An Approach to Ontology-Based Intention Recognition Using State Representations

Craig Schlenoff[1,2], Sebti Foufou[2,3] and Stephen Balakirsky[1]

[1]*Intelligent Systems Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD USA*
[2]*LE2i Lab University of Burgundy, Dijon France*
[3]*Computer Science and Engineering, Qatar University, Doha, Qatar*
[craig.schlenoff@nist.gov, sfoufou@qu.edu.qa, stephen.balakirsky@nist.gov]

Keywords: intention recognition: human-robot interaction and safety: state representation: ontology

Abstract: In this paper, we present initial thoughts on an approach to ontology/logic-based intention recognition based on the recognition, representation, and ordering of states. This is different than traditional approaches to intention recognition, which use activity recognition and the ordering of activities. State recognition and representation offer numerous advantages, including the ability to infer the intention of multiple people working together and the fact that states are easier for a sensor system to recognize than actions. The focus of this work is on the domain of manufacturing assembly, with an emphasis on human/robot collaboration during the assembly process.

1 INTRODUCTION

Safe human-robot collaboration is widely seen as key to the future of robotics. When humans and robots can work together in the same space, a whole class of tasks becomes amenable to automation, ranging from collaborative assembly, to parts and material handling and delivery. Keeping humans safe requires the ability to monitor the work area and ensure that automation equipment is aware of potential danger soon enough to avoid it. Robots are under development throughout the world that will revolutionize manufacturing by allowing humans and robots to operate in close proximity while performing a variety of tasks (Szabo, Norcross et al. 2011).

Proposed standards exist for robot-human safety, but these standards focus on robots adjusting their speed based on the separation distance between the human and the robot (Chabrol 1987). The approaches focus on where the human is at a given point in time. It does not focus on where they are anticipated to be at points in the future.

A key enabler for human-robot safety involves the field of intention recognition, which involves the process of (the robot) understanding the intention of an agent (the human(s)) by recognizing some or all of their actions (Sadri 2011) to help predict future actions. Knowing these future actions will allow a robot to plan its own actions in such a way as to either help the human perform his/her activities or, at a minimum, not put itself in a position to cause an unsafe situation.

In this paper, we present an approach for ontology-based intention recognition using state-based information. In this context, state is defined as a set of one or more objects in an area of interest that consist of specific recognizable configuration(s) and/or characteristic(s). This is different than most ontology-based approaches in the literature (as described in Section 2) as they primarily focus on activity (as opposed to state) recognition and then use a form of abduction to provide explanations for observations. In the approach presented in this paper, state information serves as the focus of the observations, which provides many advantages over the representation of activities. One such advantage is the ability to handle more than one single observed agent, which is a significant limitation of current approaches (Sadri 2011). This paper focuses on the knowledge requirements necessary to represent state information for ontology-based intention recognition.

Section 2 describes the state of the art in ontology/logic-based intention recognition. Section 3 provides an example of the initial domain of interest, namely, industrial assembly operations. Section 4 describes the newly-formed IEEE Ontologies for Robotics and Automation (ORA) Working Group and describes some information

requirements necessary to represent state-based intention recognition, including spatial relations, ordering constraints, and associations with overall intentions. Section 5 describes how this information can be put together to perform intention recognition. Section 6 concludes the paper.

2 ONTOLOGY/LOGIC-BASED INTENTION RECOGNITION

As mentioned in the introduction, intention recognition traditionally involves recognizing the intent of an agent by analyzing the actions that the agent performs. Many of the recognition efforts in the literature are composed of at least three components: (1) identification and representation of a set of intentions that are relevant to the domain of interest, (2) representation of a set of actions that are expected to be performed in the domain of interest and the association of these actions with the intentions, and (3) recognition of a sequence of observed actions executed by the agent and matching them to the actions in the knowledge base.

There have been many techniques applied to intention recognition that follow the three steps listed above, including ontology-based approaches (Jeon, Kim et al. 2008) and probabilistic frameworks such as Hidden Markov Models (Kelley, Tavakkoli et al. 2008) and Dynamic Bayesian Networks (Schrempf and Hanebeck 2005). In this paper, we focus on ontology-based approaches.

