# A Discriminative Model for Understanding Natural Language Route Directions

**Thomas Kollar**
MIT CSAIL
32 Vassar St, 32-331
Cambridge, MA 02139
Email: tkollar@mit.edu

**Stefanie Tellex**
MIT Media Lab
75 Amherst St E14-574M
Cambridge, MA 02139
Email: stefie10@media.mit.edu

**Nicholas Roy**
MIT CSAIL
32 Vassar St, 32-330
Cambridge, MA 02139
Email: nickroy@mit.edu

## Abstract

To be useful teammates to human partners, robots must be able to follow spoken instructions given in natural language. However, determining the correct sequence of actions in response to a set of spoken instructions is a complex decision-making problem. There is a "semantic gap" between the high-level symbolic models of the world that people use, and the low-level models of geometry, state dynamics, and perceptions that robots use. In this paper, we show how this gap can be bridged by inferring the best sequence of actions from a linguistic description and environmental features. This work improves upon previous work in three ways. First, by using a conditional random field (CRF), we learn the relative weight of environmental and linguistic features, enabling the system to learn the meanings of words and reducing the modeling effort in learning how to follow commands. Second, a number of long-range features are added, which help the system to use additional structure in the problem. Finally, given a natural language command, we infer *both* the referred path and landmark directly, thereby requiring the algorithm to pick a landmark by which it should navigate. The CRF is demonstrated to have 15% error on a held-out dataset, when compared with 39% error for a Markov random field (MRF). Finally, by analyzing the additional annotations necessary for this work, we find that natural language route directions map sequentially onto the corresponding path and landmarks 99.6% of the time. In addition, the size of the referred landmark varies from $0m^2$ to $1964m^2$ and the length of the referred path varies from $0m$ to $40.83m$.

## 1 Introduction

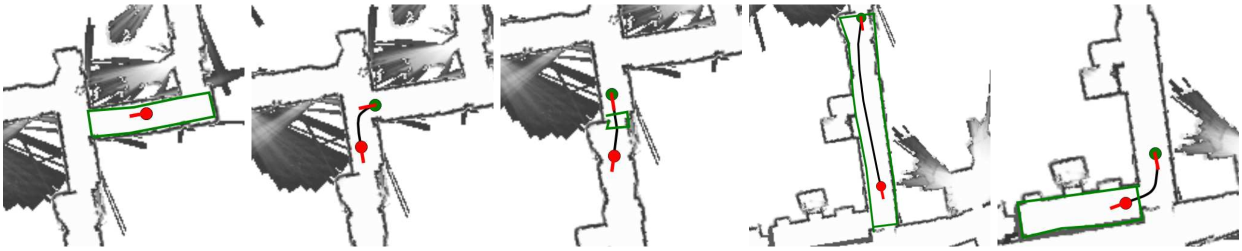
In order to achieve higher levels of autonomy, robots need the ability to interact naturally with humans in unstructured environments. One of the most intuitive and flexible interaction modalities is to command robots using natural language. To follow natural language directions such as "Go down the hallway, take a right, and go into a lounge with some couches," a robot must convert the symbolic natural language instructions to low level actions and observations that correspond to the desired motion through the environment. Associated with this language grounding problem is an inherent lack of complete information; the robot may only know about a few objects or landmarks in the environment, making it necessary to match unknown parts of the language to landmarks that the robot can already detect. In addition it must ground a variety of words such as, "right," "left," "through," and "past" in motions and paths through the environment.

In previous work Kollar et al. (2010) formulated the problem of understanding route directions as an MRF, parsing the language it into its component parts and inferring the path that corresponded to the language. Spatial relations, such as "past" or "through", were modeled using naive Bayes: learning the probability of a spatial relation such as "past" given a set of features. Each unknown component landmark, such as "monitors" or "refrigerator", was probabilistically grounded using Flickr co-occurrences that relate the language in the command to observed objects in the environment.

