Assisted Exploration of Complex State Spaces

Brandon Mechtley (CSE 574 Semester Project)

April 29, 2008

Abstract

This paper describes a work-in-progress application of recent work in assistive planning to realtime, embodied, interactive scenarios involving complex systems. In modeling these interactions, focus is placed not on biasing the user toward a specific goal, such as winning a game, but rather on exploring the extreme states of the system as he or she sees fit, necessitating that a user's goal be modeled as distribution over a number of predefined possible points of interest in the system.

1 Introduction

1.1 Behavioral Change for Sustainability

A key challenge facing sustainable growth is in promoting behavioral change amongst all involved. Perhaps the most effective mode of persuasion is to promote understanding. Unfortunately, those systems which are most threatened by unsustainable growth are often extremely complex, seemingly beyond summary. Common to the vast majority of these problems are environmental, social and political, and material costs. Traditional quantitative methods of analyzing the interplay of these domains are often limited in terms of public accessibility and also commonly leave out the possible social negotiations between those involved in the system.

1.2 Embodied Exploration for State Space Exploration

To study how users can negotiate understanding of these systems, we have employed a media-rich, multimodal environment, SMALLab, to model complex systems. SMALLab is a 10 by 10 foot floor projection with quadraphonic audio, three-dimensional object tracking, and gesture recognition. By supporting embodied interaction, multi-user choreography and social interaction, sonic feedback, and visual feedback, SMALLab is well-suited for interacting with high-dimensional spaces.

In this environment, multiple users can take on various roles, ranging from resource suppliers and regulatory bodies to consumers. Toward the understanding described above, we have sought to allow users full access to the entire state space without qualitative judgment of a user's actions or mediated biases toward any one specific state (e.g. through game points).

1.3 Assistive Planning

To assist users in exploring the full state space, I am working to develop an online strategy for assistive planning that can be used to help users explore their goals. Specifically, an assistive planner will be used to provide suggestions to (but not perform actions for) a user who has taken the role of a regulatory body. The systems described above have a number of properties, including partially observable user goals, continuous-valued state variables, and multi-agent interaction. By appropriately handling modeling issues of continuous-valued state variables, suboptimal goal-dependent user policies, and an application of assistance through suggestion rather than action, the POMDP modeling of assistive planning described in [3] can be fit to this application. The method of assistance described in [1] differs from that of [3] in that it uses a fully-observable MDP framework to assist in a singular goal. Since this application must maintain relative neutrality between multiple possible goals in this domain, it is necessary to model the partial observability of user goals described in [3].

2 Domain Modeling

2.1 Agents and Effort-based Regulation

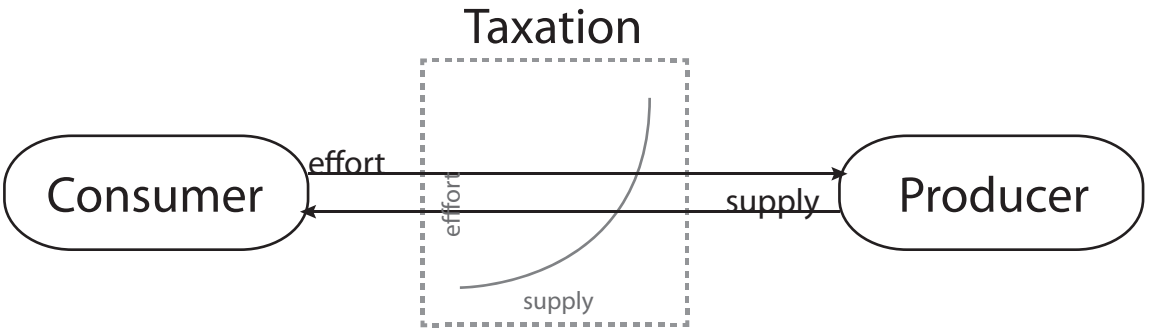

Figure 1: An example agent-to-agent transfer of supply per unit effort, where the regulatory body maintains a graduated tax curve.

At the highest level, scenarios are modeled in terms of multiple agents that have structured two-way mediated communications, such as resource transfer, in the form of provisions and requests, such as supply and demand. These transfers can be regulated by a regulatory body on both ends by modifying either the received request or provision. The goal of this regulator is generally to assist agents in obtaining certain equilibria that minimize effort across all users, one goal bias which has been included in the assistive model.

The agents whose transfers are being modified by the regulator have knowledge of his or her presence and capability of modifying his or her actions through social negotiation. Therefore, for the representation of the state space, it is sufficient to model the state of the world as the conjunction of individual agents' states and assume that the decisions of the regulator reflect the conclusions of social negotiations between all performers.

