Reinforcement Learning

Syllabus Information

CS 4995 - Reinforcement Learning

Associated Term: 2024/25 Academic Session
Learning Objectives:

The aim of the module is to introduce students to the theory and practice of reinforcement learning (RL), which is the current machine learning paradigm for learning to perform multi-stage tasks. The algorithmic study of reinforcement learning began in the late 1980s, building on established work in control by dynamic programming, developed in the late 1950s. During the last 8 years there has been startling progress in combining reinforcement learning with deep neural networks: most famously, 'deep RL' has been used to learn far superhuman performance in chess and go within hours, given only the rules of the game, without further human input. Practical applications of RL are starting to emerge in design optimisation, large-scale customer interaction, and robot control. Students want to learn how this is done. The course will start by introducing the standard theory of formalising control problems as Markov decision processes (MDPs), the Bellman optimality equations, and standard dynamic programming algorithms for calculating optimal policies and value functions. Next, reinforcement learning will be presented as incremental dynamic programming on finite MDPs: TD methods, Q learning, and sarsa will be covered, with convergence results. The policy gradient theorem will be presented, with motivation for policy gradient algorithms. Contextual bandit methods will be briefly covered. Next the analogous algorithms will be presented for deep RL. Monte-Carlo tree search will be presented (without convergence proofs) as a method of value-function smoothing. Finally, recent learning architectures, such as alphazero and muzero, will be presented. **Pre-requisites:** none **Learning Outcomes:** 1. understand the notion of formalising multi-stage tasks as Markov Decision Problems (MDPs), and understand and be able to implement standard dynamic programming algorithms for finding optimal policies and value functions to satisfy the Bellman equations in tabular MDPs 2. understand and be able to implement tabular reinforcement learning algorithms for learning optimal policies in tabular MDPs, including actor-critic and Q-learning. 3. understand and implement basic algorithms for contextual bandit problems 4. understand basic deep reinforcement learning methods 5. understand the current achievements and limitations of reinforcement learning algorithms, and how to assess RL performance.
Required Materials: Click here for the reading list system
Technical Requirements: The total number of notional learning hours associated with course are 150. **These will normally be broken down as follows:** 22 hour(s) of Lectures

across 11 week(s) 22 hour(s) of Laboratory classes across 11 week(s) 106 hour(s) of Guided Independent Study **Formative Assessment:** Lab activities (20 hours) - Verbal feedback In class participation (20 hours) - Verbal feedback **Summative Assessment:** Examination (60%) 2 Hours Written Assessments (40%) 30 Hours