In many of these efforts, abduction has been used as the underlying reasoning mechanism in providing hypotheses about intentions. In abduction, the system "guesses" that a certain intention could be true based on the existence of a series of observed actions. For example, one could guess that it must have rained if the lawn is wet, though the sprinkler could have been on as well. As more information is learned, probabilities of certain intentions are refined to be consistent with the observations.

One of the key challenges in intention recognition is pruning the space of hypotheses. In a given domain, there could be many possible intentions. Based on the observed actions, various techniques have been used to eliminate improbable intentions and assign appropriate probabilities to intentions that are consistent with the actions performed. Some have assigned weights to conditions of the rules used for intention recognition as a function of the likelihood that those conditions are true (Pereira and Ahn 2009).

There has also been a large amount of research in the Belief-Desire-Intention (BDI) community (Rao and Georgeff 1991). However, this work focuses on the intention of the intelligent agent (as opposed to the human it is observing) and the belief structure is often based on the observation of activities as opposed to inferring the intention of the human via state recognition.

Once observations of actions have been made, different approaches exist to match those observations to an overall intention or goal. (Jarvis, Lunt et al. 2005) focused on building plans with frequency information to represent how often an activity occurs. The rationale behind this approach is that there are some activities that occur very frequently and are often not relevant to the recognition process (e.g., a person cleaning their hands). When these activities occur, they can be mostly ignored. In (Demolombe, Mara et al. 2006), the authors combine probabilities and situation calculus-like formalization of actions. In particular, they not only define the actions and sequences of actions that constitute an intention, they also state which activities cannot occur for the intention to be valid.

All of these approaches have focused on the activity being performed as being the primary basis for observation and the building block for intention recognition. However, as noted in (Sadri 2011), activity recognition is a very difficult problem and far from being solved. There has only been limited success in using RFID (Radio Frequency Identification) readers and tags attached to objects of interest to track their movement with the goal of associating their movement with known activities, as in (Philipose, Fishkin et al. 2005).

Throughout the rest of this paper, we describe an approach to intention recognition that uses state information as opposed to activity information to help address some of the challenges described in this section.

3. INDUSTRIAL ASSEMBLY EXAMPLE

Imagine a situation where a person and a robot are working together to assemble furniture. There are different types of furniture that need to be assembled, and many of the pieces of furniture use the same set of interchangeable parts.

In this example, we will focus on two cabinets, as shown in Figure 1 and Figure 2. The cabinets and their subsequent assemblies were taken from the assembly instructions on the IKEA web site (http://www.ikea.com/ms/en_US/customer_service/assembly_instructions.html). Ten of each cabinet need to be assembled by the end of the shift. The order in which the assembly happens is up to the

human. He may choose to do all of the cabinet 1s first, all of the cabinet 2s first, or intermingle the two.

The robot's goal is to see which assembly the human is trying to accomplish, and then take steps to facilitate that assembly, whether it be handing the human parts or orienting the subassembly to make it easier for the human to complete his task.

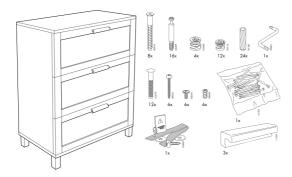Figure 1: Picture of cabinet 1 with some associated parts

Figure 2: Picture of a cabinet 2 with some associated parts

Not knowing which assembly the human is performing at any given time, the robot will observe the sequence of states that occur over time and associate those states with the overall intention of which cabinet is being created. Because many of the parts are common between the two cabinets, simply seeing which part the human picks up is not sufficient. The robot also needs to observe which other parts are used in the assembly and how those parts are attached together.

In the approach described in this paper, the robot will observe the series of states that are the results of various actions and infer the intent of the human by matching the sequence of states to an ontology of intentions with associated state orderings.

Next, we will describe an overall effort that is creating an ontology for robotics and automation and then describe how we are extending this ontology to capture state information.