Where costs of real-world systems are often multivariate in nature, the SMALLab space specifically affords social, cognitive, and physical efforts. Social efforts are understood in terms of negotiations between players, cognitive effort-based tasks are defined as those that require considerable focus from a user (such as multitasking), and physical effort is effort that is directly related to the amount of physical exertion a user must exert to perform a particular action. This effort cost will be

explicitly described in terms of a cost function, $C(s, a)$ on the agent's actions in Section 3.
The states of a regulator, ρ, can therefore be described as a tuple of regulation scheme variables for every pair of communicating agents.
The state of each agent, $\phi_i, 1 \leq i \leq n$, can also be described as a conjunction of continuous-valued state variables that represent internal properties of the agent, such as the amount of available supply for a supplier.

2.2 Discretization

In this domain, all state variables are in terms of continuous values, so to make the problem tractable, a discrete approximation of the variables is necessary that will retain certain properties of the original topology, including the presence certain interesting goal states. Existing domain-independent methods of discretization include the fitting of approximated piece-wise models [4], perhaps with assumptions about the topology of the space, such that it can be fit into a piece-wise linear model [2].
I have chosen to use a method based on previous results in this domain in estimating perceptual clusters of effort expenditure and media feedback. This offline method assumes that state variables are independent and has several limitations over the above methods, which will be mentioned. Discretization must be performed both in terms regulatory actions and agents' states, so two different methods have been employed, respectively:

1. For regulatory actions, clustering methods are used to learn levels of physical effort for users offline by fitting an ordered set of physical effort values to observed user actions in order to approximate perceptual notions of physical effort associated with gesture size, speed, and so on. Once these clusters of effort are found, their mean effort values are used to define actions that achieve the respective value of the regulation variable for the effort expended, according to a predefined mapping.

   A limitation of this method is that notions of physical effort tend to vary from person to person. A clear example is gesture size, which varies quite directly with user height. In the future, a possible direction may be to

implement an online learning scheme to customize this discretization to the user performing the role.

2. Agent states will be discretized according to the perceptual characteristics of the media that represent them, making the assumption that if a certain system state is not perceptually distinguishable from another, it will be unlikely for users to intentionally arrive at the state as a result of their own actions.

   This discretization has a large limitation that has been observed in trials with users. Although an agent may not intentionally arrive at a particular state, it is very common that an imperceptible change in one user's state may allow for a very perceptible change in another user's state. To model for this type of interplay, it is necessary to discretize the entire system holistically, instead of considering each state variable independent of others. To do this would require some sort of approximation of the state space's total topology.

2.3 Goal Identification

A generalizable model for goal identification would identify maintainable effort minima in the state space. Given a closed system with deterministic user actions, a potential goal could be one where no user performs any action. Maintainability can be described in terms of temporal dynamics in the system state independent of user actions, such as fluctuations in the supply of a resource or user necessities such as satiation, profit, and so forth. With these aspects in mind, the goal distribution may enlarge to include several local minima.

Another method that has been used with some success is to simply identify locations in the state space that receive a great deal of attention [5]. This heuristic is of course specific to the domain of travel, but it may be relevant to this domain as well, as one might expect the states users are most interested in exploring to be those in which they linger.

For the purposes of this project, I am uniquely identifying potential goal states for each domain without quantitative analysis.

3 Assistive Planning

3.1 POMDP Model

A POMPD is defined in terms of the model $\langle S, A, T, C, I, O, \mu \rangle$, where S is the set of system states, A is the set of actions, T is the set of stochastic transition distributions, C is an action cost function, I is the initial state distribution, O is the set of observations, and $\mu(o, s)$ is a distribution over observations given the current state, s [3].

The set of finite world states, W, has been defined above, each state being a tuple $(\rho, \phi_1, ...\phi_n)$. However, as in [3], the state must also be described in terms of the unobservable goal state, $g \in G$. Therefore, the final state space S is defined as $W \times G$. In addition, W contains information about the last action to be performed by the regulator performer.

Each discrete regulation variable has a respective action with cost equal to the predefined amount of effort associated with obtaining that regulation state. This predefined effort is modeled to ensure increased and broader regulation accounts for resistances not accounted for by social negotiation in the scenario. The action set is equal to this set of regulatory actions in addition to the *noop* action, as the POMDP will be used to calculate the optimal policy distribution for the regulatory assistant's actions, assuming that the regulatory user heeds the assistant's advice. The cost for performing the *noop* action in state (w, g) is equal to the expected cost of the next agent action, according to the agent's assumed goal-dependent policy $\pi(w, g)$.