The technical contribution of this work is four-fold. First, we formulate the problem of understanding natural language commands as a conditional random field (CRF), which enables the system to learn the mapping from novel words onto the referred path through the environment and the referred landmarks. Secondly, the CRF, unlike the MRF, is able to learn the relative weights of long-range word-dependent features. Third, the CRF reduces the amount of modeling effort necessary when compared with the MRF in Kollar et al. (2010), not requiring the creation of a new training dataset each time a new word, such as a spatial relation, is added. Finally, we formulate the CRF so that both the referred landmarks and referred partial paths are inferred, instead of marginalizing over the set of landmarks as in Kollar et al. (2010). For example, with a command such as "go past the bathroom", the robot will infer a polygon that describes "bathroom", and a path that corresponds to "past". The CRF is demonstrated to have 15% error on a held-out dataset, when compared with 39% error for a Markov random field (MRF). Finally, by analyzing the additional annotations necessary for this work, we find that natural language route directions map sequentially onto the corresponding path and landmarks 99.6% of the time. In addition, the size of the referred landmark varies from $0m^2$ to $1964m^2$ and the length of the referred path varies from $0m$ to $40.83m$.

## 2 Approach

To address the challenge of understanding natural language commands, there are two key insights. The first is to decompose a natural language command into a sequence of spatial description clauses (SDCs), which corresponds to a component sequence of actions that the robot should follow (Kollar et al., 2010). Each spatial description clause (SDC) consists of a figure, verb, spatial relation and landmark. For example, the command "Go down the hallway,"



(a) v:starting sr:in l:this hall

(b) v:turn left

(c) v:walk sr:through l:this metal door

(d) v:walk sr:straight down l:the hall

(e) v:turn right sr:into l:the hall in front of the elevators

Figure 1: An example from our corpus: spatial description clauses in each subfigure are paired with a path and landmark. The start location of the partial path is shown in green, the end location is shown in red and the corresponding landmark is a green polygon when it exists.

consists of a single SDC: the figure is implicitly "(you)", the verb is "go," the spatial relation is "down", and the landmark is "the hallway". Longer commands consist of a sequence of SDCs, which apply to different parts of the path that the robot should follow.

The second key insight is to model the problem of inferring a path from natural language commands as a cost function which, when optimized, yields the desired behavior of the robot. In order to do this, we break down the language into SDCs and a path through the environment into partial paths, each of which refers to a particular landmark. Thus, the natural language, given as text input, is automatically decomposed into a sequence of $n$ SDCs sdc$_i$. A path $r$, represented as a sequence of robot poses in global coordinates is broken down into a corresponding sequence of (potentially overlapping) partial paths $r_i$ and corresponding landmarks $l_i$.

Thus, our training dataset has training examples that consist of a sequence of partial paths $r_i$, landmarks $l_i$ and SDCs sdc$_i$, as seen in Figure 1. The partial path $r_i$ corresponds to a sequence of poses between the green robot pose and the red robot pose, the landmark $l_i$ corresponds to a polygon, and sdc$_i$ corresponds to the text in the description. $\phi_i$, the output variable, is Boolean and determines if the $i$th partial path and landmark corresponds to $i$th SDC. When the text corresponds to the partial path and landmark (as in each element of Figure 1(a-e)), then $\phi_i$ = True. When the partial path and landmark does not correspond to the language, then $\phi_i$ = False. $\Theta$ are the parameters of the model. Thus, the goal of inference is to minimize the cost function $C$:

$$\underset{r_1 \ldots r_n, l_1 \ldots l_n}{\text{argmin}} \quad C(r_1 \ldots r_n, l_1 \ldots l_n | sdc_1 \ldots sdc_n; \Theta)$$

$$\text{where} \quad C(r_1 \ldots r_n, l_1 \ldots l_n | sdc_1 \ldots sdc_n; \Theta) \triangleq$$
$$- \, log(p(\phi_1 \ldots \phi_n | r_1 \ldots r_n, l_1 \ldots l_n, \text{sdc}_1 \ldots \text{sdc}_n; \Theta))$$

A full example of a sequence of partial paths along with the corresponding landmarks is shown in Figure 1. The goal of learning is then to estimate the model parameters $\Theta$ from a training dataset of sequences of SDCs, partial paths, and landmarks. The goal of inference is to take as input a sequence of SDCs and pick the best sequence of partial paths and landmarks in the environment that maximizes the probability that the two correspond.

