NICTA

NICTA Course Module

Markov Decision Processes: A Decision-Theoretic Approach to Planning Under Uncertainty

By: Pascal Poupart

April 4-6, 2011

NICTA course modules build toward meeting the course requirement which is part PhD expectations at certain partner universities, and which forms a part of a NICTA Enhanced PhD program.

- Modules are designed for short intensive delivery, usually over a few days.
- A module may be coordinated by a member of academic staff from a university, by a NICTA researcher, or by a visiting researcher to NICTA or an associated university.
- Modules are designed so that coordinators do not have to commit to regular lecture times spanning an entire session.
- A typical module requires a commitment of 50 hours of work, 15-20 contact hours, reading, assignments and assessments. Such a module meets approximately one-third of the requirements for a postgraduate masters (or higher) level course.
- With the approval of Heads of School modules may accumulate to satisfy coursework requirements. Postgraduate Coordinators can advise on the precise administrative processes by which this is achieved.

Markov Decision Processes: A Decision-Theoretic Approach to Planning Under Uncertainty

Background

Over the years, Markov decision processes (MDPs) have emerged as a popular framework for sequential decision making due to its generality and principled nature. MDPs are rooted in decision theory, which combines probability theory to model uncertainty in the action effects and utility theory to quantify preferences. Today MDPs are used to optimize sequences of decisions in a wide range of domains including robotics (e.g., control), investments (e.g., portfolio management), speech technologies (e.g., spoken dialogue management), operations research (e.g., inventory management, resource allocation, dynamic pricing), assistive technologies (e.g., patient monitoring and prompting), and manufacturing (e.g., fault diagnosis). Hence this course should be of interest to a wide audience beyond researchers working in planning or decision making under uncertainty. Familiarity with probability theory and optimization at an introductory level is recommended as prior knowledge.

Coordinator: Pascal Poupart (http://www.cs.uwaterloo.ca/~ppoupart)

Course description

This course will cover the theory and practice of fully and partially observable Markov decision processes. We will first formalize Markov decision processes and go over three broad classes of offline algorithms: value iteration, policy iteration and linear programming. Then online algorithms and advanced techniques for factored and continuous MDPs will be presented. In the second half or the course, the assumption that states are fully observable will be relaxed by formalizing partially observable Markov decision processes (POMDPs). Two broad classes of offline algorithms for POMDPs will be covered: point-based value iteration and policy search. Finally, online algorithms and advanced techniques for factored, continuous and Bayes-adaptive POMDPs will be presented.

Course Outline

Textbook: Puterman, *Markov Decision Processes*, 2nd edition, MIT Press, 2005.

There will be 12 lectures of 1.5 hours each over 3 days followed by 1 assignment.

Day 1:

1. 9:00-10:30 Introduction and brief review of probability, utility and decision theory (Puterman: Chap. 1)
2. 11:00-12:30 Fully observable Markov decision processes (MDPs), value iteration and convergence properties (Puterman: Chap. 2, 3, 6.1-6.3)
3. 13:30-15:00Policy iteration and linear programming for MDPs (Puterman: Sec. 6.4, 6.5, 6.9)
4. 15:30-17:00 Factored MDPs and symbolic dynamic programming
    - J. Hoey, R. St-Aubin, A. Hu and C. Boutilier, SPUDD: Stochastic Planning using Decision Diagrams, *Proc. International Conference on Uncertainty in Artificial Intelligence (UAI)*, 1999.

Day 2:

5. 9:00-10:30 Approximate linear programming and MDPs with continuous stats and actions
    - C. Guestrin, D. Koller and R. Parr, Efficient Solution Algorithms for Factored MDPs, *Journal of Artificial Intelligence Research*, 2003.
    - B. Kveton, M. Hauskrecht, C. Guestrin. Solving Factored MDPs with Hybrid State and Action Variables. *Journal of Artificial Intelligence Research (JAIR)*, vol 27, pp. 153-201, 2006.
6. 11:00-12:30 Online MDP algorithms
    - M. Kearns, Y. Mansour and A. Ng, A sparse sampling algorithm for near-optimal planning in large Markov decision processes, *Machine Learning*, vol. 49, pp. 193-208, 2002.

7. 13:30-15:00 Partially observable Markov decision processes (POMDPs), value function properties and dynamic programming
    - L.P. Kaelbling, M. Littman and A. Cassandra, Planning and Acting in Partially Observable Domains, *Artificial Intelligence*, 1998.
8. 15:30-17:00 Point-based value iteration and factored POMDPs
    - J. Pineau, G. Gordon and S. Thrun, Anytime Point-based Approximations for large POMDPs, *Journal of Artificial Intelligence Research*, 2006.
    - P. Poupart, K.-E. Kim and D. Kim, Closing the Gap: Towards Provably Optimal POMDP solutions, *Proc. International Conference on Planning and Automated Scheduling (ICAPS)*, 2011.
    - H.S. Sim, K.-E. Kim, J.H. Kim, D.-S. Chang and M.-W. Koo, Symbolic Heuristic Search Value Iteration for Factored POMDPs, *Proc. National Conference on Artificial Intelligence (AAAI)*, pp. 1088-1093, 2008.

Day 3:

9. 9:00-10:30 Finite state controllers and policy search techniques
    - C. Amato, D. Bernstein and S. Zilberstein, Optimizing Fixed-Size Stochastic Controllers for POMDPs and Decentralized POMDPs. *Autonomous Agents and Multi-Agent Systems* 21(3): 293-320 (2010).
10. 11:00-12:30 POMDPs with continuous states, actions and observations
    - J. Porta, N. Vlassis, M. Spaan and P. Poupart, Point-Based Value Iteration for Continuous POMDPs, *Journal of Machine Learning Research*, 2006.
11. 13:30-15:00 Online POMDP algorithms
    - S. Ross, J. Pineau, S. Paquet, B. Chaib-Draa, Online Planning Algorithms for POMDPs, *Journal of Artificial Intelligence Research*, 2008.
    - D. Silver and J. Veness, Monte-Carlo Planning in Large POMDPs, *Advances in Neural Information Processing Systems*, 2010
12. 15:30-17:00 Bayesian reinforcement learning
    - P. Poupart, N. Vlassis, J. Hoey and K. Regan, An Analytic Solution to Discrete Bayesian Reinforcement Learning, *Proc. International Conference on Machine Learning (ICML)*, 2006.
    - P. Poupart and N. Vlassis, Model-based Bayesian Reinforcement Learning in Partially Observable Domains, *International Symposium on Artificial and Mathematics (ISAIM)*, 2008.

Assignment:

A. MDP (lectures 1-6)
    i. Formulate a sequential decision making problem as an MDP
    ii. Implement value iteration and policy iteration in Matlab
    iii. Answer theoretical questions about the properties of optimal policies/value functions and advanced algorithms
B. POMDP basics (lectures 7-12)

i. Formulate a sequential decision making problem as a POMDP
ii. Implement point-based value iteration and forward search in Matlab
iii. Answer theoretical questions about the properties of optimal policies/value functions and algorithms

How to enroll Download a Registration Form from the Education Web Site http://www.nicta.com.au/education/advacedict and return it to: Chrisanthy.theodosakis@nicta.com.au
Please discuss your enrolment with your supervisor and postgraduate coordinator

When: April 4-6, 2011

The course will be conducted in Sydney. You will be advised of the location, which will also be displayed on the web site.