Single Assembly Robot in Search of Human Partner: Versatile Grounded Language Generation

Ross A. Knepper[†], Stefanie Tellex[†], Adrian Li[‡], Nicholas Roy[†], and Daniela Rus[†]
[†] MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA
Email: {rak,stefie10,nickroy,rus}@csail.mit.edu
[‡]Department of Engineering, University of Cambridge, Cambridge, UK
Email: alhl2@cam.ac.uk

***Abstract*—We describe an approach for enabling robots to recover from failures by asking for help from a human partner. For example, if a robot fails to grasp a needed part during a furniture assembly task, it might ask a human partner to "Please hand me the white table leg near you." After receiving the part from the human, the robot can recover from its grasp failure and continue the task autonomously. This paper describes an approach for enabling a robot to automatically generate a targeted natural language request for help from a human partner. The robot generates a natural language description of its need by minimizing the entropy of the command with respect to its model of language understanding for the human partner, a novel approach to grounded language generation. Our long-term goal is to compare targeted requests for help to more open-ended requests where the robot simply asks "Help me," demonstrating that targeted requests are more easily understood by human partners.**

Categories and subject descriptors: I.2.7 [Artificial intelligence]: Natural Language Processing—Language generation; I.2.9 [Artificial intelligence]: Robotics

General Terms: Algorithms

Keywords: robots; furniture assembly; natural language

I. Introduction

Failures are inevitable in complex robotic systems. Engineers may design contingency plans that enable robots to handle anticipated failures in perception or manipulation, but humans remain more adept at dealing with complex or unforeseen errors. We describe an approach to enable robots to recover from failures during complex tasks: when the robot encounters failure, it asks for assistance from a human partner. After receiving assistance, it continues to execute the task autonomously.

To implement this strategy, the robot must first be able to detect its own failures and identify a strategy to recover from them. For strategies involving a human partner, it must next communicate this strategy to the human. Finally, it must detect when the human has successfully or unsuccessfully provided help to the robot, in order to plan its next actions. When articulating its help request, the robot generates a natural language description of the needed action. Because the robot might need help in different contexts and situations, a pre-specified template-based set of help requests is inadequate.

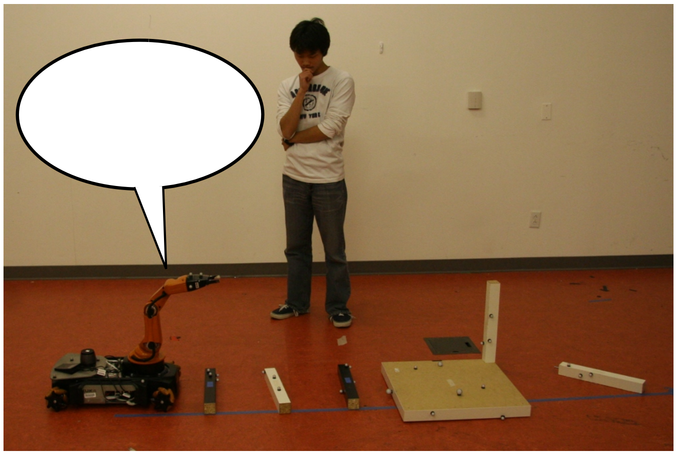

Fig. 1. When a robot needs human help with an assembly task, effective communication requires intuitive formulation of the request. Simple canned expressions like "Help me" or "Hand me white_leg_2." fail to exploit available information that could disambiguate the request. By integrating a natural language model, the robot is able to effectively communicate context-appropriate relations, such as "Hand me the white table leg near me."

We present a new algorithm for generating a natural language request for help by searching for an utterance that maximizes the probability that the person will successfully follow the request, making use of a computational model of a person's language understanding faculty. Our aim is to demonstrate that by modeling and bounding the probability of a human misinterpreting the request, the robot is able to generate targeted requests that work better than baselines involving either generic requests (e.g., "Help me") or template-based non-context-specific requests. We are able to compute the probability of error by inverting our previous work, which used language to generate a graphical model that allowed us to infer the semantic meaning of the language in the robot's frame of reference. In the present work, we utilize the semantic meaning of what we want the person to do to search for the joint graphical model and language that minimizes the probability of error of the correct plan being executed.

II. Furniture Assembly

As a test domain, we focus on the problem of assembling Ikea furniture. We assume the robot has a pre-existing model of the piece to be assembled and a plan for assembling it. However, due to perceptual and mechanical failures or to plan inadequacy, aspects of the plan might be difficult or impossible to carry out. For example, in Figure 1, the robot may have failed to grasp the white table leg because of an inability to approach it from the correct angle due to obstacles. However, using its model of what needs to be done, the robot requests help from its human partner. After receiving help, the robot continues to execute the task.

