

CMPUT 655

Reinforcement Learning

Fall 2022

Instructor: Adam White
University of Alberta
DeepMind Alberta



About me ...



What is Reinforcement Learning?

- Agent-oriented learning—learning by interacting with an environment to **achieve a goal**
- Learning by trial and error, with only **delayed evaluative feedback** (reward)
 - the kind of machine learning like natural learning (animals)
 - learning that can tell for itself when it is right or wrong

This Course

- We will use the RL MOOC for this course
- In-class time will be spent on
 - Group discussions
 - Some free-form question and answer sessions
 - Some lectures (and demos) on additional material
 - Lab sessions with TAs

This class is full!

- What to do??

Instruction Team

- Prof: Adam White
- TAs (grad students doing research in AI)
 - Jordan
 - Vincent

Contacting us

- Use slack
 - You have already been invited!
 - Email the TAs if you need to be added
 - This will be the primary way we will broadcast information to the class
- Meeting w/profs and TAs during office hours

Course Information

- Github pages with updated schedule:
<https://amw8.github.io/grad-rl-one/>
- Coursera RL Mooc (our own private session)
- Course eClass page
 - some official information, and place to submit work (e.g., course projects)
 - Practice exams
 - Slack announcements

Prerequisites

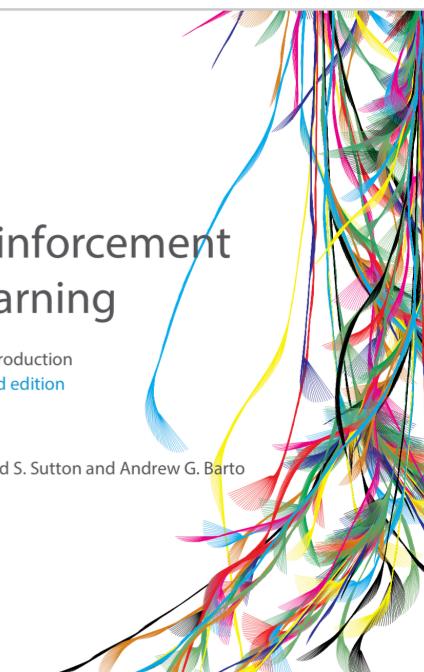
- Comfort or interest in thinking abstractly and with mathematics
- Elementary statistics, probability theory
 - conditional expectations of random variables
 - there will be a lab session devoted to a tutorial review of basic probability
- Basic linear algebra: vectors, vector equations, gradients
- Programming skills (Python)
 - We assume you can program reasonably well

Textbooks

Reinforcement
Learning

An Introduction
second edition

Richard S. Sutton and Andrew G. Barto



- Readings will be from: Reinforcement Learning: An Introduction, by R Sutton and A Barto, MIT Press.
 - available freely online
- RL Experiments Cookbook:
 - <https://sites.ualberta.ca/~amw8/cookbook.pdf>

Registering for RL on Coursera

- We have our own private session, on Coursera
- You should already be registered. Check your email
- Quick start doc for this class:
 - [https://docs.google.com/document/d/
1EnzjDA4G5NmGxnYPibvapx0CNZNEbD9hMWLCAqQw
Rg/edit?usp=sharing](https://docs.google.com/document/d/1EnzjDA4G5NmGxnYPibvapx0CNZNEbD9hMWLCAqQwRg/edit?usp=sharing)
- **It is NOT OK to use the public session. You will get ZERO for everything you submit in the public session. Lecture next week will discuss this in detail!**

Evaluation

- Assignments/Quizzes (completed in Coursera) – 30%
- In-class Participation (discussion topics) - 10%
- Midterm – 15%
- Project draft - 10%
- Final project – 35%

Weekly Quizzes and Assignments

- Each week we will cover TWO different modules, with TWO associated quizzes and/or notebooks:
 - If the deadlines on coursera don't match the class schedule (gSheet, then tell us!!)
- In preparation for class, on your own you need to:
 - Watch the lectures online (at most 1 hour of time)
 - Complete the TWO quizzes/assignments (~2-3 hrs)
- **You get one time this term that you don't have to submit notebook/quiz: ie you are marked out of 12-1=11 submissions**

