Decision Theoretic Approaches to Human-Robot Social Interaction

Allison Bruce

October 24, 2003

1 Introduction

In order for robots to help with common tasks in populated environments, the ability to navigate around obstacles is insufficient. Many aspects of getting around in crowded areas, such as passing people in the hall, standing in line, and waiting to get on and off the elevator, are largely socially regulated behaviors. A robot with no model of how to participate in these interactions will be an obstacle to the people around it. Thus, we maintain that planning in order to interact appropriately is just as important as planning how to navigate from one place to another.

The problem of planning to act in social situations has a number of characteristics in common with other planning domains, suggesting potential approaches. These characteristics are uncertainty in the outcome of actions, partial observability, time dependence of actions, arbitrary distributions of time spent in a state, and opportunities to act in order to gain information. First of all, in these decision problems the outcomes of actions are uncertain, as is always the case for robots performing real-world tasks. Also, not all of the state of the world is observable. The robot can know its own location and potentially the current location of people around it, but it doesn't know what their future actions will be or what their intentions are. The actions performed by the robot and changes in the environment may take time to complete, and the world may stay in a certain state for an arbitrary distribution in time. Because this domain necessarily includes other agents, there are many opportunities to take information-gathering actions, either by observing the actions of others or by communicating with them to clarify their intentions. Many domains possessing some subset of these characteristics have successfully used Markov decison processes or some variant of them as a representation for planning.

Our proposed approach to this class of problems to use decision-theoretic planning methods and develop computational models of social interactions that are useful for these planning representations. Decision theory is a useful framework because it is a well-developed mathematical formalism that can deal with the uncertainty present in real-world problems. However, the most common decision-theoretic representation used for planning, the Markov decision process (MDP), makes limiting assumptions that make it insufficiently expressive. We believe that successful plans for interaction must deal with both the perceptual aliasing and the non-Markov nature of these problems. Therefore, we propose to explore methods for combining the expressive power of partially observable Markov decision processes (POMDPs) and semi-Markov decision processes (SMDPs) to build models that can represent both partial observability and arbitrary state duration.

The task of riding an elevator with people is a typical example of this problem domain that illustrates its inherent complexity. Uncertainty, partial observability, complex time relationships between world state and actions are all aspects of this task. The ability to move to observe other's button presses and to ask them about their intended actions offer interesting opportunities to take information gathering actions. This is a multiagent domain, which in many cases requires a game theoretic approach to achieve good solutions. But game theory adds an additional level of complexity to the task of elevator riding that we would like to avoid. Luckily, most people feel this way about riding elevators as well, which is why social conventions exist. Social domains assume that there is a way of achieving mutually beneficial results by following a set of rules. If we can learn these rules, we can avoid the need to reason strategically.

2 Related Work

Recently, there has been a variety of work involving social interaction between robots and humans [1]. Very little of this work looks at social interaction as a planning problem. Existing approaches to producing social behavior tend to be either hand coded systems or model free learning methods. Policies for interaction are often designed by hand. Whether they are based on domain knowledge about the task to be performed [2] or a more general theory of social behavior [3], the result is a set of hand-crafted rules that trigger the appropriate actions. The main problem with this approach is that designing at set of rules that produce the desired behavior is likely to be very difficult and the method of obtaining the policy will not generalize to other tasks. Learning policies for behavior by either imitation [4] or reinforcement learning [5] is a popular method. The shortcoming of learning by imitation is that a particular way of achieving the goal is learned, not a full policy. The solution learned may be brittle and fail to generalize to future instances. The problems inherent in the model free reinforcement learning approach are the difficulty of obtaining enough training data for learning to occur and the difficulty of producing rewards for the system. In social tasks, rewards are likely to be delayed, and it is difficult to have people assign rewards during a realistic interaction.

The approaches mentioned above seem unlikely to scale to situations where robots must exhibit sophisticated behavior that may take an extended period of time to achieve the goal. We choose to focus our effort on creating a representation of these tasks that is flexible and complete rather than coming up with adhoc solutions. There is existing research that treats human-robot interaction as a decision-theoretic planning problem. One application that bears many similarities to the problems that we are interested in is dialog management for robots, in particular approaches to dialog management that attempt to deal with the ambiguity of human speech in addition to speech recognition error. Work by [6] uses of the POMDP formalism to model dialog tasks for a robotic helpmate. This approach is novel because it explicitly models ambiguity by representing the topic of conversation using hidden state. This makes it possible to find policies that allow the robot to take information gathering speech acts to resolve this ambiguity. Their problem differs from ours in that time-dependence and semi-Markov properties are not important characteristics of their task. Additionally, the dialog task doesn't contain a physical component, whereas the interaction tasks we describe involve the movements of agents and the spatial relationships between them.

