miniature version of the entire course
- Markov Decision Processes and Value Functions (Ch3, C1M2 and C1M3):
  - Our formalization of reinforcement learning and AI…no solution methods here
  - Students usually get impatient here

High-level view (2)

- MDP solution method, given a model:
  - Dynamic programming - (Ch 4, C1M4)
- MDP solution methods, if you can only learn from interaction
  - Monte Carlo (MC) - (Ch 5, C2M1)
  - Temporal difference learning (strengths of both DP and MC) - (Ch 6, C2M2,C2M3)
- Planning with learned models - (Ch 8, C2M4)

High-level view (3)

- Everything up to and including Chapter 8 is tabular solution methods:
  - The foundation of modern RL
- In Chapters 9, 10, 13 cover approximate solution methods:
  - Function approximation (including Neural Nets)
- The foundations established in chapter 3-8 will largely transfer to the function approximation case

Let’s chat

- Let’s breakup into groups, led by a TA
- Go around the circle and **answer one of** the questions
  - Q1: What you hope to get out of this course?
  - Q2: Do you think you will use RL in your future career?
  - Q3: Why are you interested in RL?

Goals of Artificial Intelligence

- Scientific goal: 
  - understand principles that make rational (intelligent) behavior possible, in natural or artificial systems.
- Engineering goal: 
  - specify methods for design of useful, intelligent artifacts.
- Psychological goal: 
  - understanding/modeling people
  - cognitive science
- Philosophical goal: 
  - Understand what it means to be a person
  - Understand humanity’s role in the universe

Intelligence (mind)

- “Intelligence is the computational part of the ability to achieve goals in the world”

  —John McCarthy

- “the most powerful phenomena in the universe”

  —Ray Kurzweil

Discussion:
Is human-level AI possible?

- If people are biological machines, then eventually we will reverse engineer them, and understand their workings

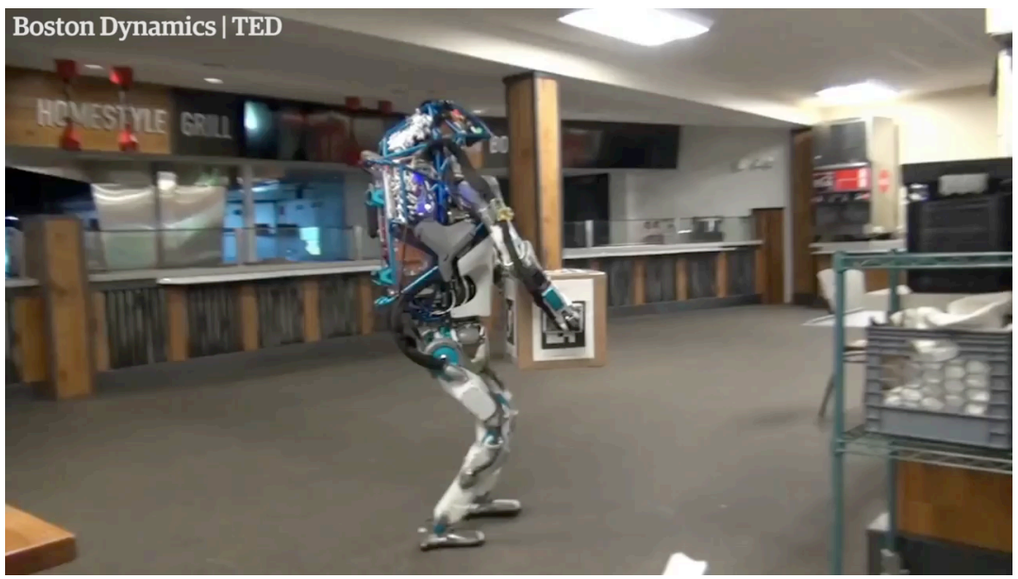
- Then, surely we can make improvements
  - with materials and technology not available to evolution
  - how could there not be something we can improve?
  - design can overcome local minima, make great strides, try things much faster than biology

Cheap computation power drives progress in AI

- Deep learning algorithms are essentially the same as what was used in ‘80s
  - only now with larger computers (GPUs) and larger data sets

- Similar impacts of computer power can be seen in recent years, and throughout AI’s history, in natural language processing, computer vision, and computer chess, Go, and other games

BUT, But! Many fundamental research questions remain unresolved



(Henderson et al, 2018)