Deadlines for Quizzes and Assignments

- We start Course 1 Module 2 (Sequential decision-making) Next week (Monday, September 5)
- This means by the end of the day, Sunday, September 4
 - OPTIONAL: complete Practice Quiz on Coursera
 - MUST: submit a Discussion Question on Slido

Marking participation

- Submit Discussion Questions for class via Slido
 - It informs lecture topics and in-class discussion
- **You get two times a term that you don't have to submit discussion questions: ie you are marked out of 12-2=10 submissions**

Project

- You can pick a project from a list of projects we provide
- There will be:
 - a short (unmarked) presentation
 - A draft
 - Final report

Attendance

- Lecture will be in person in ED 276
 - We will also record lecture for offline viewing
- Midterm will be in person: contact me early if you are sick and need to do it online
- This course is best when we are all here and in person and interacting

Grades are not based on a curve

- We will provide cut-offs at the end of the course
- You can see your approximate ranking in eClass throughout the course
- Letter grades are provided by 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

Collaboration

- Working together to solve the problems is encouraged
- But you must write-up / CODE your answers individually
- 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.

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

The coming of artificial intelligence

- When people finally come to understand the principles of intelligence—what it is and how it works—well enough to design and create beings as intelligent as ourselves
- A fundamental goal for science, engineering, the humanities, ...for all mankind
- It will change the way we work and play, our sense of self, life, and death, the goals we set for ourselves and for our societies
- It will lead to new beings and new ways of being, things inevitably much more powerful than our current selves

Discuss with your classmates:

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

If AI is possible, then will it eventually, inevitably happen?

- No. Not if we destroy ourselves first
- If that doesn't happen, then there will be strong, multi-incremental economic incentives pushing inexorably towards human and super-human AI
- It seems unlikely that they could be resisted
 - or successfully forbidden or controlled
 - there is too much value, too many independent actors

Investment in AI is way up

- Google's prescient AI buying spree: Boston Dynamics, Nest, Deepmind Technologies, ...
- Newish AI research labs at Facebook, Baidu, Allen Institute, Vicarious, Maluuba, DeepMind Alberta...
- Also enlarged corporate AI labs: Microsoft, Amazon, Adobe...
- Yahoo makes major investment in CMU machine learning department
- Many new AI startups getting venture capital
- New Canadian AI funding in Toronto, Montreal, and Edmonton
 - The Alberta Machine Intelligence Institute (AMII)

Advances in AI abilities are coming faster:

- IBM's Watson beats the best human players of Jeopardy! (2011)
- Deep neural networks greatly improve the state of the art in speech recognition and computer vision (2012–)
- Google's self-driving car becomes a plausible reality (\approx 2013)
- Deepmind's DQN learns to play Atari games at the human level, from pixels, with no game-specific knowledge (\approx 2014, Nature)
- University of Alberta program solves Limit Poker (2015, Science), and then defeats professional players at No-limit Poker (2017, Science)
- Google Deepmind's AlphaGo defeats legendary Go player Lee Sedol (2016, Nature), and world champion Ke Jie (2017), vastly improving over all previous programs
- Large language models, playing games by watching Youtube, ...