A procedure derived from a symbolic planner guides the automated furniture assembly, as in the work of Knepper et al. [2]. Each step of the plan comprises an action (pick up the table leg), a set of preconditions (hand is empty, robot can reach the table leg), and a set of postconditions (table leg is in hand). At plan time, a sequence of steps is discovered such that executing the actions in turn causes the postconditions to become true. At each step, each precondition is satisfied by either an initial condition or an earlier action's postcondition. If the symbolic problem is properly specified and the robot properly executes every step, then the procedure results in a completely assembled piece of furniture.

Of course, real robots do not always succeed on the first try, due to myriad causes. A failure can be detected by comparing the pre- or postconditions to the state of the real world. Any mismatch—for example the robot hand is not holding the specified table leg—triggers a recovery routine. Recovery can be accomplished by rerunning the symbolic planner from the current real world state. Most often, this strategy results in retrying the failed step. However, this solution is not complete due to unforeseen circumstances outside of the robot's control. For example, in Fig. 1 the white table leg is blocked by the black table legs. If the robot programmer neglected to include a contingency for moving table legs out of the way, the robot can still accomplish its goal by asking a human for help.

III. Asking For Help From a Human Partner

In order to be understood, a robot asking for help should follow the Gricean maxims [1], which dictate that communication should be truthful, relevant, clear, and should provide as much information as is needed but no more. This last maxim, especially, guides our solution approach.

The simplest possible request for help (i.e. "Help me.") may fail to result in the needed assistance from many human users because it is unclear what help is needed. A more sophisticated approach to asking for help might include a template request for each failed condition, with details filled in. An example template would be "Hand me the _____."

Such templatized approaches are likely to be most effective when all items possess unique, well-known names and appearances. In the context of furniture assembly, parts are often unnamed, and many parts look similar. In such situations, programmers may have difficulty encoding all possible situations and relations among parts in order to generate appropriate requests. Thus, we propose a more flexible framework of referring to concepts by means that either disambiguate among similar parts ("Hand me the white table leg that is near me.") or else make clear that any member of a class of parts is acceptable ("Hand me any white table leg that is unattached.").

IV. Language Generation

We generate language using the Generalized Grounding Graph (G^3) framework of Tellex et al. [4]. That work performs sentence understanding on human requests by inferring a set of groundings corresponding to phrases broken up according to the grammatical structure of the sentence. For instance, "Give me the table leg" would yield groundings corresponding to "me" – a location in space, "the table leg" – another location, and"give" – a path in space from the table leg to me. In order to find a set of groundings that best correlates to the given phrases, a corpus of annotated examples is provided as training data. From these groundings, a robot is able to execute an action that satisfies the human's request.

In the present work, we invert the process by searching over grounding graphs (i.e. sentence structures) and language elements that appropriately describe the groundings, which in this case are provided by the robot's failed condition. This process attempts to satisfy the Gricean maxims of truthfulness and relevance. However, even a sentence that correlates to the groundings with high probability may be ambiguous. In order to achieve clarity and appropriate information, we model human understanding by feeding candidate sentences back through the understanding algorithm, which returns a probability distribution over possible groundings. Tellex et al. [5] select a specific node in the grounding graph with high entropy and generate targeted clarifying questions to increase certainty on that term. In the current work, we instead examine the probability of getting back the correct grounding. The search is successful when a sentence is discovered that yields a sufficiently high probability of correct interpretation by the human helper. To keep the request succinct, we search sentence structures in sorted order from the simple to the complex.

V. Conclusion

This work represents a step toward the goal of mixed-initiative human-robot cooperative assembly. Much prior work in this area focuses on robots meeting human expectations [3]. In this work, we instead focus on setting human expectations of the robot through clear communication.

We demonstrate the capability to both understand and generate natural language commands in the context of furniture assembly—a domain in which many actions and objects are confusingly similar. The careful selection and interpretation of language within the spatial context of the problem reliably disambiguates such requests.

References

[1] H. P. Grice. *Logic and Conversation'in P. Cole and J. Morgan (eds.) Syntax and Semantics Volume 3: Speech Acts.* Academic Press, New York, 1975.

[2] R. A. Knepper, T. Layton, J. Romanishin, and D. Rus. IkeaBot: An autonomous multi-robot coordinated furniture assembly system. In *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, Karlsruhe, Germany, in submission.

[3] S. Nikolaidis and J. Shah. Human-robot teaming using shared mental models. In *IEEE/ACM International Conference on Human-Robot Interaction, Workshop on Human-Agent-Robot Teamwork*, Boston, MA, Mar 2012.

[4] S. Tellex, T. Kollar, S. Dickerson, M.R. Walter, A. Banerjee, S. Teller, and N. Roy. Understanding natural language commands for robotic navigation and mobile manipulation. In *Proc. AAAI*, 2011.

[5] S. Tellex, P. Thaker, R. Deits, T. Kollar, and N. Roy. Toward information theoretic human-robot dialog. In *Proceedings of Robotics: Science and Systems*, Sydney, Australia, July 2012.