Minor Changes to Course Descriptor: Decision Making in Robots and Autonomous Agents (INFR11090)

Dr. Subramanian Ramamoorthy, School of Informatics

I. Summary

This is a proposal to make *minor changes* to the course descriptor, updating it to better reflect the material that should be covered in future offerings. The changes will focus on course description, suggested readings and associated delivery methods descriptions.

The reasons for making these changes are as follows:

(i) The rapid increase in real world deployment of autonomous robots, and associated machine learning methods used to achieve these autonomous capabilities, has made it all the more important for computer scientists to consider issues including safety, explainability and trust. DMR is well positioned to bring in these topics into the curriculum. At the same time, the use of these robotics application-specific issues to illustrate the complete cycle of modeling and problem solving will make it easier for the students to grasp the why behind the methods.

(ii) This course was originally motivated as a complement to other algorithmic courses (e.g., RL, AGTA) by exposing students to issues of modeling, and how the real world problem can be cast in the language of these formalisms. So, it would be proper to rebalance the course content in terms of two halves: (i) mathematics/algorithms of decision making, (ii) topics and models motivated by applications considerations including safety, explainability, trust.

II. Proposed Changes

The proposal is to edit the summary, course description and reading list as follows:

Summary:

This course is intended as a specialized course on models and techniques for decision making in autonomous robots that must function in rich interactive settings involving interactions with a dynamic environment, and other agents (e.g., people). In the first part of the course, students will learn about formal models of decision making, and computational methods for automating these decisions within robots. In the second part of the course, we will consider issues arising in practical deployments of such autonomous robots, including problems of achieving safety, explainability and trust. Students will be exposed to current thinking on models and algorithmic methods for achieving these attributes in autonomous robots.

The content of this course has connections to other courses within our existing curriculum, such as Reinforcement Learning and Algorithmic Game Theory. A noteworthy difference is that RL and AGTA are primarily focussed on broad coverage of algorithmic methods, whereas this course will emphasize issues of modelling, with some focus on problems arising in practical robotics applications.

Pre-requisities:

This course is open to all Informatics students including those on joint degrees. However, students will benefit from prior exposure to robotics at the level of the Robotics: Science and Systems or Intelligent Autonomous Robotics.

For external students where this course is not listed in your DPT, please seek special permission from the course organiser.

Other requirements:

Prior exposure to mathematical models; Multivariate Calculus (Jacobian), Probability (expectation, conditional probability), Stochastic Processes (Markov chains), Principles of Optimization (linear programming, gradient descent methods)

Ability to program in a high level language, such as C/C++ or Python.

Course Description:

The course will cover the following major themes, although specific topics could vary from year to year.

I. Motivation
   a. Problems involving interaction: Strategically rich human-robot interaction; Multi-robot interactions
   b. How have decisions been modelled in different disciplines: probability theory, machine learning, psychology and cognitive science

II. Mathematics of decisions
   a. The utility maximization framework, Bayesian choice models
   b. Causality, Causal learning
   c. Bandit problems, Markov Decision Processes, and associated analysis methods
   d. Dynamic programming principle, and associated approximation and learning algorithms
   e. Incomplete information, Game theoretic models and solution concepts

III. Computer science of decisions
   a. Representations for planning – tradeoffs in modelling hierarchy, uncertainty, etc.
   b. Safety and trust in autonomous systems
   c. Explainability in AI
   d. Bounded rationality and cognitive biases

Relevant QAA Computing Curriculum Sections: Artificial Intelligence, Intelligent Information Systems Technologies

Reading List:

There is no single textbook for this course.

The instructor will provide lecture notes/slides, which will be complemented by readings from books and research articles.

Readings indicative of the course content include:

- B. Christian, T. Griffiths, Algorithms to Live By, William Collins Press, 2016.
- W.B. Powell, Approximate Dynamic Programming, Wiley, 2011.