RL + Deep Learning Performance on Atari Games



Space Invaders



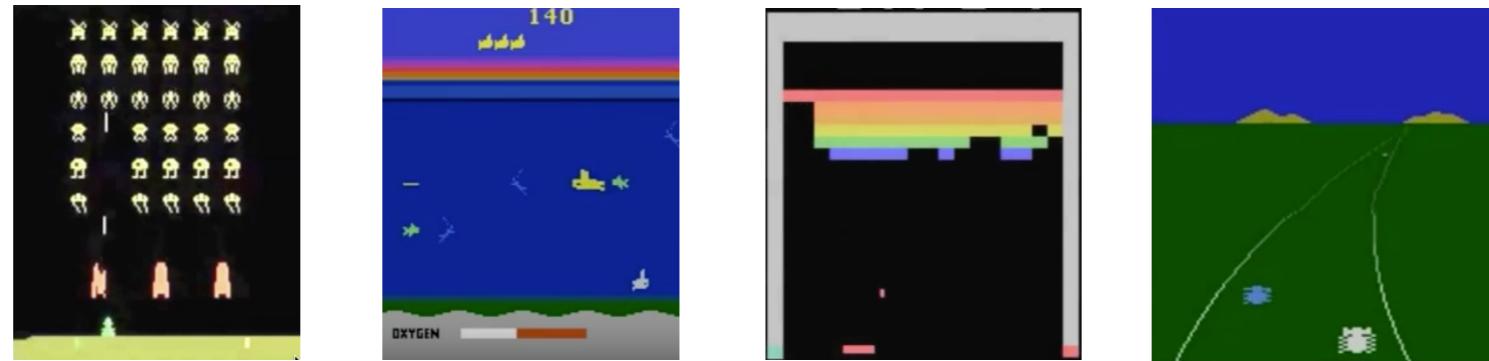
Breakout



Enduro

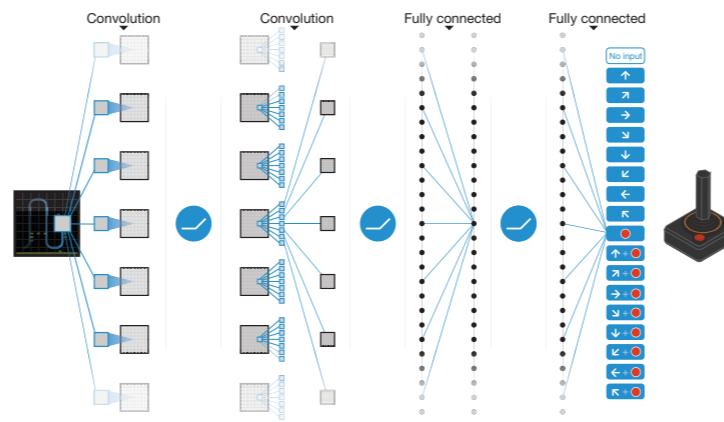
RL + Deep Learning, applied to Classic Atari Games

Google Deepmind 2015, Bowling et al. 2012



- Learned to play 49 games for the Atari 2600 game console, without labels or human input, from self-play and the score alone

mapping raw screen pixels



to predictions of final score for each of 18 joystick actions

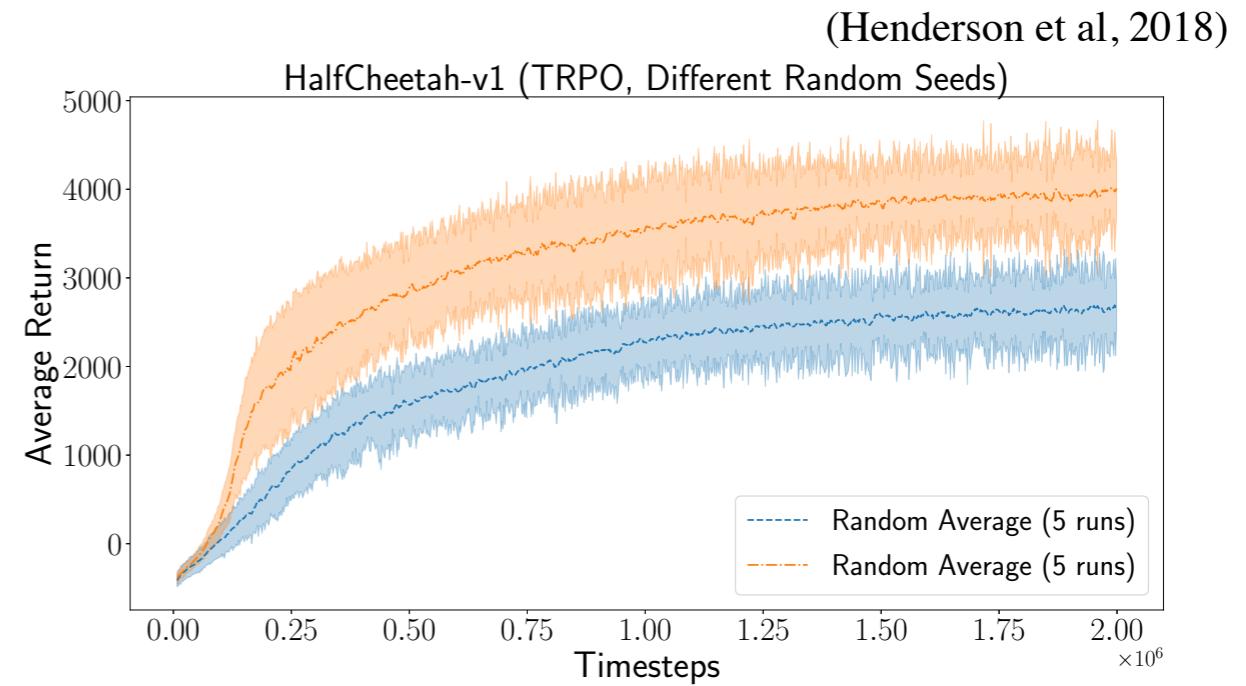
- Learned to play better than all previous algorithms and at human level for more than half the games

