CSE 691 COURSE DESCRIPTION FOR SPRING SEMESTER 2023 (3 Credits)

TOPICS IN REINFORCEMENT LEARNING

ASU/SCAI

Prof. Dimitri P. Bertsekas

Wednesdays 4:30 to 7:00 ASU time

This course will focus on Reinforcement Learning (RL), a currently very active subfield of artificial intelligence, and it will discuss selectively a number of algorithmic topics couched on approximate Dynamic Programming (DP) methods: approximation in value and policy space, approximate policy iteration, rollout (a one-time form of policy iteration), model predictive control, adaptive control, multiagent methods, applications to challenging combinatorial optimization problems, implementations using simulation and neural network architectures, and engineering and artificial intelligence applications, such as the high-profile successes of the AlphaZero and TD-Gammon programs, which play chess and backgammon, respectively.

The course will center around a sequential decision architecture, called *approximation in value space*, which involves two algorithms that are designed largely independently of each other but operate in synergy. We call these the *off-line training and the on-line play algorithms*; the names are inspired by the context of games, such as AlphaZero and TD-Gammon. In the context of these games, the off-line training algorithm is the method used to teach the program how to evaluate positions and to generate good moves at any given position, while the on-line play algorithm is the method used to play in real time against human or computer opponents. A key fact is that the on-line play algorithm tends to improve substantially the performance predicted by the off-line play algorithm, and thus its proper implementation, through one-step or multistep minimization, rollout, and terminal cost approximation, is the critical part of the overall architectural design. For this reason, our course will place primary emphasis on the on-line algorithm implementation possibilities.

Starting from the AlphaZero and TD-Gammon methodology, we will aim to show that approximation in value space applies very broadly to deterministic and stochastic optimal control problems, involving both discrete and continuous search spaces, as well as finite and infinite horizon. Moreover, we will show that our conceptual framework can be effectively integrated with other important methodologies such as model predictive and adaptive control, multiagent systems and decentralized control, discrete and Bayesian optimization, and heuristic algorithms for discrete optimization.

The primary emphasis of the course is to encourage graduate research in reinforcement learning through directed reading and interactions with the instructor over zoom. Prerequisites are a full course on calculus, and a first course in probability.

The course will leverage a series of videolectures, slides, and other material from previous ASU offerings of the course, which are posted at

http://web.mit.edu/dimitrib/www/RLbook.html

All class material (videolecture recordings, slides, class notes, etc) will be posted at this website on Fridays, as well as at ASU/Canvas.

Textbooks (optional):

(1) D. Bertsekas, "Reinforcement Learning and Optimal Control," Athena Scientific, 2019.
(2) D. Bertsekas, "Rollout, Policy Iteration, and Distributed Reinforcement Learning," Athena Scientific, 2020.
(3) D. Bertsekas, "Abstract Dynamic Programming," 3rd Edition, Athena Scientific, 2022 (on-line).
(4) D. Bertsekas, "Lessons from AlphaZero from Optimal, Model Predictive, and Adaptive Control," Athena Scientific, 2022 (on-line).
(5) D. Bertsekas, On-Line Class Notes.

Supplementary material:

(1) Sutton, R., and Barto, A., "Reinforcement Learning," 2nd Edition, MIT Press, Cambridge, MA (on-line; a valuable resource that approaches the subject from the AI point of view).
(2) Research papers and monographs, available on-line.

Structure:

One 2.5-hour lecture per week by the instructor (divided in two parts with a 15 min break), except for the last lecture, which will involve research presentations by student participants. Three to four homeworks (25 percent of the grade), and a research project or term paper (75 precent of the grade). Office hours are by appointment with the instructor: dimitrib@mit.edu

Topics to be covered:

(1) Introduction to exact and approximate dynamic programming
(2) Approximation in value and policy space
(3) Off-line training, on-line play, and Newton's method
(4) Rollout and approximate policy iteration
(5) Model predictive and adaptive control
(6) Multiagent and multiprocessor reinforcement learning
(7) Training of feature-based approximation architectures and neural networks
(8) Policy networks and approximation in policy space

The first four lectures will aim to provide an introduction and overview of the subject, which will facilitate selecting and focusing on some research area. The remaining lectures will develop the topics listed above in greater depth.