AAAI Technical Report FS-12-07
Robots Learning Interactively from Human Teachers

Using Spatial Language to Guide and Instruct Robots in Household Environments

Juan Fasola and Maja J Matarić
Interaction Lab
Computer Science Department
University of Southern California
Los Angeles, CA 90089
{fasola, mataric}@usc.edu

Abstract

We present an approach for enabling in-home service robots to follow natural language commands from non-expert users, with a particular focus on spatial language understanding. Specifically, we propose an extension to the semantic field model of spatial prepositions that enables the representation of dynamic spatial relations involving paths. The relevance of the proposed methodology to interactive robot learning is discussed, and the paper concludes with a description of how we plan to integrate and evaluate our proposed model with end-users.

Introduction

In order for autonomous service robots to become ubiquitous in household environments, they will need to be capable of interacting with and learning from non-expert users in a manner that is both natural and practical for the users. In particular, robots will need to be capable of understanding natural language instructions in order to learn new tasks and receive guidance and feedback on task execution from non-technical users. This necessity is especially evident in assistive contexts, where robots are interacting with people with disabilities, age-related (e.g., reduced mobility, limited eyesight) or otherwise (e.g., individuals post-stroke), as the users may not be able to teach the robot new tasks and/or provide feedback by demonstration.

We present an approach for enabling in-home service robots to follow natural language commands from non-expert users, with a particular focus on spatial language understanding.

Spatial Language in Human-Robot Communication

Service robots designed to interact with non-expert users in household environments would benefit from being endowed with a certain set of primitive actions and/or tasks. Furthermore, enabling these robots with a knowledge base of commonly used words that refer to the specific tasks or actions they can perform, would greatly facilitate spoken language communication and understanding with the user. For example, consider the following:

(1) Go to the kitchen

If the user says (1), the robot should understand, in principle, what that means. That is, it should understand which task among those within its task/action repertoire the user is referring to. In this example, the robot may not know where the kitchen is located in the user's specific home environment, but it should be able to understand that (1) expresses a command to physically move to a desired goal location that fits the description of the noun phrase "the kitchen".

Spatial language plays an important role in instruction-based natural language communication. In (1), the spatial preposition "to" was used in the instruction of the task. The following sentence contrasts minimally in the preposition employed, using "away from" (compound preposition) instead of "to":

(2) Go away from the kitchen

Yet, the meaning of the command specified by (2), or rather the implied goal task/action sequence, is completely different from (1), even though the verb and place noun are the same. The same holds for the spatial prepositions “around”, “through”, “behind”, etc. Spatial relations expressed by language are often expressed by prepositions (Landau and Jackendoff 1993). Therefore, the ability for in-home service robots to understand and differentiate between spatial prepositions in spoken language is crucial for the interaction to be successful.

Spatial language understanding is especially relevant for interactive robot task learning and task modification. For example, the human might teach the robot the complex task “Clean up the room”, by using only natural language and specifying the subgoals of that task individually, each represented by its own spatial language instruction (e.g., “Put the clothes in the laundry basket”, “Stack the books on top of the desk in the right-hand corner”, “Put all toys underneath the bed”, etc.). In addition, user modification of known robot tasks can also readily be accomplished with the use of spatial language. For example, the user might modify the task defined by (1) (i.e., robot movement to “the kitchen”), by providing spatial constraints, or rules, for the robot to obey during task execution, such as “Don't go through the hallway,” or “Move along the wall.” These user-defined constraints do not change the meaning of the underlying task, but rather allow the user to interactively modify the manner in which the robot executes the task for the specific instance. Finally, spatial language can be used to provide teacher feedback during task execution, to further correct or guide robot behavior. Within the context of the same example, as the robot is moving along the wall en route to the kitchen, the user may provide additional feedback by saying “Move a little further away from the wall,” or perhaps “Move forward close to the wall but stay on the center rug”. These examples illustrate the importance of spatial language in the instruction and teaching of in-home service robots by non-expert users. Our methodology for spatial language understanding was developed to address these fundamental research issues.

Related Work

Previous work that has investigated the use of spatial prepositions, and spatial language in general, includes Skubic et al. (2004), who demonstrated a robot capable of understanding static spatial relations (e.g., “to the right”, “in front of”, etc.) in natural language instruction and producing such spatial language in information passing. Sandamirskaya et al. (2010) investigated the use of Dynamic Neural Fields theory in a static spatial language architecture for use in human-robot cooperation in object manipulation tasks on a tabletop workspace. Both of these works implemented pre-defined notions of spatial relations, however researchers have also investigated learning these types of static spatial relations automatically from training data (e.g., Roy 2002; Chao, Cakmak, and Thomaz 2011). Our work aims to extend upon this related work by encoding not only static spatial relations in robot language understanding, but also dynamic spatial relations involving paths, discussed in more detail in the next section.