4 A MANUFACTURING ROBOT ONTOLOGY

In this section we describe current work in developing a standard robotics ontology and how it can be expanded to capture state information.

4.1 IEEE Ontologies for Robotics and Automation Working Group

In late 2011, IEEE formed a working group entitled Ontologies for Robotics and Automation (ORA) (Schlenoff, Prestes et al. 2012). The goal of the working group is to develop a standard ontology and associated methodology for knowledge representation and reasoning in robotics and automation, together with the representation of concepts in an initial set of application domains. The working group understood that it would be extremely difficult to develop an ontology that could cover the entire space of robotics and automation. As such, the working group is structured in such a way as to take a bottom-up and top-down approach to address this broad domain. This group is comprised of four sub-groups entitled: Upper Ontology/Methodology (UpOM), Autonomous Robots (AuR), Service Robots (SeR), and Industrial Robots (InR). The InR, AuR, and SeR sub-groups are producing sub-domain ontologies that will serve as a test case to validate the upper ontology and the methodology developed by UpOM.

The industrial robots group is focusing on manufacturing kitting operations as a test case, which is extremely similar in concept to manufacturing assembly operations. This kitting ontology is focusing on activities that are expected to be performed in a sample kitting operation along with pertinent objects that are expected to be present.

4.2 Expanding the Ontology to Include State Information

The current version of the IEEE Industrial Robots Ontology contains minimal information about states. Initial efforts will look to expand the information that is already represented to include more detailed state information.

A comprehensive literature review was performed in (Bateman and Farrar 2006) which explored the way that spatial information was represented in a number of upper ontology efforts, including Standard Upper Merged Ontology (SUMO), OpenCyc, DOLCE (A Descriptive

Ontology for Linguistic and Cognitive Engineering), and Basic Formal Ontology (BFO). The general findings of the study concluded that, in order to specify the location of an entity, the following four items are needed:

1. A selection of an appropriate granular partition of the world that picks out the entity that we wish to locate
2. A selection of an appropriate space region formalization that brings out or makes available relevant spatial relationships
3. A selection of an appropriate partition over the space region (e.g., qualitative distance, cardinal direction, etc.)
4. The location of the entity with respect to the selected space region description

Item #1: World partition is provided in many manufacturing assembly applications via a parts list. Additional environmental information may be necessary as well.

Item #2: Space region formalization is important in manufacturing assembly, among other reasons, in that it provides a point of reference. When describing spatial relations such as *to_the_right_of*, it provides a point of reference such that all observers are interpreting the relation similarly.

Item #3: Partition of the space region is perhaps one of the most important items to represent in the ontology as it pertains to manufacturing assembly. This is because an assembly operation is based on the ability to combine pieces together to form an overall structure. The location of each piece, whether on a table or attached to other pieces, is key to determining what actions a person has performed and what actions they are likely and able to be performed next.

Item #4 Absolute location of objects in the space regions is perhaps the least important of the four items because the absolute location of objects is often not essential for intention recognition. Note that for controlling robots, the absolute location of objects is extremely important, but that is not the focus of this paper.

4.3 Identifying and Ordering of States to Infer Intention

In just about any domain, there are an extremely large number of states that can occur. However, most of those states are not relevant to determine what activity is occurring. By pre-defining (in the ontology) the activities that are relevant and of interest in the domain, one can then infer the states that are associated with these activities and train the

sensor system to only track and report when those states exist in the environment.

