Course Syllabus

Course Description

This is a **graduate level** class on dynamic decision making and reinforcement learning (RL). RL is one of the three basic machine learning paradigms, alongside supervised learning and unsupervised learning. RL concerns with how an "agent" aims to maximize its cumulative reward by dynamically interacting with the "environment", where the focus is to balance between the *exploration* of uncharted territory and the *exploitation* of current knowledge.
The goal of this class is to teach theories and solution methods of RL in the context of *industrial engineering*, *operations research*, and *operations management*. Unlike most extant RL textbooks which often use examples arising from computer science, robotics, psychology and physics, in this class we will discuss how to apply RL methods to solve IE and OR problems including: machine maintenance and replacement, resource allocation, investment, gambling (e.g., blackjack, two slot machine problem), pricing an American option, inventory control, revenue management, queueing control, and multi-armed bandits problems.

This course (a.k.a. "**stochastic models for systems analytics**") is the second course of the ISE department's **system analytics** series:

- Optimization models for systems analytics (Prof. Shu-Cherng Fang)
- **Stochastic models for systems analytics (this course)**
- Statistical models for systems analytics (Prof. Xiaolei Fang)
- Simulation models for systems analytics (Prof. Hong Wan)

This special topics course is designed mainly for Ph.D. students with proper background, research interests and self-learning capabilities.

Prerequisites

This course is intended for graduate students in industrial engineering, operations research and related fields. Student are expected to

- have completed a first course on stochastic models and optimization at the level of the first-year doctoral courses ISE760 and ISE505;
- have completed a graduate-level course on computer simulation such as ISE762;
- be comfortable with implementing RL algorithms using a programming language such as MatLab, Python, etc.

Instructor

Yunan Liu
Online Office hours: Monday/Wednesday 4:15–5:15
Office hours Zoom link: https://ncsu.zoom.us/j/4711428118
Email: *yliu*48@*ncsu.edu*
Homepage: *http://yunanliu.wordpress.ncsu.edu*

Time, Place and Delivery Method

Monday and Wednesday 3:00–4:15. Online delivery.

Class Zoom link: https://ncsu.zoom.us/j/4711428118

Teaching Assistant

Yining Huang
Office hours: 9:00–10:00 Tuesday/Thursday.
Personal Zoom link: https://ncsu.zoom.us/j/3637028099
Email: *yhuang*43@*ncsu.edu*

Textbooks (recommended)

(i) S. M. Ross, *Introduction to Stochastic Dynamic Programming.* Academic Press, 2014.

(ii) R. Sutton and A. Barto, *Reinforcement Learning: An Introduction.* 2nd Edition, Bradford Books, 2018.

(iii) C. Szepesvári, *Algorithms for Reinforcement Learning.* Morgan & Claypool, 2010.

(iv) D. Bertsekas, *Reinforcement Learning and Optimal Control.* Athena Scientific, 2019.

Homework

There will be biweekly homework assignments. (No homework, no learning!)

- Students are encouraged to collaborate with other students in the class, as long as each person writes his/her own solutions and codes. But any such collaboration should be clearly **noted** (If some ideas of your solutions come from the discussion with another person, write his/her name on your solution).
- Copying homework from another student (past or present) is **forbidden**.
- Single **PDF file** of scanned homework solutions should be **submitted at the class moodle site** before the deadline. PDF files may be compiled using mobile apps such as "camscanner" (Do NOT submit photos of homework solutions to moodle).
- Late homework will **NOT** be accepted.
- Homework solutions will be posted at the moodle site.

Exams

All exams are open notes (closed everything else: HW solutions, textbooks, googling, etc).

- Midterm: TBD (in-class, online proctoring).
- Final: TBD (possibly takehome).

Project

The group project has both modeling and coding components. Each group will be composed of at most two students and will be responsible for

- choosing a topic (after the midterm);
- submitting a project report (by the last day of class);
- giving a project presentation (during the last 2 weeks of class).

