Safe and Interactive Robotics

Logistics

Autumn Quarter, 3-4 Units

Instructor Name: Dorsa Sadigh

Course Assistant: Toki Migimatsu

Contact Information: `dorsa@cs.stanford.edu`

Lectures: Tue, Thu 3:00 PM - 4:20 PM, McMurtry Building 360

Office Hours: Gates 142 (by appointment)

Description

Once confined to the manufacturing floor, robots are quickly entering the public space at multiple levels: drones, surgical robots, service robots, and self-driving cars are becoming tangible technologies impacting the human experience. Our goal in this class is to learn about and design algorithms that enable robots to reason about their actions, interact with one another, the humans, and the environment they live in, as well as plan safe strategies that humans can trust and rely on.

This is a project-based graduate course that studies algorithms in formal methods, control theory, and robotics, which can improve the state-of-the-art human-robot systems. We focus on designing new algorithms for enhancing safe and interactive autonomy.

Format

The course is a combination of lecture and reading sessions. The lectures discuss the fundamentals of motion planning, formal methods applied in robotics, learning from demonstration, intent inference, and shared control required for modeling and design of safe and interactive autonomy for human-robot systems. During the reading sessions, students present and discuss recent contributions in this area. Throughout the semester, each student works on a related research project that they present at the end of the semester.

Prerequisites

There are no official prerequisites, but an introductory course in artificial intelligence and robotics is recommended.

Learning Objectives

At the end of this course you will have gained knowledge about applications of various topics in designing safe and interactive autonomous systems: temporal logics, reactive synthesis, planning and control, learning and human modeling, game theoretical foundations of interactive systems, safe learning, etc.

You will also have hands-on experience working on a research project and it is expected that you will gain the following research skills: analyzing literature related to a particular topic, critiquing papers, and presentation of research ideas.

Grading

Final Project (50%) Each student is required to work individually or in groups of two on a research project. The project requires a 2-page proposal including the relevant literature survey, a 6-10 page-long final report, and a final presentation/demo. Students who are taking the class for 4 units are required to work individually on their projects.

Student Presentations & Paper Reviews (30%) All students will get a chance to present multiple papers throughout the class during the reading days. Each paper will have two presenters each discussing the pros or cons of the paper. The presenters need to send the reviews (conference style) of their reading assignments 48 hours before the class. The presentation grade is based on how well the material is presented, how well it is connected to the rest of the papers or class, and how prepared the student is in answering questions from the class.

Scribing & Class Participation (20%) Students take turns to scribe the lecture notes of each lecture. The scribe notes are due within a week of every lecture. Students are also encouraged to participate in class as well as Piazza discussions.

Project Instructions

The research project throughout the class should study a new research problem, i.e., design a new algorithm, study a new application, etc. Literature surveys are not acceptable. The main deliverables of the project are:

Project Proposal (5%) A 2-page proposal that has identified the problem definition, literature survey, and a timeline.

Project Proposal Presentation (10%) A short presentation discussing the proposal, limitations, and challenges.

Project Presentation, Possibly with Demo (15%) A 15-min presentation reporting the final findings of the project.

Final Project Report (20%) A 6-10 page-long project report.

	Date	Topic	Readings
1	Sep 26	**Introduction to Safe and Interactive Robotics**	
2	Sep 28	**Motion Planning (lecture)**	• Spatial Planning: A Configuration Space Approach. *Lozano-Perez.* (1983). • Analysis of Probabilistic Roadmaps for Path Planning. *Kavraki, et al.* (1988). • Randomized Kinodynamic Planning. *LaValle, et al.* (2001). • Path Planning in Expansive Configuration Spaces. *Hsu, et al.* (1997). • Belief Space Planning Assuming Maximum Likelihood Observations. *Platt, et al.* (2010).

3	Oct 03	**Trajectory Optimization (lecture)**	• No readings.
4	Oct 05	**Formal Methods in Robotics (lecture)**	• Church's Problem Revisited. *Kupferman and Vardi*. (1999). • Temporal Logic-based Reactive Mission and Motion Planning. *Kress-Gazit, et al.* (2009). • A Fully Automated Framework for Control of Linear Systems from Temporal Logic Specifications. *Kloetzer, et al.* (2008). • Synthesis for Human-in-the-Loop Control Systems. *Li, et al.* (2014). • Synthesis of Human-in-the-Loop Control Protocols for Autonomous Systems. *Feng et al.* (2016).
5	Oct 10	**Formal Methods in Robotics (lecture)**	• Optimization-based Trajectory Generation with Linear Temporal Logic Specifications. *Wolff, et al.* (2014). • Model Predictive Control with Signal Temporal Logic Specifications. *Raman, et al.* (2014). • Safe Control under Uncertainty with Probabilistic Signal Temporal Logic. *Sadigh, et al.* (2016). • Provably Safe and Robust Learning-based Model Predictive Control. *Aswani, et al.* (2013). • Diagnosis and Repair from Signal Temporal Logic Specifications. *Ghosh, et al.* (2016).
6	Oct 12	**Safe Learning and Control (reading)**	• Concrete Problems in AI Safety. *Amodei, et al.* (2016). • Reachability-based Safe Learning with Gaussian Processes. *Akametalu, et al.* (2014). • Safe Exploration for Optimization with Gaussian Processes. *Sui, et al.* (2015). • Bayesian Optimization with Safety Constraints. *Berkenkamp, et al.* (2016). • Safe Visual Navigation via Deep Learning and Novelty Detection. *Richter, et al.* (2017).

