Imitation Learning for Natural Language Direction Following

Alexander Grubb* AGRUBB@CMU.EDU
Felix Duvallet* FELIXD@CMU.EDU
Stefanie Tellex† STEFIE10@CSAIL.MIT.EDU
Thomas Kollar† TKOLLAR@MIT.EDU
Nicholas Roy† NICKROY@MIT.EDU
Anthony Stentz* TONY@CMU.EDU
J. Andrew Bagnell* DBAGNELL@RI.CMU.EDU

*Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA

†Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA

Abstract

Using natural language directions to generate plans that can be executed by robots requires an understanding of the environment, the objects within it, and the structure of language. Building and training the complex systems that can do this is difficult.

We formulate this problem as an instance of imitation learning and learn a cost function from demonstrated examples of people following directions. This cost function is then used to produce the desired direction following behavior from an optimal controller or planner, providing a straightforward and principled way to reproduce human behavior. Furthermore, unlike the previous supervised approaches, this method allows for the learning of policies, not just specific plans, enabling operation in cases such as those where the map is not known *apriori*.

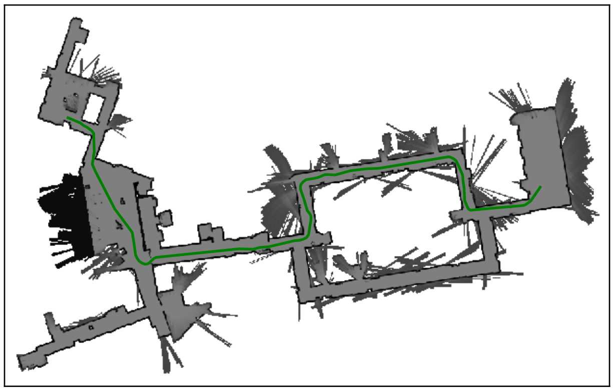

Figure 1. Example path for direction text 'Start with back to long wall of windows. Walk through door on the farthest left side of opposite wall (says "rle"). Take right at photocopier, stay left at refrigerator, take left at end of hall and right at intersection. Go straight through several sets of doors until you are in a large open area facing windows. Head forward and to your right, beyond the spiral staircase, towards some cubbies. Go straight down that hall, turning left when you see orange couches.'

1. Introduction

Robots that can be commanded through spoken instructions are a critical component to effective human-robot teams. Spoken directions are an intuitive way for users to interact with robots which requires no extensive training or specialized interfaces. As such, enabling robots to understand and follow natural language instructions is a key challenge in robotics.

As an example, a robot could be tasked to deliver a package through directions such as the ones shown in Figure 1. Though robots are very good at executing planned actions, generating those plans automatically and without imposing a high cognitive load on the user is a difficult problem.

Previous work (Kollar et al., 2010a;b) has used a probabilistic graphical model approach coupled with supervised learning to build direction following systems. We instead formulate this as an inverse optimal control problem, where the goal is to learn a cost function which, when coupled with an optimal controller for the domain, will generate a path through the environment that matches the spoken directions.

Appearing in *Proceedings of the ICML Workshop on New Developments in Imitation Learning*, Bellevue, WA, USA, 2011. Copyright 2011 by the author(s)/owner(s).

- **Verb:** Go
- **Spatial Relation:** through
- **Landmark:** the door near the elevators.

Figure 2. SDC for the sentence, "Go through the door near the elevators."

In contrast to the graphical model approach, using the inverse optimal control approach can learn behaviors based on a wide variety of optimal controllers. For example, in the case where the robot is given a complete model of the world, a reasonable controller is a planner which searches over all possible paths for the minimum cost option. In the case of incomplete world information the same technique can be coupled with a controller which instead generates policies corresponding to the direction commands. By learning a policy instead of a specific plan, the robot can adapt its behavior to the actual state of the world as it executes the natural language directions.

2. Previous Work

2.1. Spatial Description Clauses

Following previous work (Tellex, 2010), we use *spatial description clauses* (SDCs) to break down spatial language discourse into simpler subcomponents. Each SDC consists of a figure, a verb, a spatial relation, and a landmark, as shown in Figure 2.

2.2. Supervised Learning Approach

Using the language and structural information extracted from natural language directions in the form of SDCs, previous work (Kollar et al., 2010a) has used a probabilistic graphical model approach for selecting paths corresponding to natural language directions. More concretely, this work uses a graphical model to estimate the probability of seeing a particular path, given the natural language command (in the form of SDCs), the features of the world (in the form of a map of objects), and allowable paths. Using this learned graphical model, the most likely path is selected as the path corresponding to the commanded direction.

Their work learns the given graphical model by factorizing the probability distribution over paths into a number of separate components, such as a factor which measures the probability that a specific path geometry corresponds to the spatial relation keywords in a given text, or a component which estimates the probability that the objects seen by a specific path correspond to the landmark descriptions in the text. Each of these components is then individually trained using a separate labeled training set. For example, the path geometry component is trained on datasets consisting of spatial labels such as 'towards' and 'around' and sample path geometries for each label.

More recently, Kollar et al. (2010b) have introduced a conditional random field (CRF) approach to estimating the conditional probabilities of paths in environment, given the natural language commands, path and world geometry, and groundings of objects in the extracted SDCs to real world objects. This model can be trained given supervised labelings of all the necessary variables using standard CRF techniques.

3. Imitation Learning Approach

In contrast to the previous supervised learning approaches, we frame the problem as an imitation learning problem using the LEARCH framework of Ratliff et al. (2009) for inverse optimal control. In this approach we use demonstrated behaviors to learn a cost function which, when used with an optimal controller, results in similar robot behavior.