3 Proposed Research

The model that we propose to use to represent these problems is complex, even moreso than the POMDP, which notoriously expensive to compute policies for. Is it really necessary to both represent partial observability and to explicitly model time in a non-Markov manner? Can't good solutions to these problems be obtained by using either a POMDP or SMDP model? Using elevator riding as an example, we argue that time dependence and partial observability are essential characteristics of these types of interaction problems that must be explicitly modelled and reasoned about.

In many decision-theoretic planning problems, the world is assumed to be static. In interactions with people, this is clearly not the case. Changes in state are not event-driven by the agent. The state of the world will continue to change whether the robot acts or not. Therefore, temporal indexing is a critical aspect of the state. In addition to the dynamics introduced by the presence of people, in the elevator riding task, the location of the elevator and the state of the doors are also dynamic and changing in a way that the robot cannot completely control. Because the time at which an action is executed determines the state of the world that it is executed in, it is necessary to model the duration of time that it takes to make observations and take actions. For example, if the elevator door has just opened, the robot can head straight for it to get on. But if the door has been open for a longer period of time and the robot is still far from it, the robot will be unable to reach the door before it closes. Another action, such ask asking someone to hold the door or pressing the call button again, must be taken before it can pass through the doorway. Partial observability only adds to the complexity of the problem. There is perceptual aliasing because the robot cannot fully observe the physical state of the world (ie, where is the elevator now, how long will it take to come). The goals of the people interacting with the robot are another unobservable aspect of the dynamics of the environment. But in addition to being able to directly change the state of the world, the robot can choose take actions in order to gain more information about the world's true state. The robot can get information about people's intentions either by observing their interactions with the elevator (which buttons they press) or by directly asking them which floor they plan to exit on.

The challenge to using partially observable semi-Markov decision processes (POSMDPs) for planning is two-fold. First, the specifics of the representation must be formalized. Then methods for computing policies from these models must be developed. In order to develop the representation, we will decide on a way of expressing the time-dependent nature of state transitions based on existing work on SMDPs. Once the representation has been formalized, we can begin to explore how existing methods for computing policies that can cope with partial observability can be extended to find solutions using this new model.

Semi-Markov decision processes have been formalized in a number of different ways in the literature. SMDPs have been explored as a form of temporal abstraction for reinforcement learning [7]. They have also been used to model planning problems that allow concurrent actions [8]. An interesting formulation that is closely related to SMDPs is the time-dependent Markov decision process, which models both time-dependence and arbitrary distributions over action durations and was used for planning in a transportation domain [9]. The exact way that the non-Markov nature of state transitions should be represented in our problem will require that we examine closely exactly what kinds of temporal relationships we need to express.

The major offline MDP algorithms (value iteration, policy iteration, and linear programming) all have well-known SMDP versions [10]. Reinforcement learning algorithms, such as Q-learning and temporal distance learning, have been extended to SMDPs [11], and convergence results for SMDP Q-learning have been proven [12]. The naturalness with which MDP solution methods can be extended to SMDPs suggests to us that extensions of approximate solution methods for POMDPs to partially observable semi-Markov decision processes should also be possible. There are many different kinds of algorithms for finding approximately optimal POMDP policies, either by approximating the value function [13],[14] or by policy search [15],[16]. There are also greedy heuristics that solve the underlying MDP and make simplifying assumptions about belief during execution [17], [18]. Another class of heuristics makes use of the convexity of the value function over the belief space in order to take information-gathering actions. The Q-MDP heuristic makes the simplifying assumption that the state will become observable after one step [19]. The weighted-entropy heuristic combines the Q-value function of the completely unobservable MDP and the completely observable MDP weighted according to the entropy of the belief state [20]. Belief compression, another heuristic technique that allows for information gathering actions, represents the belief space by a set of sufficient statistics [21]. It is not yet clear which of these algorithms is the best choice to extend to our representation, but there are a wide variety of approaches to draw from.