## 3 Corpus

For the purpose of understanding what people might want to tell their robots, we have used a corpus collected as part of Kollar et al. (2010). In it, Kollar et al. (2010) performed an extensive study on how people give route directions, collecting a corpus of natural language route directions through an office environment in two adjoining buildings at MIT. They asked fifteen subjects to write directions between 10 different starting and ending locations, for a total of 150 directions in each environment. Subjects were solicited by placing fliers around MIT and were selected for inclusion in the study if they were between the ages of 18 and 30 years old, were proficient in English, and were unfamiliar with the test environment. The pool was made up of 47% female and 53% male subjects from the MIT community, primarily students or administrators. When collecting directions, subjects were first given a tour of the building to familiarize themselves with the environment. Then subjects were asked to write down directions from one location in the space to another, as if they were directing another person. Subjects were allowed to wander around the floor as they wrote the directions and were not told that this data was for a robotics research experiment.

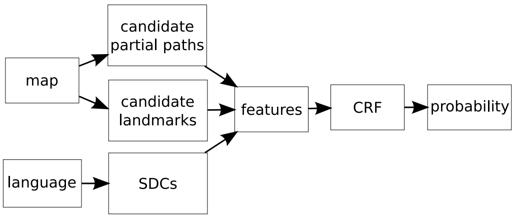


Figure 2: An overview of the system architecture.

In this work, we extended the dataset by annotating each command with the corresponding path stated in the natural language command. In addition, we annotated each natural language command with a sequence of SDCs and the corresponding partial paths and landmarks that were referenced in each SDC, as shown in Figure 1. This gives us a dataset from which we can both learn how to follow natural language commands and also to understand geometrically the types of landmarks and paths that people use in route directions. The results of this effort are discussed in Section 5. In addition to annotating the correct SDCs, partial paths, and landmarks for the 150 directions, paths that do *not* correspond to the SDCs stated in the command were annotated as well, resulting in a dataset of 300 commands. Half of this dataset corresponds to good paths and landmarks and half of it corresponds to bad paths and landmarks for a given sequence of SDCs. Statistics about the partial paths can be seen in Table 2 and Figure 5.

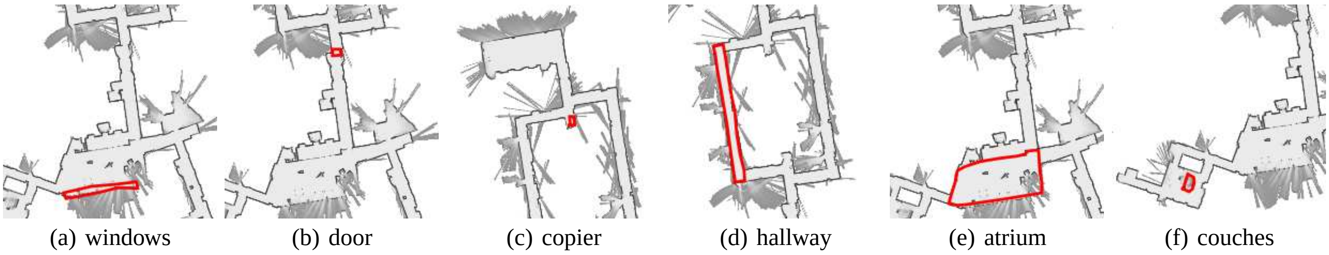

(a) windows (b) door (c) copier (d) hallway (e) atrium (f) couches

Figure 3: Examples of candidate landmarks.

## 4 System

The input to our system is a semantic map of the environment (e.g., a gridmap with the locations of detected objects) along with a natural language command. The output of our system is a cost, which, when optimized, returns the sequence of partial paths and landmarks that correspond to a natural language command. In order to achieve this, we extract a set of candidate destinations from a map of the environment and plan a path from the current location of the robot to each final destination. Extracting destinations from the map is performed by first creating a rapidly exploring randomized graph (RRG) (Karaman and Frazzoli, 2010), and then clustering this graph using a graph-based clustering technique (Ng, Jordan, and Weiss, 2001). Since extracting landmarks is an open research problem, we currently optimize over a set of annotated landmarks, as in Figure 3. Finally, given a natural language command, we have previously shown the ability to automatically extract a sequence of SDCs from a natural language command (Kollar et al., 2010).

Given a sequence of SDCs from a natural language command, a sequence of landmarks from the candidate set, and a path to one of the destinations in the candidate set, we then extract a set of features from each element in the sequence, resulting in a sequence of features. The goal is then to learn when these features correspond to a good sequence of SDCs, partial paths and landmarks, and output the corresponding probability. An overview of the system can be seen in Figure 2. The rest of this section will focus on the conditional random field, how the features are extracted and how we can compute the optimal sequence of partial paths and landmarks.