Same learning algorithm applied to all 49 games! w/o human tuning

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



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)

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 $t = -9.0916$, $p = 0.0016$.

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
- Borealis AI
- Deepmind Alberta
- Amii
- Several new labs and startups on the horizon

Discuss with your classmates:

Why are you here?

- What do you expect to learn?

Discuss with your classmates

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?

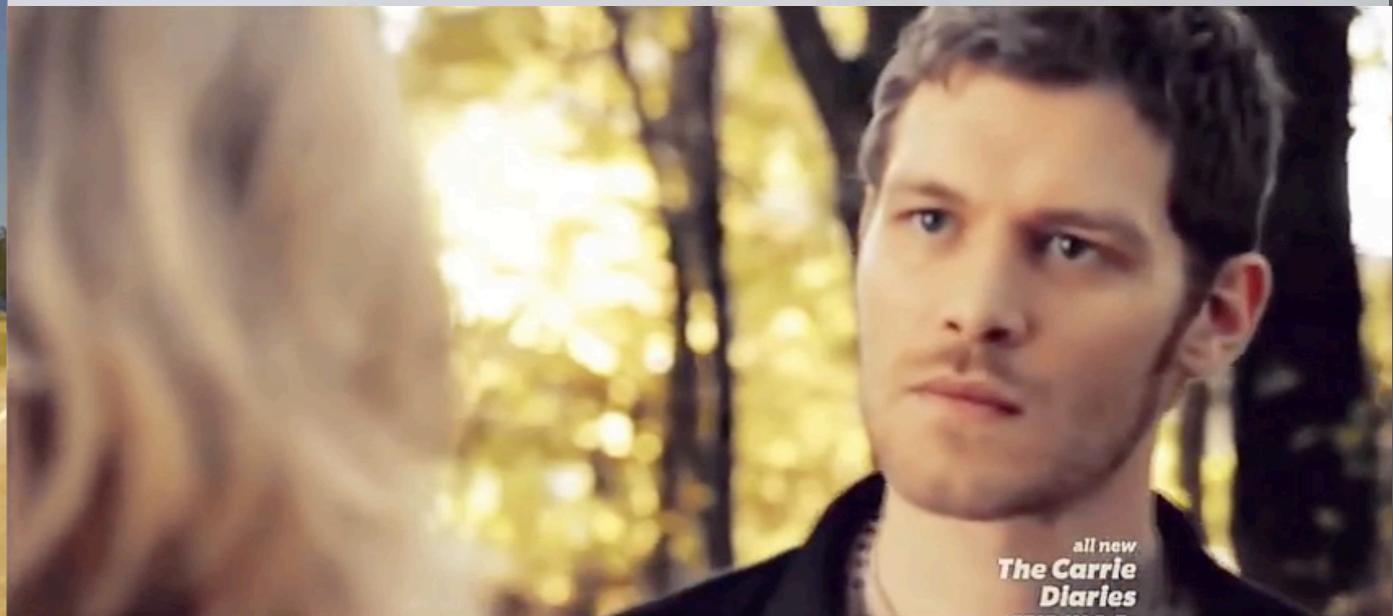
For you, which of the following are essential abilities of an intelligent system that you would like to learn about (say in this course)?