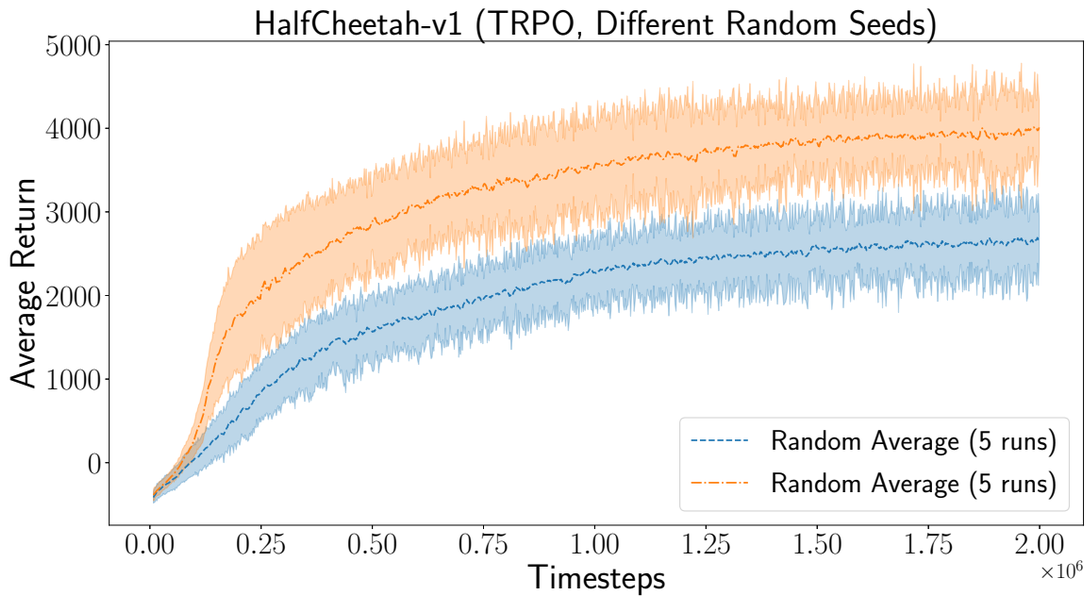


Figure 5: TRPO on HalfCheetah-v1 using the same hyperparameter configurations averaged over two sets of 5 different random seeds each. The average 2-sample t-test across entire training distribution resulted in $t = -9.0916$, $p = 0.0016$.

, while DQN was trained on only 10M frames.

Game	ES	DQN w/ ϵ-greedy	DQN w/ param noise
Alien	994.0	1535.0	**2070.0**
Amidar	112.0	281.0	**403.5**
BankHeist	225.0	510.0	**805.0**
BeamRider	744.0	**8184.0**	7884.0
Breakout	9.5	**406.0**	390.5
Enduro	95.0	1094	**1672.5**
Freeway	31.0	**32.0**	31.5
Frostbite	370.0	250.0	**1310.0**
Gravitar	**805.0**	300.0	250.0
MontezumaRevenge	**0.0**	**0.0**	**0.0**
Pitfall	**0.0**	-73.0	-100.0
Pong	**21.0**	**21.0**	20.0
PrivateEye	100.0	**133.0**	100.0
Qbert	147.5	**7625.0**	7525.0
Seaquest	1390.0	8335.0	**8920.0**
Solaris	**2090.0**	720.0	400.0
SpaceInvaders	678.5	1000.0	**1205.0**
Tutankham	130.3	109.5	**181.0**
Venture	**760.0**	0	0
WizardOfWor	**3480.0**	2350.0	1850.0
Zaxxon	6380.0	**8100.0**	8050.0

(Plappert et al, 2017)

Algorithmic advances in Alberta

- World's best computer games group for decades (see Bowling's talk) including solving Poker

- Created the Atari games environment that our alumni, at Deepmind, used to show learning of human-level play

- Trained the AlphaGo & AlphaStar team that beat the world Go champion

- World's leading university in reinforcement learning algorithms, theory, and applications, including TD, MCTS

- ≈20 faculty members in AI

Job opportunities in Alberta

- Huawei Edmonton Research lab
- Amii
- Deepmind Alberta
- Several new labs and startups on the horizon

Discussion

For you, which of the following are essential abilities of an intelligent system that you would like to learn about (say in this course)?

The ability to:

A. sense and perceive the external world
B. choose actions that affect the world
C. use language and interact with other agents
D. predict the future
E. fool people into thinking that you are a person
F. have and achieve goals
G. reason symbolically, as in logic and mathematics
H. reason in advance about courses of action before picking the best
I. learn by trying things out and subsequently picking the best
J. have emotions, pleasure and pain
K. other?

