ECE 457C Reinforcement Learning

Spring

Instructor information

Name	
Contact Info	
Office location	
Office hours	

TA Information [If applicable]

TA name	
TA Contact Info	
Office location	
Office hours	

Course Description

The field of Artificial Intelligence intersects many areas of knowledge which engineers can utilize for building robust, dynamic systems in a world filled with large amounts of data yet also containing uncertainty and hidden information. In this course we focus from the ground up on the concepts and skills needed to build systems that can reason, learn and make decisions under uncertainty using a type of Machine Learning called Reinforcement Learning. This begins with reviewing concepts from Bayesian probability and learning how to perform probabilistic inference with exact and more efficient approximate methods such as MCMC for simple decision problems called Multi-armed Bandits.

Reinforcement Learning (RL) is a general framework for decision making where agents learn how to act from their environment without any prior knowledge of how the world works or possible outcomes. We will explore the classic definitions and algorithms for RL and see how it has been revolutionized in recent years through the use of Deep Learning. Recently, impressive AI algorithms have been demonstrated which combine all of these concepts along with Monte-Carlo Tree Search to learn to

play video games (such as Star Craft) and board games (such as Go and chess) from scratch. More practical applications of these methods are used regularly in areas such as customer behaviour modelling, traffic control, automatic server configuration, autonomous driving and robotics.

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.

Learning Objectives

This course complements other AI courses in ECE by focusing 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.

- Explain and apply the basics methods of Bayesian modelling including inference
- Explain and apply the basic methods of Bayesian Optimization to specific problems
- Explain, evaluate and implement Reinforcement Learning algorithms for given problem descriptions

Topics Covered by Week

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 and Value Function approximation
8. Basics of Neural Networks
   - fully connected, multi-layer perceptrons
   - supervised training, back-propagation
   - stochastic gradient descent
   - regularization methods
9. Deep Reinforcement Learning
   - Deep Q- Networks (DQN)
   - Experience replay buffers and mini-batch training
   - A3C, A2C, DDPG
10. Other Challenges
    - Partially Observable MDPs (POMDPs)
    - Multi-Agent RL (MARL)
11. Other ways to solve (PO)MDPs
    - Monte-Carlo Tree Search, Explaining AlphaGo
    - Curiosity based learning
12. Wrap-up and Review

Assessments

- Assignment 1: 7.5% - 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: 7.5% - Implement RL algorithms for simple domain, SARSA, Q-Learning, Eligibility Traces.
  - *Evaluation: Short report with graphs, automated grading and code review.*

- Midterm: 30%

- Assignment 3: 7.5% - Implement Policy Gradient, Actor Critic, Value Function Approximations for larger RL domains.
  - *Evaluation: Short report with graphs and code review.*

- Assignment 4: 7.5% - 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: 40% - all topics, leaning towards latter half of course

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.

See http://www.uwaterloo.ca/academicintegrity/

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. When in doubt please be certain to contact the departments administrative assistant who will provide further assistance.

See http://www.adm.uwaterloo.ca/infosec/ Policies/policy70.htm

Discipline: A student is expected to know what constitutes academic integrity, to avoid committing an academic offence, and to take responsibility for his/her actions.

See http://www.uwaterloo.ca/academicintegrity/

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.

See http://www.adm.uwaterloo.ca/infosec/Policies/policy71.htm.

For typical penalties check Guidelines for the Assessment of Penalties.

See 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)

See http://www.adm.uwaterloo.ca/infosec/Policies/policy72.htm.

Note for Students with Disabilities: The Office for Persons with Dis-abilities (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.