In the case where the environment is known *apriori* this imitation learning formulation has a simple formulation. Given a piece of direction text parsed into a sequence of SDCs $(\mathrm{sdc}_1, \ldots, \mathrm{sdc}_n)$, the optimal control (planning) problem in this domain is to find a path ξ made up of segments $(\xi_1, \ldots, \xi_n)$ where each segment ξ_i corresponds to sdc_i. This optimal planner selects the path $\hat{\xi}$ which minimizes some cost function c evaluated over each segment:

$$\hat{\xi} = \arg\min_{\xi \in \Xi} \sum_{i=1}^{N} c(\phi(\xi_i, \mathrm{sdc}_i))$$

where Ξ is the set of all segmented paths and $\phi(\xi_i, \mathrm{sdc}_i)$ is a vector of features corresponding to that particular path segment and the parsed text.

Given a demonstrated behavior in the form of a segmented path ξ^*, we now wish to find the cost function c which best reproduces the demonstration. Using the behavior of the optimal planner above, we can then formulate the LEARCH objective function, which corresponds to the difference between the cost of the planned path and the demonstrated path:

$$\mathcal{F}[c] = \sum_{i=1}^{N} c(\phi(\xi_i^*, \mathrm{sdc}_i)) - \min_{\xi \in \Xi} \sum_{i=1}^{N} c(\phi(\xi_i, \mathrm{sdc}_i))$$

This objective function is convex in c, so we can use

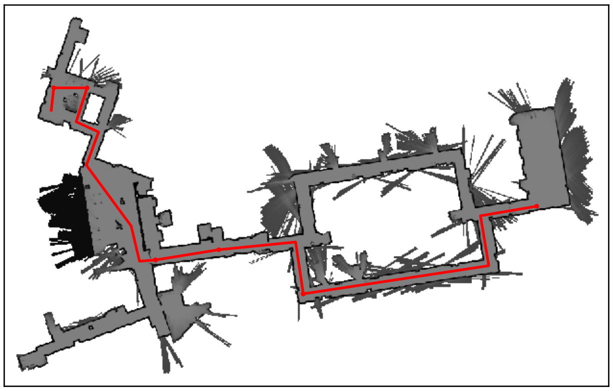

Figure 3. Resulting planned behavior for the directions in Figure 1, using only landmark information in the parsed SDCs and objects present in the environment for cost function evaluation.

standard convex optimization techniques to find the optimal cost function $c^* = \arg\min_c \mathcal{F}[c]$ for a wide variety of different cost function representations.

For cost functions c parametrized as linear sums of the features, this model generalizes the previous CRF work (Kollar et al., 2010b; Tellex et al., 2011), where the cost is the log-probability computed by the CRF. However, this approach can also extend to more general cost functions and offers a number of new advantages. The previous work must be trained using difficult to generate negative examples, while our approach need not. Our work also allows for approximate planners and, as we will discuss shortly, can even be used to learn policies which can react to unexpected information instead of just learning open-loop plans.

Preliminary results show that this imitation learning approach competes with the supervised approaches used previously. An example path planned by the system is shown in Figure 3. Here the features used in the cost function are entirely based on landmark information found in the direction text from Figure 1 and the semantic tags given to objects in the world. This particular path has low cost due to the pairing of the direction text and objects in the world (*e.g.*, the direction text "orange couches" and an object in the world labeled "sofa")

3.1. Direction Following Without a Map

When a semantically-labeled map of the environment is not present, enumerating paths to perform inference is impossible. Previous approaches have used a greedy local evaluation of visible objects. This approach is more realistic, as a robot navigating an unknown environment is unlikely to have previous knowledge of a complete map. However, this approach only considers two SDCs at a time and disregards any information potentially contained in the complete directions.

One current research direction is to learn *policies* directly from demonstrations, to enable the robot to reason about what it might see in the future and use the entire set of SDCs. Imitation learning is especially applicable in this setting as people are quite good at following directions, even in buildings they have not seen before.

An additional benefit of learning policies is the ability to give directions that include a way to recover from mistakes. For example, consider the direction "go down the hall and take a right, if you've reached the elevators you've gone too far." Being able to follow these directions requires the ability to back-track and reason about potential failures, neither of which the current planning-based system can handle.

4. Discussion

We believe inverse optimal control is a good approach for this domain, as it mirrors the way humans naturally solve the same problem. Furthermore, collecting demonstrations is straightforward, and they can be used in a principled manner to train the system. Learning policies directly is a promising direction for this work, which will enable us to reason about unseen portions of the environment and plan for contingencies.

References

Kollar, T., Tellex, S., Roy, D., and Roy, N. Toward understanding natural language directions. In *International Conf. on Human-Robot Interaction*, 2010a.

Kollar, T., Tellex, S., and Roy, N. A discriminative model for understanding natural language route directions. In *AAAI Fall Symposium Series*, 2010b.

Ratliff, N., Silver, D., and Bagnell, J. A. Learning to search: Functional gradient techniques for imitation learning. *Autonomous Robots*, July 2009.

Tellex, S. *Natural Language and Spatial Reasoning*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2010.

Tellex, S., Kollar, T., Dickerson, S., Walter, M. R., Banerjee, A. G., Teller, S., and Roy, N. Understanding natural language commands for robotic navigation and mobile manipulation. In *Conference on Artificial Intelligence*, 2011.