For you, which of the following are essential abilities of an intelligent system that you would like to learn about (say in this course)?

The ability to:

A. sense and perceive the external world
B. choose actions that affect the world
C. use language and interact with other agents
D. predict the future
E. fool people into thinking that you are a person
F. have and achieve goals
G. reason symbolically, as in logic and mathematics
H. reason in advance about courses of action before picking the best
I. learn by trying things out and subsequently picking the best
J. have emotions, pleasure and pain
K. other?

What is Reinforcement Learning?

- **A statistical approach to AI**

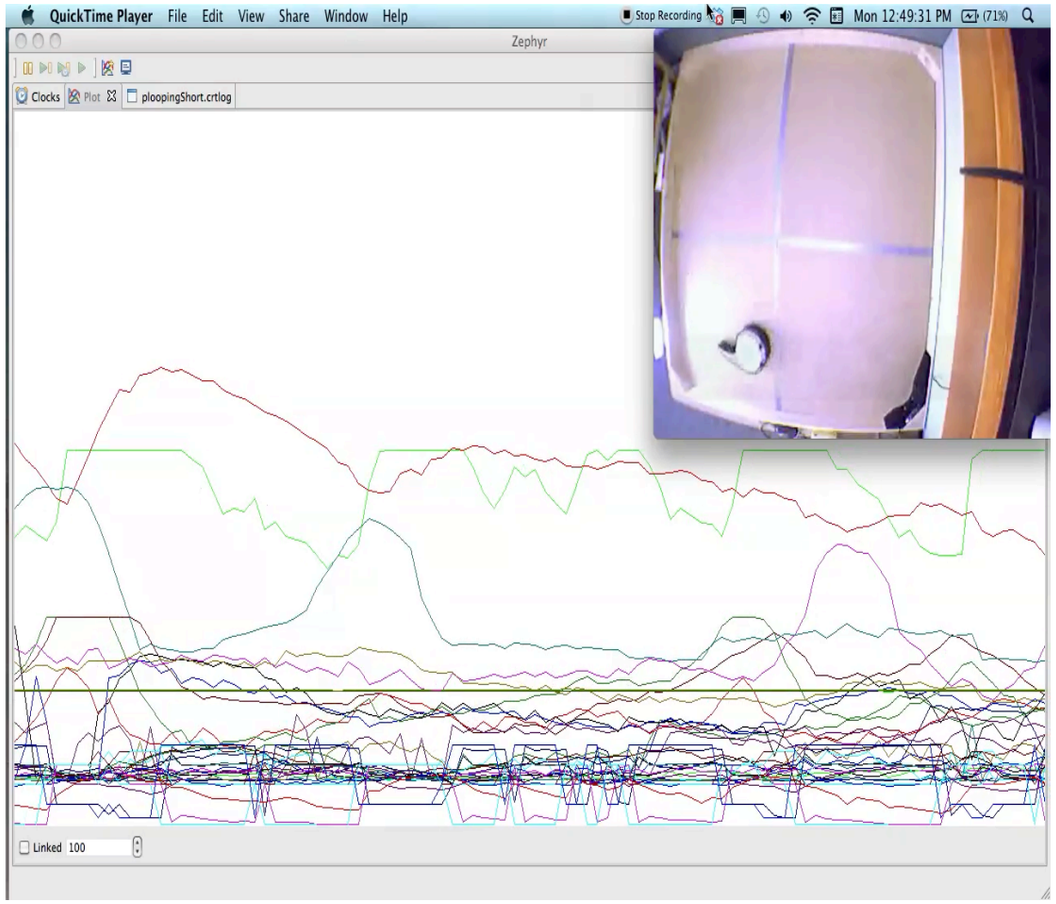
- An agent continually interacting with the world around it, learning to make decisions that result in better outcomes

  - Learning is statistical, because it extracts patterns from the experience it has already gathered

  - A key choice in RL is that goodness of outcomes is measured by a scalar reward signal

  - The agent generates its own training data! It is dependent on prior learning!! No fixed datasets!!!

An Example of an RL System: The Critterbot



Demo of Bandits

- https://www.coursera.org/learn/fundamentals-of-reinforcement-learning/ungradedWidget/44Z9R/lets-play-a-game

- https://www.coursera.org/learn/fundamentals-of-reinforcement-learning/ungradedWidget/jEYTO/whats-underneath