Grading

Define the following random variables:
$HW \equiv$ homework, $M \equiv$ midterm, $F \equiv$ final exam, $Pr \equiv$ project, $Pa \equiv$ in-class participation and $G \equiv$ overall grade.
Then the overall grade is given by

$$G \equiv HW \times 25\% + M \times 25\% + F \times 25\% + Pr \times 20\% + Pa \times 5\%.$$

Tentative Course Topics

The course topics include:

0. Introduction & preliminaries
    - Intro to RL
        - What's RL?
        - Main applications of RL
        - RL vs. other machine learning areas
    - Preliminaries
        - Conditional probability and Bayesian updating
        - Law of total probabilities and expectations
        - Convexity and concavity
        - Jensens inequality
        - Fixed-point equations and contraction mapping
        - Stochastic ordering
        - Markov chain and Markov decision process (MDP)
1. Finite-time MDP
    - Optimality equation
    - Examples
        - Gambling problem
        - Exercising an American option
        - Machine replacement
        - Searching for the best candidate (the famous 37% rule)
        - Sequential resource allocation
        - Slot machine problem

  - Inventory management
- Extensions and overview of the class

2. Discounted-reward infinite-time MDP
   - Policy evaluation
     - Bellman equation: existence and uniqueness
     - Iterative algorithm and convergence
   - Optimality equation
     - Contraction mapping representation
   - Policy improvement
   - Solution techniques:
     - Value iteration
     - Policy iteration
     - Linear programming
     - Examples: machine replacement, the rental car problem
     - Further discussions
3. Undiscounted-reward infinite-time MDP
   - Maximizing rewards
     - Optimality equation
     - Solution methods
     - Example: gambler's problem
   - Minimizing costs
     - Optimality equation
     - Solution methods
   - Optimal stopping
     - one-step look-ahead policy
     - Examples: Burglar problem, selling an asset, Baysian sequential problem.
4. Averaged-reward infinite-time MDP
   - Policy evaluation
   - Optimality equation (AR version)
   - Average-reward MDP vs. discounted MDP and undiscounted MDP
   - Solution techniques:
     - reduction to discounted MDP
     - policy iteration
     - value iteration
     - linear programming (state action frequency)
   - Examples: machine maintenance, rental car, inventory management.
5. Monte-Carlo (MC) methods

- MC simulation
- MC policy evaluation
  - first-visit and every-visit methods
  - Examples: gambler's problem, exercising American option, blackjack
- MC policy iteration
  - MC estimation of action values
  - On-policy methods
    - exploring start
    - ϵ-greedy
  - Off-policy methods
    - Importance sampling
    - IS-based off-policy control
- Rollout algorithm and MC tree search
- MC vs. DP

6. Temporal-difference (TD) methods

   - TD(0): one-step TD policy evaluation
     - Stochastic approximation and convergence of TD(0)
     - Examples: Gambler's problem, inventory management, rental car problem
   - TD policy iteration
     - SARSA: TD-based control
     - Q-learning
     - expected TD
   - Multi-step TD
     - Policy evaluation: TD(n)
     - On-policy learning: n-step SARSA
     - Off-policy learning: n-step tree backup
   - Comparing TD, MC and DP
   - Categorizing modern RL methods

7. Function approximation (unifying RL and supervised learning)

   - FA policy evaluation
     - Mean squared value error (MSVE)
     - Stochastic gradient descent (SGD)
     - Linear function approximation
     - Nonlinear function approximation: deep Q-learning network (DQN)
   - FA policy iteration
     - FA-SARSA
   - Policy gradient method
   - Actor critic method

8. Multi-armed bandits
    - Algorithms
    - Regret analysis
    - Extensions
9. Project Presentations (last week of class)
    - Each group will prepare and give a 15 min presentation
    - Group information (tentative title and group members) should be sent to instructor and TA right after the midterm
    - Presentation slides should be sent to instructor and TA before last week of class