### Conditional Random Field

We formulate the problem of understanding natural language commands as inference in a linear-chain conditional random field (CRF) (Lafferty, McCallum, and Pereira, 2001). We use CRFs because they do not suffer from the label-bias problem, allow for long-range features, and directly learn the probability of the output class instead of expending modeling effort on the probability of particular features which may (or may not) help with discriminating between examples (Sutton and McCallum, 2007). Combined, this usually leads to better performance at learning the quantity of interest Ng and Jordan (2001).

Thus, using the annotated corpus from the previous section, the goal is to learn when a sequence of SDCs corresponds to a sequence of partial paths and landmarks. Again, $\phi_i$ is a Boolean correspondence variable to determine if the $i$th partial path and landmark correspond to $i$th SDC, $r_i$ is the $i$th partial path, $l_i$ is the $i$th landmark and sdc$_i$ is the $i$th spatial description clause. The corpus from the previous section is used for training, which has examples of paths that correspond to the language ($\phi_i =$ True) and examples of paths that do not correspond to the language ($\phi_i =$ False). Assuming that we have a set of $K$ features $s_k$ which depend only on the $i$th SDC, path, and landmark and $J$ features $t_j$, which depend on pairs of SDCs, paths, and landmarks in the sequence, then we would like to learn the following distribution:

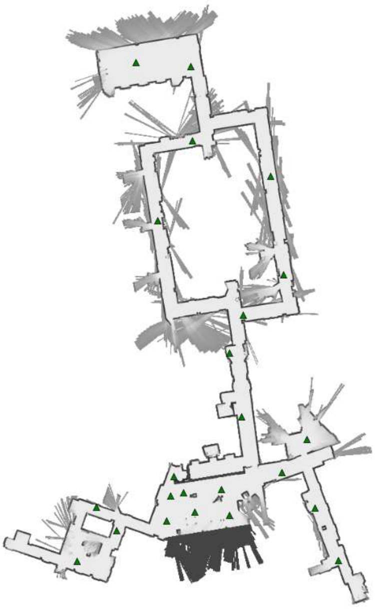


Figure 4: Candidate destinations. Destinations (triangles) are proposed by using spectral clustering on the graph of the RRG.

$$p(\phi_1 \ldots \phi_N | r_1 \ldots r_N, l_1 \ldots l_N, \text{sdc}_1 \ldots \text{sdc}_N) =$$

$$\frac{1}{Z} \exp\left(\sum_{i,j} \lambda_j t_j(\phi_{i-1}, \phi_i, r_{i-1}, r_i, l_i, l_{i-1}, \text{sdc}_{i-1}\text{sdc}_i)\right)$$

$$\times \exp\left(\sum_{i,k} \mu_k s_k(\phi_i, r_i, l_i, \text{sdc}_i, i)\right)$$

The set of features $s_k$ for this work is of three types, and extends the feature set presented in other Kollar et al. (2010). Feature types include those that depend only on the path, those that depend only on the landmark, and those that de-

| Word $p(\text{sdc}_i)$ | Path $q_1(r_i)$ | Landmark $q_2(l_i, \text{sdc}_i)$ | Landmark-Path $q_3(r_i, l_i)$ | Pairs of Paths: $q_4(r_{i-1}, l_{i-1}, r_i, l_i)$ |
|---|---|---|---|---|
| word$_1$<br>word$_2$<br>. . .<br>word$_N$ | path length<br>orient. change | detected object == word$_i$<br>detected object seen with word$_i$<br>maximum Flickr co-occurrence between objects and word$_i$<br>maximum Flickr co-occurrence between object$_i$ and words | distance of $r_i$ to $l_i$<br>$l_i$ in front of $r_i$ start/end<br>$l_i$ in behind of $r_i$ start/end<br>$l_i$ in right of $r_i$ start/end<br>$l_i$ in left of $r_i$ start/end<br>area of landmark<br>perimeter of landmark<br>no landmark | distance between $r_{i-1}$, $r_i$<br>overlap of $r_{i-1}$, $r_i$ |

Table 1: The set of features used when learning the CRF.

pend on the landmark and the path. The partial path features $t_j$ are new in this work and include features that are unable to be captured by an MRF. Thus referring to Table 1, the set of features used in the CRF is:

- $p(\text{sdc}_i) \times q_1(r_i)$ - the Cartesian product of words and path features
- $q_2(l_i, \text{sdc}_i)$ - Context features
- $p(\text{sdc}_i) \times q_3(r_i, l_i)$ - the Cartesian product of words and landmark-path features
- $p(\text{sdc}_i) \times q_4(r_{i-1}, l_{i-1}, r_i, l_i)$ - the Cartesian product of words and “pairs of paths” features.