In the context natural language robot instruction, however, the use of dynamic spatial relations has in fact been explored by recent work. Tellex et al. (2011) developed a probabilistic graphical model to infer task/actions for execution by a forklift robot from natural language commands. Kollar et al. (2010) developed a Bayesian framework for interpreting route directions on a mobile robot, using learned models of dynamic spatial relations such as “past” and “through” from schematic training data. In both of these works there was no explicit definition of the spatial relations used, static or otherwise, and instead they were learned from labeled training data. However, these approaches typically require the programmer to provide an extensive training data set of natural language input for each new application context, without taking advantage of the domain-independent nature of spatial prepositions. Our proposed approach aims to develop novel, pre-defined templates for spatial relations, static and dynamic, that facilitate use and understanding across domains, and whose computational representations enable guided robot execution planning, discussed further in the next section.

Approach and Methodology

We propose a methodology for human-robot dialogue and natural language instruction understanding for household service robots. Specifically, we aim to extend the semantic field model of spatial prepositions, proposed by O’Keefe (2003), to include dynamic spatial relations and to develop a computational framework for human-robot interaction which integrates the proposed model. The *semantic field* of a spatial preposition is analogous to a probability density function (pdf), parameterized by schematic figure and ground objects, that assigns weight values to points in the environment depending on how accurately they capture the meaning of the preposition (e.g., points closer to an object have higher weight for the preposition ‘near’). While appropriate for static relations, this method is not sufficient for dynamic spatial relations that involve paths. *Paths* are comprised of a set of points connected by direction vectors that define sequence ordering. To account for paths in the spatial representation of prepositions, we propose to add a weighted vector field to

Figure 1. The PR2 mobile robot platform

each point in the traditional semantic field model. As an example, the preposition “along” denotes not only proximity, but also a path parallel to the border of a ground object. Thus, in the proposed model, the semantic field for “along” would contain not only weights for each point in the environment to encapsulate proximity, but also weighted direction vectors at each point (the more parallel to the ground object the higher the weight). The advantage of modeling spatial relations as pdfs, as opposed to using classification-based methods (e.g., Kollar et al. 2010), is that generating robot action plans for instruction following is as simple as sampling the pdf; there is no need to search the action space randomly for appropriate solutions, which may be prohibitive in time-complexity. Furthermore, user teaching, feedback, and refinement of the robot task execution plan can easily be incorporated as an alteration of the pdf. For example, the feedback statement “Move closer to the wall” could alter the pdf by attributing higher weight to points closer to the wall from the robot’s current location.

We plan to validate our methodology with a study with non-expert users engaging with the PR2 mobile robot platform (shown in Figure 1, above) in a lab-simulated household environment. Participants will instruct the robot to perform a variety of tasks from the robot task repertoire involving spatial relations (e.g., robot movement, object placement tasks, object search, etc.). The problem formulation is as follows: the robot must infer the most likely task type/command given the natural language input. Our subsequent work will focus on human teaching of new robot tasks using spatial language instructions and modification of known tasks through dialogue. Performance of the system will be evaluated according to both subjective (post-session surveys) and objective evaluation measures (quantifiable goal achieving metrics).

Summary

We have described the need for enabling in-home service robots with spatial language understanding to facilitate natural communication with non-expert users for interactive task teaching, task modification, and user feedback on robot task execution, and have presented a general approach we are developing toward addressing this research challenge.

References

Landau, B., and Jackendoff, R. 1993. What and where in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2):217–265.

O’Keefe, J. 2003. Vector grammar, places, and the functional role of the spatial prepositions in English, E. van der Zee and J. Slack, Eds. Oxford: Oxford University Press.

Tellex, S.; Kollar, T.; Dickerson, S.; Walter, M. R.; Banerjee, A. G.; Teller, S.; and Roy, N. 2011. Approaching the Symbol Grounding Problem with Probabilistic Graphical Models. *AI Magazine*. 32(4): 64-76.

Kollar, T.; Tellex, S.; Roy, D.; and Roy, N. 2010. Toward Understanding Natural Language Directions. In Proc. ACM/IEEE Int’l Conf. on Human-Robot Interaction (HRI), 259–266.

Sandamirskaya, Y.; Lipinski, J.; Iossifidis, I.; and Schöner, G. 2010. Natural human-robot interaction through spatial language: a dynamic neural field approach. 19th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 600-607.

Skubic, M.; Perzanowski, D.; Blisard, S.; Schultz, A.; Adams, W.; Bugajska, M.; and Brock, D. 2004. Spatial Language for Human-Robot Dialogs. *IEEE Transactions on SMC Part C*, Special Issue on Human-Robot Interaction, 34(2):154-167.

Chao, C.; Cakmak, M.; and Thomaz, A.L. 2011. Towards grounding concepts for transfer in goal learning from demonstration. In Proceedings of the Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob).

Roy, D.K. 2002. Learning visually grounded words and syntax for a scene description task. In *Computer Speech & Language*. 16(3-4): 353-385.