Towards State Summarization for Autonomous Robots

Daniel Brooks, Abraham Shultz, Munjal Desai, Philip Kovac, and Holly A. Yanco
Computer Science Department
University of Massachusetts Lowell
One University Avenue
Lowell, Massachusetts 01854
{dbrooks, ashultz, mdesai, pkovac, holly}@cs.uml.edu

Abstract

Mobile robots are an increasingly important part of search and rescue efforts as well as military combat. In order for users to accept these robots and use them effectively, the user must be able to communicate clearly with the robots and obtain explanations of the robots' behavior that will allow the user to understand its actions. This paper describes part of a system of software that will be able to produce explanations of the robots' behavior and situation in an interaction with a human operator.

Motivation

The Urban Search and Rescue (USAR) field is beginning to adopt teleoperated systems, such as those used in the wake of Hurricane Katrina to check partially-collapsed buildings for survivors (Micire 2008). Because they have little or no autonomy, these systems require the dedicated attention of a trained operator. This requirement restricts the availability of such systems to teams which have a trained teleoperator among their members.

To reduce the difficulty of using telepresence and robotic systems in USAR, the operator should be able to issue orders to an autonomous system and expect that those orders will be followed. This form of supervisory control helps reduce the amount of time and effort the operator spends on navigation and controlling individual motions (Fong and Thorpe 2001).

One question of particular importance is how operators of autonomous systems will be able to maintain or regain control of a robot when its status becomes unclear or it exhibits unexpected behaviors. If the robot has been moving autonomously, the user will not have the same degree of situation awareness that they would have had if they had been explicitly controlling the robot's motions. In such a case, it would be helpful if the operator could query the robot for information about its state and behavior, receive answers, and develop an informative dialog with the robot.

Operators of autonomous robots will be able to better understand how their instructions were interpreted if the robot is capable of recounting recent events or explaining some specific action that it took. We call this process of generating a concise, detailed, and useful representation of the robot's current state and world model *state summarization*. We believe that having access to this type of information will increase operator trust of autonomous systems and free the operator from having to "babysit" each machine to understand what it is doing.

Related Work

It is not uncommon for people who work with computers on a regular basis to desire an explanation for some unexpected negative behavior that the user finds irrational. Previous research in the fields of artificial intelligence, data mining, and machine learning has sought to provide reasonable ways of having an autonomous system explain its decisions and subsequent actions.

Explaining events in a simulation is part of the motivation for explainable AI (XAI) as proposed in Gomboc et al. (2005) and Core et al. (2006). That research focuses on *a posteriori* methods of extracting representations of behavior from the data generated by a simulation that did not originally include a detailed, accessible behavior model. Because we are the developers of our system, we can change the system to produce a maximally useful behavior representation.

Lemon et al. (2001) focused on generation of speech for interaction with users. The paper describes an interface that combines natural language commands and dialog with a computer-based map interface. This system allows the user and robot to agree on pronoun referents without specific names, such as the command "Go here," coupled with a click on the map interface for disambiguation.

System Design

Our system uses a modular, message-passing infrastructure. The messages are units of data called "parcels". Parcels can be created from incoming sensor data, or by processing that takes place in one of the modules of the system. When a module of the infrastructure creates a new parcel based on one or more older parcels, it includes references to the parcels that provoked the creation of the new parcel. As items of data progress through the system, the set of parcels to which they are linked expands. The complete set can be viewed as a tree, with the action that the robot takes as the root of the tree. The parcels immediately preceding the root are the first level of the tree, the parcels that went into them are the next level, and so forth, extending out to the leaf nodes, which will mostly be raw sensor data. This set of links provides a context for decisions that proceed from the data. Using this context, the robot can account for its perception of the environment at the time that a particular action changed the state of the robot. This means that explanations such as "Why is the robot not moving forward?" can be answered succinctly with a message such as "The way forward is blocked", and with more detail using messages such as "There is an object located 11 degrees left of the robot's center and 0.3 meters away."

An explanation can be made useless by being either too succinct to capture important details, or by being so verbose that details are obscured. In order to support a flexible level of detail, each parcel will be tagged with a level of abstractness and a level of interestingness.

