CS 6756: Learning for Robot Decision Making

Sanjiban Choudhury
WHAT A TIME TO BE ALIVE!
Exciting time for Artificial Intelligence

Deep Q Networks

Playing Atari with Deep Reinforcement Learning

AlphaGo

Transformers

2013

2016

Today
Where are the robots?
Robots are not far behind!
Robots are not far behind!

Self-driving companies going driverless ...
Robots are not far behind!

Boston Dynamics are starting to sell their robots ...
Robots are not far behind!

Drones are getting more reliable ...
But ...

... robots are not in millions of homes yet.

Why?
Why are robots not in millions of homes yet?

PollEv.com/sc2582
The way we program robots today is … rigid!

Engineers hand-craft behaviors

Ship robot

Choose option
1. Start
2. Clean
3. Stop

Frustrate users!

Not flexible enough to be used by everyday users for everyday tasks
This restricts robots to a CLOSED world
How can we get robots out of the factory into the OPEN WORLD?
Robotics 2.0: Scale and improve with data

Formulate as a learning problem

Invest in good ML pipelines
Robotics 2.0: Scale and improve with data

Formulate as a learning problem

Invest in good ML pipelines

Self-driving led the way!
What is special about learning for robot decision making?

Standard learning

\[
\min_{\theta} \mathbb{E}_{x,y} \ell(y, \theta(x))
\]

\(x\) is a sequence of inputs, \(y\) is a sequence of outputs, \(\theta\) is a model

In decision making:

\(x\) is the sequence of observations
\(y\) is the sequence of decisions (plan)
What is special about learning for robot decision making?

\[
\min_{\theta} \mathbb{E}_{x,y} \ell(y, \theta(x))
\]

\(x\) is a sequence of inputs, \(y\) is a sequence of outputs, \(\theta\) is a model

Transformers are pretty standard choice for the model.
What is special about learning for robot decision making?

$$\min_{\theta} \mathbb{E}_{x,y} \ell(y, \theta(x))$$

$x$ is a sequence of inputs, $y$ is a sequence of outputs, $\theta$ is a model

Problem 1: What’s special about the data?
**What is special about learning for robot decision making?**

\[
\min_{\theta} \mathbb{E}_{x,y} \ell(y, \theta(x))
\]

- $x$ is a sequence of inputs,
- $y$ is a sequence of outputs,
- $\theta$ is a model

**Problem 2: What’s special about the loss?**
WHY this course?

Formulate as a Markov Decision Problem (MDP)

Analyze and Solve MDPs
(uniﬁed framework + algorithmic toolkit)

Take any robot application

Develop a uniﬁed framework
(that ties old and new ideas)
Belonging
The Crew
"Sanjiban"
He / Him

Office hours:
Tues 11:30 – 1:30pm
Gates 413B

Build robots that can learn from humans!

Undergrad

Research Engineer

PhD

PostDoc
Kushal Kedia, PhD Student

Research Interests: Forecasting, Imitation Learning

Fun Fact about me: I love collecting merchandise for my favourite sports team, Chelsea! Let’s chat about soccer :)

Office hours:
Thursday 12:30 – 2:30pm,
Rhodes 402
We are PoRTaL (People and Robots, Teaching and Learning)

https://portal.cs.cornell.edu/
Everyday Robots for Everyday Users
HAL & DORA
Helping Out In the Kitchen
Let’s get started!
How should robots learn to make good decisions?
Grounded in two “personal” applications

Self-driving

Home Robots
Self-Driving
Activity!
Brainstorm: What is “good” behavior in a left turn?
Brainstorm: What is “good” behavior in a left turn?
How should robots learn to make good decisions?
Three fundamental questions
Values

What are good / bad states?
Models
How do decisions affect states?

Values
What are good / bad states?
Optimization
How do we efficiently find the optimal sequence of decisions?

Models
How do decisions affect states?

Values
What are good / bad states?
Optimization

How do we efficiently find the optimal sequence of decisions?

Models

How do decisions affect states?

Values

What are good / bad states?

Learning
5 Levels of Robot Learning
Values
What are good / bad states?

Models
How do decisions affect states?

Optimization
How do we efficiently find the optimal sequence of decisions?
What are good / bad states?