The ability to:

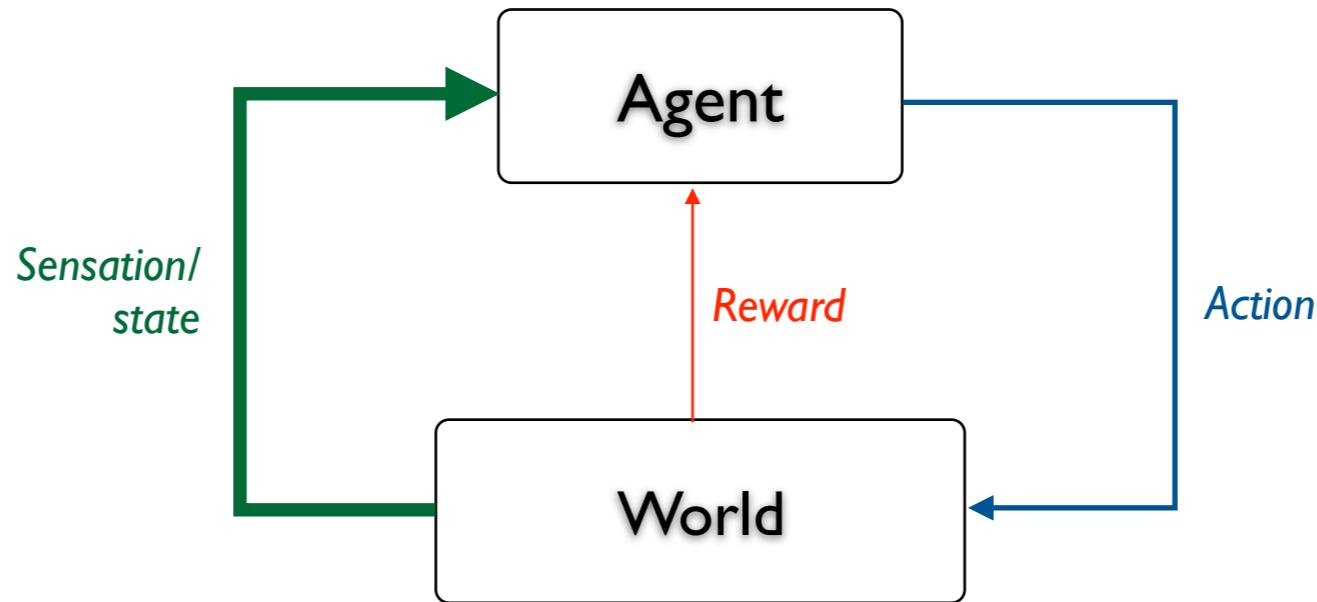
- A.sense and perceive the external world
- B.choose actions that affect the world
- C.use language and interact with other agents
- D.predict the future
- E. fool people into thinking that you are a person
- F. have and achieve goals
- G.reason symbolically, as in logic and mathematics
- H.reason in 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?

the mind's first responsibility is
real-time sensorimotor information processing

- Perception, action, & anticipation
 - as fast and reactive as possible



Reinforcement learning is ***more*** autonomous learning



- Learning that requires less input from people
- AI that can learn for itself, during its normal operation

Kinds of Reinforcement Learning

- ***Model-free RL*** — learning what to do by trying different things, remembering the best
- ***Model-based RL*** — learning how the world works, then computing what to do
- ***Prediction learning*** — learning what will happen next
- ***Representation learning*** — learning the features of state that generalize well
- ***RL architectures*** — putting it all together with massive computation

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

Order of Presentation

- Control: Bandits and Markov decision processes
- Stochastic planning (dynamic programming)
- Model-free reinforcement learning
- Planning with a learned model
- Learning with approximations

Schedule

- Week-by-week schedule on github
 - Includes topics
 - Includes assignments and deliverables

High-level view

- Bandits and online learning (ch2):
 - formalizing a problem and discussing solution methods
 - A miniature version of the entire course
- Markov Decision Processes (ch3):
 - Our formalization of reinforcement learning and AI...no solution methods here
 - Students usually get impatient here

High-level view (2)

- Classic MDP solution methods (ch's 4,5,6):
 - Dynamic programming (what if you knew how the world worked?)
 - Monte Carlo (what if you only learned from interaction)
 - Temporal difference learning (strengths of both)
- More advanced stuff:
 - planning with learned models