All of the features which are continuous are discretized into a finite set. In order to learn the parameters $\lambda_j$ and $\mu_k$ that maximize the likelihood of the training dataset, we compute the gradient, as described in Sutton and McCallum (2007), and use L-BFGS to optimize the parameters of the model via gradient descent.

### Path Optimization

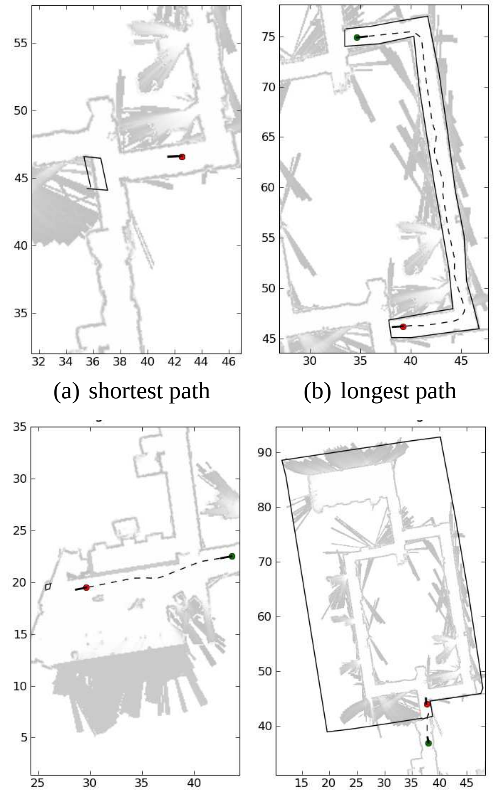
Thus far, we have learned a function that will take as input a sequence of extracted spatial description clauses, partial paths, and landmarks and output the probability that the language corresponds to this path and landmark. Given a map of a novel environment, candidate destinations are extracted. For each candidate destination, a path is planned and for each SDC, a landmark is picked. However, since each SDC corresponds to a partial path, the full path $r$ must be partitioned. For the $j$th SDC, this requires an alignment variable $a_1^j$ and $a_2^j$, which indexes into the path $r$ at starting location $a_1^j$ and ending location $a_2^j$ to divide it into a partial path $r_{a_1^j, a_2^j}$. For example, $r_{0,T} = r$ would correspond to the full path and $r_{0,0} = \emptyset$ would correspond to the empty set. Thus, the goal of the optimization is to perform the following minimization of the cost function:

$$\underset{r, l_1 \ldots, a_1^1 a_2^1 \ldots}{\text{argmin}} \quad C(r_{a_1^1, a_2^1}, \ldots r_{a_1^N, a_2^N}, l_1 \ldots l_N, \text{sdc}_1 \ldots \text{sdc}_N)$$

In order to solve this optimization efficiently, dynamic programming can be used. However, because feature extraction currently dominates the computational expenditure we only consider the uniform alignment (e.g. dividing the path uniformly) and maximum probability landmark.

## 5 Experimental Insights

To evaluate the feasibility of our system and to understand the structure of route directions further, we have performed an analysis of the corpus of partial paths and landmarks. Initially we hypothesized that route directions would occasionally have reordered language relative to the path that was being referred to, that there would be a relatively contiguous sequence of partial paths associated with the language (sequence of SDCs) and that people would refer to a wide range of corresponding paths and landmarks, which would additionally vary greatly in size.



(a) shortest path (b) longest path

(c) smallest landmark (d) biggest landmark

Figure 5: Specific SDCs and their corresponding partial paths for the maximum and minimum landmark sizes and longest and shortest path length. In (a) the corresponding SDC was *turn towards the large cement-colored double doors*, in (b) was *follow the hall*, in (c) was *walk towards the M&M doll* and in (d) was *enter building 36*. The solid line corresponds to the landmark and the dotted line corresponds to the path of the robot.