Abstractness is a unitless measurement of how distant the parcel of data and action that uses it are from raw sensor data. The actual sensor data coming into the robot is the least abstract information. As it is interpreted, it gains abstraction. A raw laser sweep is not abstract, but interpreting sections of it as fiducial markers gives them a more abstracted meaning. Interpreting those fiducial markers as objects that are related to the robot's current goal is yet another level of abstraction.

Levels of interestingness are an orthogonal measure to abstractness, in that the abstractness of a unit of data is not a function of its interestingness. Interestingness is a unitless measure of how related a specific piece of information is to the completion of the robot's goals. A fiducial marker that does not identify the sought object may be ignored, but will still have a higher "interestingness" rating than a laser sweep that did not detect any fiducial markers, because it had a higher chance of being something relevant. As a consequence, if the robot is searching for a particular fiducial, it may choose to prioritize reporting that fiducial over other things it sees, or only mention the other things if it is prompted with a question. The linkage of parcels into trees provides a convenient first step towards automating the detection of interestingness, as it provides a method to assess the degree of relevance of a given element of the data to the robot's behavior. Database entries that are are referenced frequently in decisions made by action modules contain data that was relevant for making decisions that affected the state of the robot, and so is more interesting. Data that is not referenced was examined by the robot and determined not affect the robot's state.

The magnitude of the difference between the expected action and the actual action could also be considered as part of interestingness. If the instruction sent by the operator is to go forward into a room, and the robot turns to go in a different door, the difference between the expected behavior and the observed behavior is quite large. When the robot "disobeys" the operator in this manner, the large difference will more likely result in a demand for an explanation where a small difference might have passed unnoticed.

Beyond trees of parcels, the recorded data is also timestamped, and can be sequenced chronologically. This provides a means of supporting not only causal dialogs, where the user asks why an event occurred, but also chronological dialogs, where the user asks when an event occurred or where the event falls in a sequence. We believe that the most useful dialogs will occur when both causal and chronological dialogs are used together. The user will be able to ask about events in the robot's past with chronological queries, and find more detail about those events with causal queries.

The ease of interpretation of the summarized data is a problem which must be addressed regardless of the medium in which it is presented. It is imperative that the information be comprehensible and useful to a user with little or no formal technical training and a limited understanding of the machine and its programming.

Current Work

The current interface for the visualization is a timeline with each action that the robot has performed displayed as a tree. This interface will be extended to allow each new action to be added to the timeline as the action occurs. The user can expand each tree to in the timeline to view more or less detail around a certain period, and slide the timeline back and forth to view more recent or later events.

Interestingness calculators based on the difference between the user selected heading and the actual heading, the magnitude of recent changes in the heading, and the amount of fiducial data found have already been implemented. Abstractness calculators based on number of referenced parcels and parcel type are also complete. These calculators are implemented as modular units that can be configured prior to running the system, so various approaches can be quickly tested.

Acknowledgements

Thanks to Mikhail Medvedev for his help in developing PASSAT. This work was funded in part by an Army Research Office MURI (W911NF-07-1-0216) and the National Science Foundation (IIS-0546309).

References

Core, M.; Lane, H.; Van Lent, M.; Gomboc, D.; Solomon, S.; and Rosenberg, M. 2006. Building explainable artificial intelligence systems. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, 1766.

Fong, T., and Thorpe, C. 2001. Vehicle teleoperation interfaces. *Autonomous Robots* 11(1):9–18.

Gomboc, D.; Solomon, S.; Core, M. G.; Lane, H. C.; and Lent, M. V. 2005. Design recommendations to support automated explanation and tutoring. In *Proceedings of the 14th Conference on Behavior Representation in Modeling and Simulation (BRIMS05), Universal*.

Lemon, O.; Bracy, A.; Gruenstein, A.; and Peters, S. 2001. Information states in a multimodal dialogue system for human-robot conversation. In *5th Wkshp on Formal Semantics and Pragmatics of Dialogue (Bi-Dialog 2001)*, 57–67.

Micire, M. J. 2008. Evaluation and field performance of a rescue robot. *Journal of Field Robotics* 25(1-2):17–30.