4 Summary

Problems in the domain of social interaction have until recently been given little attention by the planning community. Approaches to designing policies for interaction have to date been mostly simplistic and ad hoc. We argue that human-robot interaction is not only interesting in terms of yielding compelling and useful robot behavior, but also because it has characteristics that make it a very challenging domain. Because of the hidden state and complex time dependencies exhibited by these problems, the ideal representation would be one that combines the modeling power of POMDPS and SMDPs. To our knowledge, there is no active research dedicated to the integration of these models. This may be because exact solutions to POMDPs are intractable, and a model that includes semi-Markov state would only add further computational complexity. However, semi-Markov extensions to MDPs are relatively straightforward, and there are a number of useful algorithms for generating approximately optimal POMDP policies. We believe that extending these heuristic algorithms to partially observable semi-Markov decision processes should be feasible and should prove useful for producing high quality solutions to problems that involve both semi-Markov characteristics and hidden state.

References

[1] Terrence W. Fong, Illah Nourbakhsh, and Kerstin Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, 2003.

[2] Yasuchi Nakauchi and Reid Simmons, "A social robot that stands in line," *Autonomous Robots*, , no. 3, pp. 313–324, May 2002.

[3] C. Breazeal. and B. Scassellati, "How to build robots that make friends and influence people," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1999.

[4] M. Nicolescu, "Learning and interacting in human-robot domains," 2001.

[5] H. Ayai F. Iida and F. Hara, "Behavior learning of face robot using human natural instruction," in *8th IEEE International Workshop on Robot and Human Communication* , 1999.

[6] N. Roy, J. Pineau, and S. Thrun, "Spoken dialogue management using probabilistic reasoning," in *38th Annual Meeting of the Association for Computational Linguistics*, 2000.

[7] Richard S. Sutton, Doina Precup, and Satinder P. Singh, "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning," *Artificial Intelligence*, vol. 112, no. 1-2, pp. 181–211, 1999.

[8] Khashayar Rohanimanesh and Sridhar Mahadevan, "Decision-Theoretic planning with concurrent temporally extended actions," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 2001, pp. 472–479.

[9] Justin A. Boyan and Michael L. Littman, "Exact solutions to time-dependent MDPs," in *Neural Information Processing Systems*, 2000, pp. 1026–1032.

[10] L. M. Puterman, *Markov Decision Processes*, Wiley, New York, 1994.

[11] Steven J. Bradtke and Michael O. Duff, "Reinforcement learning methods for continuous-time Markov decision problems," in *Advances in Neural Information Processing Systems*, G. Tesauro, D. Touretzky, and T. Leen, Eds. 1995, vol. 7, pp. 393–400, The MIT Press.

[12] R. Parr, "Hierarchical control and learning for markov decision processes," 1998.

[13] Sebastian Thrun, "Monte Carlo POMDPs," in *Advances in Neural Information Processing 12*, 2000, pp. 1064–1070.

[14] J. Pineau, G. Gordon, and S. Thrun, "Point-based value iteration: An anytime algorithm for pomdps," 2003.

[15] Andrew Y. Ng and Michael Jordan, "PEGASUS:A policy search method for large MDPs and POMDPs," in *Sixteenth Conference on Uncertainty in Artificial Intelligence*, 2000, pp. 406–415.

[16] James Bagnell and Jeff Schneider, "Covariant policy search," in *Proceeding of the International Joint Conference on Artifical Intelligence*, August 2003.

[17] I. Nourbakhsh, R. Powers, and S. Birchfield, "Dervish an office navigating robot," *AI MAGAZINE*, vol. 16, no. 2, pp. 53–60, 1995.

[18] Reid Simmons and Sven Koenig, "Probabilistic robot navigation in partially observable environments," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 1995, pp. 1080–1087.

[19] Michael L. Littman, Anthony R. Cassandra, and Leslie Pack Kaelbling, "Learning policies for partially observable environments: Scaling up," in *Proceedings of the Twelfth International Conference on Machine Learning*, Armand Prieditis and Stuart Russell, Eds., San Francisco, CA, USA, 1995, pp. 362–370, Morgan Kaufmann publishers Inc.: San Mateo, CA, USA.

[20] Anthony R. Cassandra, Leslie Pack Kaelbling, and James A. Kurien, "Acting under uncertainty: Discrete bayesian models for mobile robot navigation," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1996.

[21] N. Roy and G. Gordon, "Exponential family pca for belief compression in pomdps," in *Advances in Neural Information Processing Systems*, 2002.