In Table 2 we can see a summary of the results for these hypotheses. In particular, we can see that in general most

| Partial Path Statistics | |
|---|---|
| Number of route directions | 150 |
| Number of SDCs | 1460 |
| Overlap | 92.3% |
| Strictly overlap | 38.5% |
| Contained | 5% |
| Reordered | 0.4% |

Table 2: Statistics about partial paths from the corpus. Strictly overlap means that two partial paths share more than a common endpoint and reordered means that although two SDCs were sequential in the language, they were not sequential relative to the ground-truth path.

of the partial paths are contained in or overlap with other partial paths. This is indicative of the fact that directions tend to be sequential and redundant: people describe multiple landmarks of interest that overlap with one another in order to not get lost. However, people infrequently reordered SDCs relative to the ground-truth path, indicating that people speak about a path sequentially. By annotating the landmarks and partial paths, we are able to see the wide variety of landmarks and spatial relations that people use in natural language route directions. In this work, we found that the landmarks that people use in route directions ranged from $1964m^2$ to $0m^2$, with an average size of $42.94m^2$, which indicates that the size of landmarks can vary widely. In addition, the length of a partial path varies from $0m$ to $40.83m$, with a mean of $5.53m$, which is also quite large. Some examples of this variation can be seen in Figure 5.

We have evaluated the CRF on a balanced test dataset consisting of 29 positive examples of directions, paths and landmarks and 30 negative examples of directions, paths, and landmarks from a floor of the Stata center at MIT. In Figure 6, we show the test error (the number of examples whose output class do not corresponded to the correct class) for this held-out dataset. For the given set of test examples, we are able to robustly learn when the language corresponds to the partial paths and landmarks, even when we have just a few training examples. For 120 training examples, the CRF has an error of 15% on the held-out dataset, while the MRF has 39% error on the same dataset.

In addition, we are able to introspect the features that are deemed most important to the learning (e.g. the features with highest weight) out of the set of 174,000 features. If we look at the top 30 features when the SDCs correspond to the paths and landmarks, then we would hope to see the CRF learning features that intuitively correspond to good paths. This can be seen in the following:

- **left_left_st** - when the word “left” is used, then the landmark is physically to the left of the robot at the start location.
- **through_ratioFigureToAxes_0.9_1.0** - when “through” is said the robot goes through the landmark polygon.
- **orient_no_landmark** - when “orient” is said, there is typically no landmark.
- **left_orient_st_end_sp_77_84** - when “left” is said the partial path orientation should change between 77 and 84 degrees.
- **to_distFigureEndToGround_0.0_0.1** - when “to” is used the path most likely ends in the landmark.

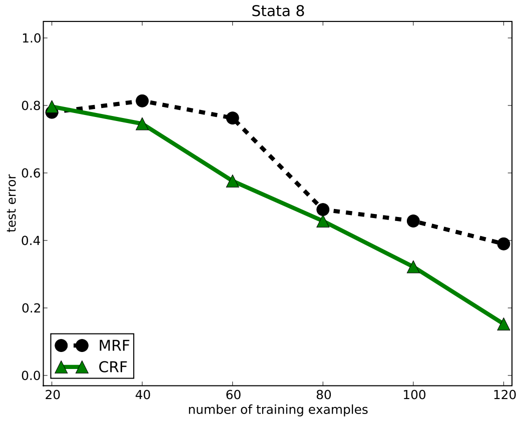


Figure 6: Error on a test dataset consisting of 59 positive and negative examples.

- **room_rel_chair** - when the word “room” is said, a chair is likely visible.

If instead we look at SDCs that do not correspond to the paths and landmarks (e.g. are bad paths), and look at the top 30 features, then we would hope to see features that do not intuitively correspond to good paths. Some examples of these include:

- **left_right_st** - when the word “left” is used, then the landmark is to the right of the partial path.
- **room_lmark_area_0_10** - when the word “room” is used, then the landmark area tends to be small (e.g. less than $2.5m^2$... most rooms were approximately $100m^2$).
- **orient_int_path_sp_14_15** - when the word “orient” is used, then the partial path should be very long (usually the partial path is very small when people say “orient” and only the orientation of the robot changes).
- **left_orient_st_end_sp_-92_-84** - When the word “left” is used, then you should turn -90 (e.g. turn right).

