Planning for Social Interaction with Sensor Uncertainty

Planning for Social Interaction with Sensor Uncertainty

Mary Ellen Foster
School of Mathematical and Computer Sciences
Heriot-Watt University
Edinburgh EH14 4AS, Scotland, UK
M.E.Foster@hw.ac.uk

Ronald P. A. Petrick
School of Informatics
University of Edinburgh
Edinburgh EH8 9AB, Scotland, UK
rpetrick@inf.ed.ac.uk

Abstract

A robot coexisting with humans must not only be able to perform physical tasks, but must also be able to interact with humans in a socially appropriate manner. In this paper, we describe an extension of prior work on planning for task-based social interaction using a robot that must interact with multiple human agents in a simple bartending domain. We describe how the initial state representation developed for this robot has been extended to handle the full range of uncertainty resulting from the input sensors, and outline how the planner will use the resulting uncertainty in the state during plan generation.

Introduction

A crucial aspect in the design of an interactive system is *state management*: transforming the noisy, continuous hypotheses produced by the low-level input processing components into a form that can be used as the basis for higher-level action selection by a component such as a planner. Intuitively, states represent a point of intersection between low-level sensor data and the high-level structures used for action selection. Since states are induced from the mapping of sensor observations to property values, the challenge of building an effective state manager rests on defining appropriate mapping functions. A state representation that considers only the highest-confidence inputs is straightforward to maintain and reason with, but discards a great deal of potentially useful information. On the other hand, a representation that takes into account the full set of input possibilities—along with their estimated confidence scores—can be more robust and informative, but requires more sophisticated methods of maintenance and more complex forms of reasoning and planning.

The particular application we consider here is a robot bartender called JAMES (Figure 1), which has the goal of supporting socially appropriate multi-party interaction in a bartending scenario.[1] In particular, the robot's sensors monitor two primary input modalities: vision and speech. Based on observations about the agents in the bar provided by these sensors, the system maintains a model of the social context, and decides on effective and socially appropriate responses in that context. Key to our approach is the use of a high-level planner for action selection in the robot system, in the place of a traditional interaction manager (Larsson and Traum 2000). Specifically, we use the knowledge-level planner PKS (Petrick and Bacchus 2002; 2004), a choice that is motivated by PKS's ability to work with incomplete information and sensing actions, since the robot will often have to gather information from its environment (e.g., by asking a customer for a drink order) in addition to performing physical tasks such as handing over drinks.

[1] See www.james-project.eu for more information.

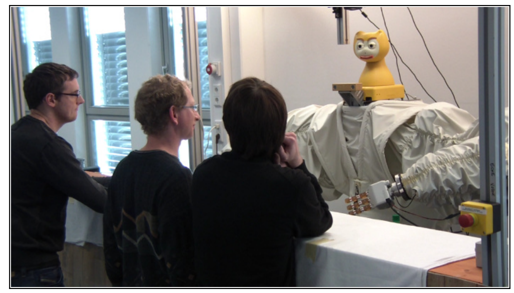

Figure 1: The JAMES robot bartender

In this paper, we describe how the initial, deterministic state representation has been extended to incorporate the full data from the robot's input sensors, and how the planner is using this enhanced representation during plan generation.

State Management with Uncertain Input

The task of the *state manager* in the robot bartender system is to keep track of information about the agents in the scene: for example, their locations, whether they are currently seeking the bartender's attention, and their drink orders. The state is derived from the continuous stream of messages produced by the low-level input and output components. In addition to storing low-level sensor information, we also infer additional relations not directly reported by the sensors; for example, we fuse vision and speech to determine which user should be assigned a recognised speech hypothesis, and use the vision data to estimate each customer's attention-seeking state (Foster, Gaschler, and Giuliani 2013).

Since the input provided by the vision and speech processing components is uncertain, there is an inherent uncertainty about the state. However, for simplicity, the state representation used in the initial JAMES system (Petrick and Foster 2013) stored only the highest-probability hypotheses, with no

seeksAttention(A1)	true	0.75
seeksAttention(A2)	false	0.45
lastSpeaker()	A1	1.0
lastEvent()	userSpeech(A1)	1.0
drinkOrder(A1)	green lemonade	0.677
	blue lemonade	0.322
lastAct(A1)	greet	0.25

Table 1: State excerpt, showing both the old discrete representation (highlighted portion) and the new representation

action ask-drink(?a : agent)
preconds: K(inTrans = ?a) $\wedge$ $\neg$K(ordered(?a)) $\wedge$ $\neg$K(otherAttnReq) $\wedge$ $\neg$K(badASR(?a))
effects: add(K_f, ordered(?a)), add(K_v, drinkOrder(?a))

Figure 2: Example PKS action in the bartender domain

confidence measures; a sample state using this representation is shown in the highlighted portion of Table 1.