7	Oct 17	**Adversarial Neural Networks (reading)**	• Exploring the Space of Adversarial Images. *Tabacof, et al.* (2015). • The Limitations of Deep Learning in Adversarial Settings. *Papernot, et al.* (2016). • Adversarial Manipulation of Deep Representations. *Sabour, et al.* (2015).
8	Oct 19	**Models of Cognition (reading)**	• Probabilistic Models of Cognition: Conceptual Foundations. *Chater, et al.* (2006). • Probabilistic Models of Cognition: Exploring representations and inductive biases. *Griffiths, et al.* (2010). • Helping People Make Better Decisions Using Optimal Gamification. *Lieder, et al.* (2016). • A Reward Shaping Method for Promoting Metacognitive Learning. *Lider, et al.* (2016). • Prospect Theory. *Kahneman and Tversky*. (1979)
9	Oct 24	**Learning from Demonstration (lecture)**	• Maximum Margin Planning. *Ratliff, et al.* (2006). • Maximum Entropy IRL. *Ziebart, et al.* (2010). • Movement Primitives via Optimization. *Dragan, et al.* (2015). • Active Preference-Based Learning of Reward Functions. *Sadigh, et al.*(2017).
10	Oct 26	**Learning from Demonstration (reading)**	• Socially Compliant Mobile Robot Navigation via IRL. *Kretzschmar, et al.* (2016). • Predicting Human Reaching Motion in Collaborative Tasks using Inverse Optimal Control and Iterative re-planning. *Mainprice, et al.* (2015). • Trajectories and Keyframes for Kinesthetic Teaching. *Akgun, et al.* (2012).
11	Oct 31	**Project Proposal Presentationsss, Experimental Design (lecture)**	

12	Nov 02	**Learning from Demonstration (reading)**	• Designing Robot Learners that Ask Good Questions. *Cakmak, et al.* (2012). • Using perspective taking to learn from ambiguous demonstrations. *Breazeal, et al.* (2012) • Understanding Intentions of Others. *Meltzoff, et al.* (1995). • Infant Imitation After a 1-Week Delay. *Meltzoff, et al.* (1988) • Rational Imitation in Preverbal Infants. *Gergely, et al.* (2002).
13	Nov 07	**Intent Inference (lecture)**	• Planning Based Prediction for Pedestrians. *Ziebart, et al.* (2009). • Goal Inference as Inverse Planning. *Baker, et al.* (2007). • Intention-Aware Motion Planning. *Bandyopadhyay, et al.* (2013). • Shared Autonomy via Hindsight Optimization. *Javdani, et al.* (2015).
14	Nov 09	**Intent Expression (reading)**	• Generating Legible Motion. *Dragan, et al.* (2013). • Manipulating Mental States through Physical Action. *Gray, et al.* (2014). • Generating Plans that Predict Themselves. *Fisac, et al.* (2016).
15	Nov 14	**Guest Lecture by Mo Chen**	
16	Nov 16	**Intent in HRI (reading)**	• Expressing Thought: Improving Robot Readability with Animation Principles. *Takayama, et al.* (2011). • Anticipation in Robot Motion. *Gielniak, et al.* (2011). • Robot Navigation in Dense Human Crowds. *Trautman, et al.* (2013).
17	Nov 21	**Interaction as an Underactuated System (lecture)**	• Planning for Autonomous Cars that Leverages Effects on Human Actions. *Sadigh, et al.* (2016). • Information Gathering Actions over Human Internal State. *Sadigh, et al.* (2016). • Robot Planning with Mathematical Models of Human State and Action. *Dragan.* (2017). • Formalizing Human-Robot Mutual Adaptation: A Bounded Memory Model. *Nikolaidis, et al.* (2016).
18	Nov 23	**Thanksgiving (no lecture)**	

19	Nov 28	**Communication and Coordination (reading)**	• Learning Behavior Styles with Inverse Reinforcement Learning. *Lee, et al.* (2010). • Asking for Help Using Inverse Semantics. *Tellex, et al.* (2014). • Knowledge and implicature: Modeling language understanding as social cognition. *Goodman, et al.* (2013). • Coordination Mechanisms in Human-Robot Collaboration. *Mutlu, et al.* (2013). • Conversational Gaze Aversion for Humanlike Robots. *Andrist, et al.* (2014). • Robot Deictics: How Gesture and Context Shape Referential Communication. *Sauppé, et al.* (2014).
20	Nov 30	**Collaboration (reading)**	• Implicitly Assisting Humans to Choose Good Grasps in Robot to Human Handovers. *Bestick, et al.* (2016). • Cooperative Inverse Reinforcement Learning. *Hadfield-Menell, et al.* (2016). • Human-Robot Cross-Training. *Nikolaidis, et al.*(2013).
21	Dec 05	**Project Presentations**	
22	Dec 07	**Project Presentations**	