States can be ordered in a similar way as activities to create an overall plan (and therefore an overall intention). In fact, some of the same types of constructs that are used for activities can also be used for states. OWL-S (Web Ontology Language – Services) (Martin, Burstein et al. 2004) is one example of an ontology for describing semantic web services. OWL-S contains a number of constructs for not only representing activities, but also for specifying the ordering of processes. Some of these ordering constructs (as applied to states) include:

- Sequences – a set of states that must occur in a specific order
- Choice – a set of possible states that can occur after a given state
- Join – two or more states that must be true at the same time for a subsequent state to be possible
- Count – a state that needs to be present multiple times. One example of this could be having multiple screws inserted to attach two parts together. Note that this is similar to the iterate construct for processes
- Any-Order – a set of states that must all occur but may occur in any order

All of the constructs that are stated in OWL-S are relevant to states apart from those that deal with concurrency. In this work, states are different than activities in that they are true or false at a given instance of time. Activities occur for a duration, which can cause them to have concurrency constraints, such as starting at the same time (as represented by the OWL-S Split Construct) and/or having to complete at the same time (as represented by the OWL-S Split+Join construct).

An intention is represented as an ordering of states. At the highest level, the overall intention could be to build a specific type of cabinet. This intention can be made up of sub-intentions which could be to build the frame, build the drawers, etc. Each of these sub-intentions would have its own ordering of states which would be a subset of overall intention.

5 APPLYING THE APPROACH TO THE CABINETS EXAMPLE

For the remainder of this paper, we will simplify the assembly operation by using two types of spatial relationships, namely:

$$attached(x,y,z) \qquad (1)$$

which means that part x and part y are attached together by part z, where part z could be a screw, a nail, or any other securing mechanism, and

$$partially\text{-}within(x,y) \qquad (2)$$

which means that an aspect of part x is within part y.

The first step in both cabinet assembly operations is placing a wooden peg into each of four legs, as shown in Figure 3.
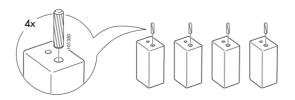

Figure 3: Step 1 of both assembly operations.

The state would be represented by:

$$partially\text{-}within(wooden\_peg, leg) \qquad (3)$$

and the sequence would be:

$$count(partially\text{-}within(wooden\_peg, leg),4) \qquad (4)$$

which indicates that there must be four instances of the state of the wooden peg within a leg. Because this series of states is true for both assemblies, more information is needed for the robot to infer which cabinet the human is assembling.

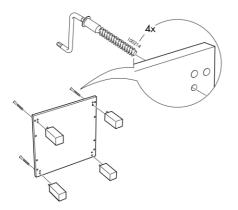The second state of the assembly for the small cabinet in Figure 1 is shown in Figure 4.

Figure 4: Step 2 of the small cabinet assembly.

In this case, the small base is attached to the legs via screw14 (based on Figure 1). Because this has to be performed four times, this would be represented by:

$$count(attached(leg,small\_base,screw14),4) \qquad (5)$$

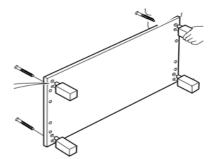The second state of the assembly for the larger cabinet in Figure 2 is shown in Figure 5.

Figure 5: Step 2 of the larger cabinet assembly.

In this case, the large base is attached to the legs via screw14. Because this has to be performed four times, this would be represented by:

$$count(attached(leg,large\_base,screw14),4) \qquad (6)$$

The overall sequence for cabinet 1 and 2 (respectively) up to this point would be:

sequence(
count(partially_within(wooden_peg, leg),4), $\qquad$ (7)
count(attached(leg,small_base,screw14),4))

sequence(
count(partially_within(wooden_peg, leg),4), $\qquad$ (8)
count(attached(leg,large_base,screw14),4))

The type of formalisms shown in (7) and (8) would serve as the basis for the state ordering specification that would be represented in the ontology. Spatial relations such as attached() and partially-within() (such as in equations (1 and 2)) would be represented as subclasses of the general *spatial_relation* class. Specific occurrences of the state (such as in equation (3)) would be represented by instances of the appropriate class. Sequence information would be represented as in OWL-S, by overall *ControlContruct* class, containing subclasses of the appropriate sequence constructs (e.g., count, sequence). As a robot makes observations about the state of the environment, these observations would

be compared to the ontology to find possible state matches. Constraints on state ordering in the ontology will guide the robot's sensor system to areas that should contain the logical next states. With this state information, a robot could track the ordering of observed states over time and compare that observed ordering to predefined state sequences in the ontology to infer the intention of the human.