This initial representation simplified action selection considerably, but also discarded potentially relevant information from the input sensors. We have therefore extended the initial version of the state manager to associate each hypothesis with a confidence score, and to include alternative hypotheses about a customer's drink order (Foster, Keizer, and Lemon 2014). Table 1 shows a full state using this expanded representation: in addition to the old-style state information in the highlighted portion, this state adds confidence scores to all properties—meaning that low-confidence relations like lastAct(A1) can now be included—and also includes multiple values for relations like drinkOrder(A1). The resulting representation is similar to the *Discrete* distribution used in RDDL (Sanner 2011), the language for the recent probabilistic tracks of the International Planning Competition.

Planning under Sensor Uncertainty

To generate plans for the robot, PKS uses a knowledge-level domain model that includes a specification of the physical, sensory, and linguistic (speech) actions available to it. The current domain supports simple interactions with individual agents for ordering drinks from the robot, as well as socially motivated behaviour such as group ordering and multi-party turn-taking. For example, Figure 2 shows the PKS representation for the ask-drink(?a) action ("ask an agent ?a for a drink order"), which is modelled as a sensing action that returns a placeholder (the function drinkOrder) for information that will become known at execution time.

We are currently improving our ability to plan with sensor uncertainty in the states described above. Since PKS does not (currently) work directly with probabilistic representations, we are modelling disjunctive state information like drinkOrder in Table 1 using PKS's ability to use "exclusive or" formula of the form $(\phi_1|\phi_2|\ldots|\phi_n)$ (which is interpreted as "one, and only one, of the ϕ_is is true"), ordered by decreasing confidence values. To incorporate confidence information for single-value relations, such as seeksAttention, we instead employ empirically determined confidence thresholds to determine whether to accept the current state information or to make an effort to gather more information before continuing. Updated state information is regularly sent to the planner from the state manager after action execution, and used for monitoring and replanning purposes.

Once the extended state information is available in the planner's knowledge state, it can be directly used during plan construction. In practice, such knowledge often has the effect of introducing additional sensing actions into a plan, to disambiguate between disjunctive alternatives. To aid this process, we are adding new actions which correspond to information-gathering (clarification) questions that the robot can ask to help clarify uncertain beliefs—without asking an agent to simply repeat an utterance, which is often interpreted by humans as a poor dialogue move (Skantze 2005).

Future Work

We will shortly carry out a user study to assess the impact of the new state representation and updated planning approach on user interactions with the system, comparing a version of the bartender that deals with all of the above forms of uncertainty to one that does not. Based on the behaviour of previous versions of the system—which did not incorporate state uncertainty but still performed reasonably well—we expect to see a positive impact in task performance (i.e., the number of drinks correctly served), since the bartender should clarify lower-confidence or ambiguous state hypotheses instead of simply serving what it believes to be the requested drink. On the other hand, it may be that the subjective user judgements will be negatively affected if the system clarifies too frequently in contexts where the top hypothesis is correct.

Acknowledgements

This research has received funding from the European Union's 7th Framework Programme under grant No. 270435.

References

Foster, M. E.; Gaschler, A.; and Giuliani, M. 2013. How can I help you? Comparing engagement classification strategies for a robot bartender. In *Proceedings of ICMI 2013*.

Foster, M. E.; Keizer, S.; and Lemon, O. 2014. Towards action selection under uncertainty for a socially aware robot bartender. In *Proceedings of HRI 2014*.

Larsson, S., and Traum, D. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering* 6(3–4):323–340.

Petrick, R. P. A., and Bacchus, F. 2002. A knowledge-based approach to planning with incomplete information and sensing. In *Proceedings of AIPS 2002*, 212–221.

Petrick, R. P. A., and Bacchus, F. 2004. Extending the knowledge-based approach to planning with incomplete information and sensing. In *Proceedings of ICAPS 2004*, 2–11.

Petrick, R. P. A., and Foster, M. E. 2013. Planning for social interaction in a robot bartender domain. In *Proceedings of ICAPS 2013, Novel Applications Track*.

Sanner, S. 2011. Relational Dynamic Influence Diagram Language (RDDL): Language description.

Skantze, G. 2005. Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication* 45(3):325–341.