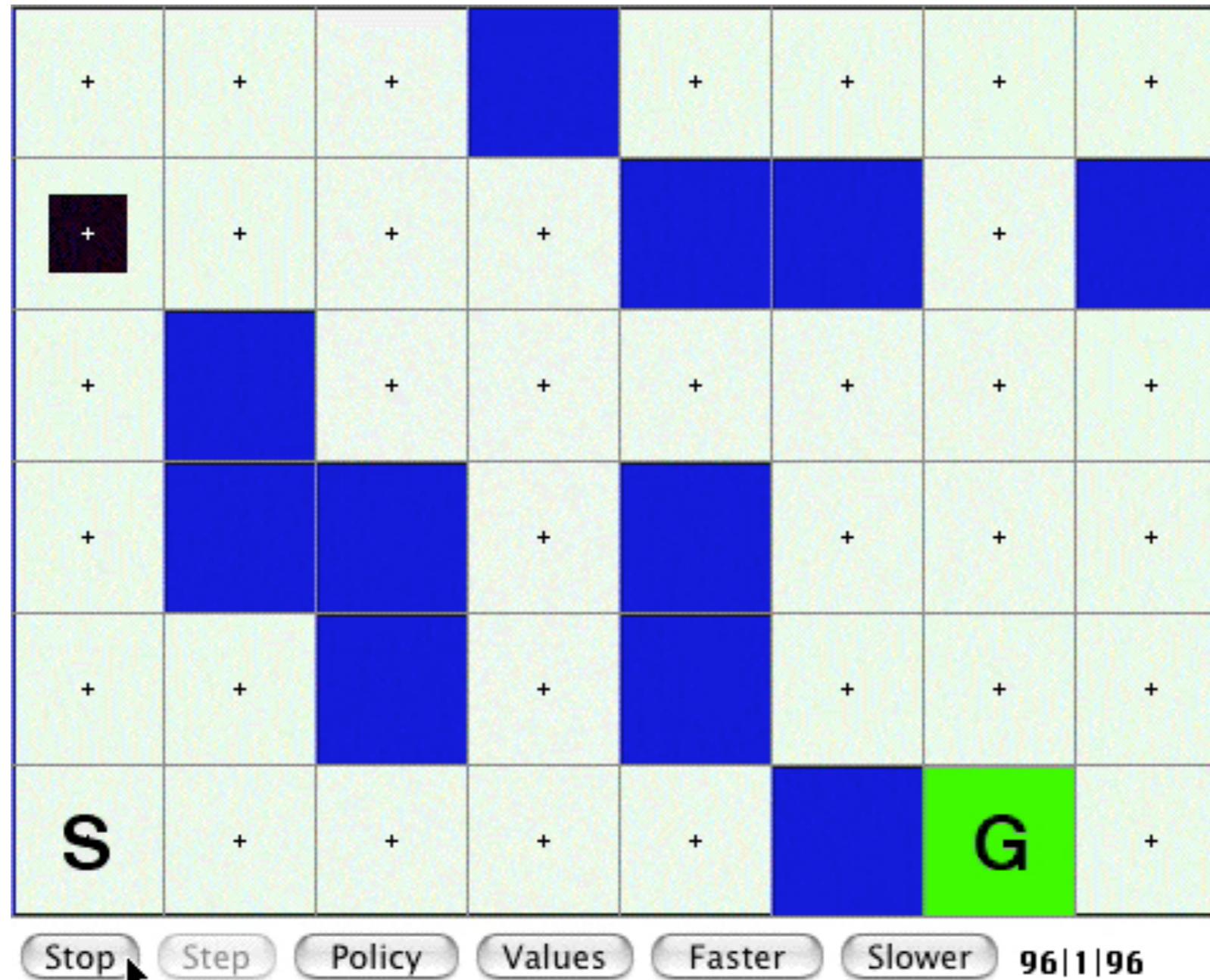
High-level view (3)

- 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

The dimensions of RL

- Problems
 - Prediction and control
 - MDPs
 - Solutions
 - Bootstrapping
 - Tabular methods
 - On-policy reinforcement learning
 - Model-based reinforcement learning
 - Value function approximation
 - Primitive actions and temporal abstraction
- We have many of the pieces already:*
- ... but there is so much left to be done :)*

Model-based RL: GridWorld Example





UAlberta RLAI Group

Thanks for listening

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>