Bad
- Collision
- Cutting off pedestrians
- Cutting off oncoming car
- Getting stuck in intersection when light turns red
- Excessive braking / braking speed limit

Good
- Completing the turn quickly
Question:

How do we program in these values?
Why don’t we simply imitate good human driving?
SUPERVISED LEARNING

#1 Get Expert Data

#2 Train Policy

\[ \pi : s \rightarrow a \]

#3 Deploy!
Train ≠ Test
Lesson #1
Feedback drives covariate shift
Models

How do decisions affect states?
Optimization
How do we efficiently find the optimal sequence of decisions?

Models
How do decisions affect states?

Values
What are good / bad states?
Activity!
DORA Reacts to Human Forecasts
Think-Pair-Share

Think (30 sec): How do we train a model of how humans move? Data? Model? Loss?

Pair: Find a partner

Share (45 sec): Partners exchange ideas
Lesson #2

Models are useful fictions
Optimization

How do we efficiently find the optimal sequence of decisions?

Models

How do decisions affect states?

Values

What are good / bad states?
High-dimensional, continuous trajectory optimization
(With hard constraints!)
The journey ahead!
<table>
<thead>
<tr>
<th>Date</th>
<th>Lecture</th>
<th>Preread</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>08/22/23</td>
<td>Introduction: How should robots learn to make good decisions?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>08/24/23</td>
<td>Interactive Online Learning</td>
<td>Shai Shalev-Shwartz (Pg.108-111)</td>
<td>Arora et al. &quot;Multiplicative Weights&quot;, Generalized Weighted Majority video</td>
</tr>
<tr>
<td>08/29/23</td>
<td>Markov Decision Process</td>
<td>MACRL Ch. 1</td>
<td>Dan Klein's slides I</td>
</tr>
<tr>
<td>08/31/23</td>
<td>Linear Quadratic Regulator: The Analytic MDP</td>
<td>MACRL (Ch 2, Pg. 23-27)</td>
<td>Underactuated robotics, Ch. 8, History of Optimal Control</td>
</tr>
<tr>
<td>09/05/23</td>
<td>Iterative Linear Quadratic Regulator</td>
<td>MACRL (Ch 2, Pg. 28-33)</td>
<td>iLQR paper, DDP for helicopter flight</td>
</tr>
<tr>
<td>09/07/23</td>
<td>Solving Hard MDPs: Constraints, Long Horizons, and more!</td>
<td>MACRL (Ch 4)</td>
<td>Gordon's notes on Lagrange, ALTRO: AuLa + iLQR,</td>
</tr>
<tr>
<td>09/12/23</td>
<td>Imitation Learning: Feedback and Covariate Shift (Assignment 2 Released)</td>
<td>MACRL (Ch 6, Pg. 53-57)</td>
<td>Three regimes of covariate shift</td>
</tr>
<tr>
<td>09/14/23</td>
<td>DAgger: A Reduction to No-Regret Learning</td>
<td>MACRL (Ch 6, full)</td>
<td>DAGGER, Agnostic SysId</td>
</tr>
<tr>
<td>09/19/23</td>
<td>Imitation Learning as Inferring Latent Expert Values</td>
<td></td>
<td>EIL, Youtube lec</td>
</tr>
<tr>
<td>09/21/23</td>
<td>Inverse Reinforcement Learning: From Maximum Margin to Maximum Entropy</td>
<td>MACRL (Ch 7)</td>
<td>LEARCH, MaxEntIRL, Youtube lec</td>
</tr>
<tr>
<td>09/26/23</td>
<td>Scaling Inverse Reinforcement Learning</td>
<td></td>
<td>Guided cost learning, f-divergence IL, Youtube lec</td>
</tr>
<tr>
<td>09/28/23</td>
<td>Imitation Learning: The Big Picture</td>
<td></td>
<td>Of Moments and Matching, Youtube lec</td>
</tr>
<tr>
<td>Date</td>
<td>Topic</td>
<td>MACRL References</td>
<td>Other References</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>10/03/23</td>
<td>Approximate Dynamic Programming : Temporal Difference, Q-learning (Assignment 3 Released)</td>
<td>MACRL (Ch 8, full), MACRL (Ch 9, full)</td>
<td>Sutton&amp;Barto (Ch. 5, 6), DQN, Rainbow DQN</td>
</tr>
<tr>
<td>10/05/23</td>
<td>Black-box vs White-box Policy Optimization</td>
<td>MACRL (Ch 10, full)</td>
<td></td>
</tr>
<tr>
<td>10/12/23</td>
<td>Nightmares of Policy Optimization</td>
<td>MACRL (Ch 11, full)</td>
<td></td>
</tr>
<tr>
<td>10/17/23</td>
<td>Actor Critic Methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/19/23</td>
<td>Model-based Reinforcement Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/24/23</td>
<td>Dealing with Uncertainty (Extended Abstracts Due)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Frontiers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/26/23</td>
<td>Meta Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/31/23</td>
<td>Offline Reinforcement Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/02/23</td>
<td>Diffusion Models and Imitation Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/07/23</td>
<td>Large Language Models and Task Planning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/09/23</td>
<td>Multi-agent Forecasting and Imitation Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/14/23</td>
<td>Learning Visual Representations from Ego Videos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/16/23</td>
<td>Visuomotor Skill Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/21/23</td>
<td>Causal Representation Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/28/23</td>
<td>Project presentations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/30/23</td>
<td>Project presentations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/04/23</td>
<td>Recap Lecture</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Logistics
Logistics