Though this example is simplistic, it does show the formalism that one could use to represent a sequence of states as a mechanism to perform intention recognition based on state ordering.

6   CONCLUSION

In this paper, we present initial thoughts on a form of intention recognition that is based on states as opposed to actions. State-based intention recognition offers some interesting advantages of activity-based recognition, including:

- States are often more easily recognizable by sensor systems than actions.
- Using activities, intention recognition is often limited to inferring the intention of a single person. State-based intention recognition eliminates this shortfall, in that the state is independent of who created it.
- State information is often more ubiquitous than activity information, thus allowing for reusability of the ontology.

Because of the similarity of state representation with activity representation, many of the same approaches that were described in Section 2 can be applied to this approach as well. Future work will explore explicitly representing which states cannot occur for a subsequent state to be possible as in (Demolombe, Mara et al. 2006) and assigning probabilities to various states similar to (Kelley, Tavakkoli et al. 2008).

DISCLAIMER

Certain commercial software and tools are identified in this paper in order to explain our research. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the software tools identified are necessarily the best available for the purpose.

REFERENCES

Bateman, J. and S. Farrar (2006). Spatial Ontology Baseline Version 2.0. OntoSpace Project Report - Spatial Cognition SFB/TR 8: I1 - [OntoSpace], University of Bremen.

Chabrol, J. (1987). "Industrial robot standardization at ISO." Robotics **3**(2).

Demolombe, R., A. Mara, et al. (2006). Intention recognition in the situation calculus and probability theory frameworks. Computational Logic in Mulit-Agent Systems (CLIMA) Conference**:** 358-372.

Jarvis, P. A., T. F. Lunt, et al. (2005). "Identifying terrorist activity with AI plan-recognition technology." AI Magazine **26**(3): 9.

Jeon, H., T. Kim, et al. (2008). Ontology-based User Intention Recognition for Proactive Planning of Intelligent Robot Behavior. International Conference on Multimedia and Ubiquitous Engineering Busan, Korea**:** 244-248.

Kelley, R., A. Tavakkoli, et al. (2008). Understanding Human Intentions Via Hidden Markov Models in Autonomous Mobile Robots. 3rd ACM/IEEE International Conference on Human Robot Interaction Amsterdam**:** 367-374.

Martin, D., M. Burstein, et al. (2004). "OWL-S: Semantic Markup of Web Services." from http://www.w3.org/Submission/OWL-S/.

Pereira, L. M. and H. T. Ahn (2009). Elder care via intention recognition and evolution prospection. 18th International Conference on Applications of Declarative Programming and Knowledge Management (INAP'09). Evora, Portugal.

Philipose, M., K. Fishkin, et al. (2005). "Inferring ADLs from interactions with objects." IEEE Pervasive Computing.

Rao, A. S. and M. P. Georgeff (1991). Modeling rational agents within a BDI-architecture. Second International Conference on Principles of Knowledge Representation and Reasoning. J. Allen, R. Fikes and E. Sandewall. San Mateo, CA.

Sadri, F. (2011). Logic-Based Approaches to Intention Recognition. Handbook of Research on Ambient Intelligence and Smart Environments: Trends and Perspectives. N.-Y. Chong and F. Mastrogiovanni**:** 346-375.

Schlenoff, C., E. Prestes, et al. (2012). An IEEE Standard Ontology for Robotics and Automation. to be published in Bridges Between the Methodological and Practical Work of the Robotics and Cognitive Systems Communities - From Sensors to Concepts. A. Chibani, Springer-Verlag.

Schrempf, O. and U. Hanebeck (2005). A Generic Model for Estimating User-Intentions in Human-Robot Cooperation. 2nd International Conference on Informatics in Control, Automation, and Robotics ICINCO 05 Barcelona**: 250-256.

Szabo, S., R. Norcross, et al. (2011). "Safety of Human-Robot Collaboration Systems Project." from http://www.nist.gov/el/isd/ps/safhumrobcollsys.cfm.