Since the system is not designed to impose any bias on goal distribution, the assumption can be made, as in [3], that $T((w, g), a, (w', g'))$ will equal zero for $g \neq g'$ and that the transition probability $T(w, noop, w')$ will equal the probability that $T(w, \pi(w, g)) = w'$, that is the probability that the next action performed by the agent will bring the world state to w'.

The observations account for observed world and action states. For the *noop* action, the observation is equal to the new world state and action performed by the regulator performer immediately after the *noop*. For other actions, $\mu'(w', g) = (w', a)$, that is the observation is equal to the observed world state (with ensuing

action), as it is fully observable, and therefore deterministic. Since the goal state component is completely unobservable, it is not included in observations.

3.2 Estimating User Policy and Prior Goal Distribution

As in [3], it is possible to estimate the agent policy $\pi(w, g)$ by solving for a separate MDP for each goal, $g \in G$, ahead of time. However, in this domain, it is not possible to assume that the regulatory user is near optimal, so it may be necessary to estimate the regulator user policy through offline analysis, where the probability of performing an action given a world state and specified goal is simply equal to the normalized frequency with which it emerges in archived data. The benefit of this, of course, is that it will more closely resemble user actions. Even with optimal MDP approximations of the user, transition probabilities would need to be estimated using the same type of analysis—the only extra requirement of modeling the entire user policy this way being that each episode needs to be labeled with a goal.

In previous user trials, we have annotated dialogs of user narration, which could yield this labeling as well information about prior knowledge of initial goal distribution (G_0). [3] suggests that the goals of each episode can be assumed to be those goals that are reached. However, if the user is operating without assistance, it can often be that the actions that seem most optimal to the user actually lead him or her to an incorrect goal, so it is necessary to have some idea of what the user's intended goal is. If the user is operating with assistance in the trials, it would be necessary to either already have a somewhat optimal assistant, or it would be necessary to perform an exponentially larger set of trials given a large span of possible assistive actions at each state.

3.3 Goal Estimation and Action Selection

Once the the state space and assistive actions are discretized and the goal-dependent user policies, $\pi(w, g)$, and initial goal distribution, G_0, are obtained, the techniques for goal estimation and action selection follow the assistive POMDP exactly. Namely, when the agent performs an action, the probability of the goal being g,

given the observation at time t, is updated as

$$P(g|O_t) = (1/Z) \cdot P(g|O_{t-1}) \cdot \pi(a|w, g), \tag{1}$$

given normalizing constant Z [3]. The probability of the goal is updated to be the previously-calculated probability that the goal was selected in the previous action step *and* that the action would be performed in the current world state by the user policy given said goal.

Action selection can then be performed through a greedy heuristic that optimizes the expected cost of performing the assistive action and then following the user policy for a goal, weighted by the probability of the goal actually being selected at the current point in time, that is

$$H(w, a, O_t) = \sum_g Q_g(w, a) \cdot P(g|O_t). \tag{2}$$

The next action suggested will be that which maximizes H given the current world state and observations.

4 Conclusion

Assistive planning has many great potential uses in interactive media, including being used as a way to assist users in navigating the extremities complex, high-dimensional state spaces. The method explored here seeks to find a way to create a generalizable representation of these complex systems through representing them with modular agents with communication channels that can be regulated. By using a POMDP framework to estimate a regulatory player's goal, an suggestive assistant can reduce the cognitive effort needed to explore the system.

There is still much work to be done on this project; it is very much incomplete apart from having yet to obtain results.

1. The discretization method described poses serious limitations on the accuracy of the model, as it may lose potential goal states or methods of achieving them. A method of discretization that is more informed by (or less affected by) the interdependencies between state variables and can map the space

more holistically would avoid these problems.

2. Although plenty of data is available from user studies, the current method of user policy and initial goal distribution estimation may require yet more to converge upon a stable policy. Especially for highly complex systems, the number of episodes required may be immense. It would be a project in itself, perhaps to do a study of the tradeoffs between assuming user optimality and estimating policies via evaluating goal-dependent MDPs and using statistical methods.

References

[1] J. Boger, P. Poupart, C. Boutilier, G. Fernie, J. Hoey, and A. Mihailidis. A decision-theoretic approach to task assistance for persons with dementia. In *Proceedings of the International Joint Conference in AI*, 2005.

[2] Z. Feng, R. Dearden, N. Meuleau, and R. Washington. Dynamic programming for structured continuous markov decision problems. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 2004.

[3] A. Fern, S. Natarajan, K. Judah, and P. Tadepalli. A decision-theoretic model of assistance. In *Proceedings of the International Joint Conference in AI*, 2007.

[4] G. Gordon. Stable function approximation in dynamic programming. In *Proceedings of the Twelfth International Conference on Machine Learning*, 1995.

[5] L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. In *Proceedings of the International Conference on Artificial Intelligence*, 2004.