Website: https://www.cs.cornell.edu/courses/cs6756/2023fa/

Lectures
Interactive lectures, please read assigned book chapters / papers

Assignments [3 assignments * 15% grade = 45%]
Programming heavy. HW2, HW3 involve PyTorch. Done individually!

Project [45%]
Final project. Pick a research problem, apply techniques from class. Be creative!
Groups of 2. Extended abstract, final presentation, final paper.

Participation [10%]
Activities in lectures, in class polls
Book!

https://macrl-book.github.io/

Modern Adaptive Control and Reinforcement Learning, James A. Bagnell, Byron Boots, and Sanjiban Choudhury

(Please send me feedback)
Expectations

Assignments are programming / ML heavy!

For final project: Check if you have the resources to train models

Familiar with ML concepts and modern tools (transformers, CNNs). Also familiarity with linear algebra (SVD etc)

Please pre-read book chapters / pre-watch supplementary video
Which course should you take?

CS 4756 / 5756 (Spring!)
More focus on fundamentals, also covers robot perception.
Good course to take before taking this one.

CS 6756 (this class)
Builds on CS 4756, goes deeper in decision making,
assumes familiarity with ML tools and concepts
Generative AI

The work you do for CS 6756 consists of writing code and natural language descriptions.

To some extent, the new crop of “generative AI” (GAI) tools can do both of these things for you.

However, we require that the vast majority of the intellectual work must be originated by you, not by GAI. You may use GAI to look up helper functions, or to proofread your text, but clearly document how you used it.
Generative AI

In this class, for every assignment and final project, you can choose between two options:

**Option 1: Avoid all GAI tools.** Disable GitHub Copilot in your editor, do not ask chatbots any questions related to the assignment, etc. If you choose this option, you have nothing more to do.

**Option 2: Use GAI tools with caution** and include a one-paragraph description of everything you used them for along with your writeup. This paragraph must:

1. Link to exactly which tools you used and describe how you used each of them, for which parts of the work.
2. Give at least one concrete example (e.g., generated code or Q&A output) that you think is particularly illustrative of the “help” you got from the tool.
3. Describe any times when the tool was unhelpful, especially if it was wrong in a particularly hilarious way.
4. Conclude with your current opinion about the strengths and weaknesses of the tools you used for real-world compiler implementation.

Remember that you can pick whether to use GAI tools for every assignment, so using them on one set of tasks doesn’t mean you have to keep using them forever.
Assignment 0

Simple survey

Link in the website:
https://docs.google.com/forms/d/e/1FAIpQLSdMMyzh1o3hA0j4S1sDEjnN91vjcxRBPaeTaoVw82apMCDyg/viewform?usp=sf_link
Questions?

PINs issued every Tues / Thursday

Plenty of space in class so waitlist should get cleared
**tl;dr**

**How should robots **learn** to make **good** decisions?**

---

**Optimization**
How do we efficiently find the optimal sequence of decisions?

**Models**
How do decisions affect states?

**Values**
What are good / bad states?

---

**WHY this course?**

- **Formulate as a Markov Decision Problem (MDP)**
- **Analyze and Solve MDPs** (unified framework + algorithmic toolkit)
- **Develop a unified framework** (that ties old and new ideas)

- **Take any robot application**