## 6 Related Work

Many authors have proposed formalisms similar to spatial description clauses for enabling systems to reason about the semantics of natural language directions. For example, Bugmann et al. (2004) identified a set of 15 primitive procedures associated with clauses in a corpus of spoken natural language directions. Levit and Roy (2007) designed *navigational informational units* that break down instructions into components. MacMahon, Stankiewicz, and Kuipers (2006) represented a clause in a set of directions as a compound action consisting of a simple action (move, turn, verify, and declare-goal), plus a set of pre- and post-conditions.

Others have created language understanding systems that follow natural language commands, but without using a corpus-based evaluation to enable untrained users to interact with the system (e.g., Dzifcak et al. (2009); Skubic et al. (2004)). Bauer et al. (2009) built a robot that can find its way through an urban environment by interacting with pedestrians using a touch screen and gesture recognition system.

In previous work, we have built direction understanding systems (Wei et al., 2009; Kollar et al., 2010). This work is able to learn the meaning of words from example paths and directly estimates the landmarks, as well as adding a number of features. Matuszek, Fox, and Koscher (2010) and Shimizu and Haas (2009) use statistical methods to understand directions. We use a richer set of features for our conditional model and estimate both the corresponding partial paths and landmarks directly from real-world directions. Also related is Vogel and Jurafsky (2010), who use reinforcement learning to follow directions in pictorial maps (MapTask). Cantrell et al. (2010) attempt to handle the challenges of dialog (such as disfluencies). Finally, there is a large set of psychological studies on understanding the structure of route directions (Fontaine and Denis, 1999; Vanetti and Allen, 1988; Tversky and Lee, 1998; Bugmann et al., 2004). In this work, we believe we have expanded on this body of work by analyzing the geometry of the the referred landmark and partial paths.

## 7 Conclusions and Future Work

Challenges for future work include more general forms of commands, which are more difficult because there are complex verbs and a wide variety of events that must be recognized. Two example commands include:

- "Wait by the staircase next to 391, and bring her to my room when she comes in."
- "Please go to the question mark and wait for her. When she gets here, bring her back here."

There are also challenges that we hope to address in the route direction domain. These include handling the multi-scale nature of some commands, such as *from the entrance of stata closest to the question mark all the way to the other end....*, fine-grained commands that are not amenable to the current set of candidate destinations, and commands that require significant amounts of backtracking.

The contribution of this work is to formulate understanding natural language directions as a CRF that enables the system to infer both the referred landmark and the referred partial path for each SDC in the natural language. We additionally introduce new features into the learning and show an analysis of the route direction corpus that indicates route directions are sequential and referred landmarks and partial paths vary greatly in size and length.

## 8 Acknowledgments

We are grateful for the support of the Office of Naval Research, which supported Thomas Kollar and Stefanie Tellex under MURI N00014-07-1-0749.

## References

Bauer, A.; Klasing, K.; Lidoris, G.; Mhlbauer, Q.; Rohrmller, F.; Sosnowski, S.; Xu, T.; Khnlenz, K.; Wollherr, D.; and Buss, M. 2009. The Autonomous City Explorer: Towards natural human-robot interaction in urban environments. *International Journal of Social Robotics* 1(2):127–140.

Bugmann, G.; Klein, E.; Lauria, S.; and Kyriacou, T. 2004. Corpus-based robotics: A route instruction example. *Proceedings of Intelligent Autonomous Systems* 96—103.

Cantrell, R.; Scheutz, M.; Schermerhorn, P.; and Wu, X. 2010. Robust spoken instruction understanding for hri. In *Proceedings of the 2010 Human-Robot Interaction Conference*.

Dzifcak, J.; Scheutz, M.; Baral, C.; and Schermerhorn, P. 2009. What to do and how to do it: Translating natural language directives into temporal and dynamic logic representation for goal management and action execution. In *IEEE International Conference on Robotics and Automation*, 4163–4168.

Fontaine, S., and Denis, M. 1999. The production of route instructions in underground and urban environments. *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science* 747–747.

Karaman, S., and Frazzoli, E. 2010. Incremental sampling-based algorithms for optimal motion planning. *CoRR* abs/1005.0416.

Kollar, T.; Tellex, S.; Roy, D.; and Roy, N. 2010. Toward understanding natural language directions. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*.

Lafferty, J. D.; McCallum, A.; and Pereira, F. C. N. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning*, 282–289. San Francisco, CA, USA: Morgan Kaufmann Publishers.

Levit, M., and Roy, D. 2007. Interpretation of spatial language in a map navigation task. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on* 37(3):667–679.

MacMahon, M.; Stankiewicz, B.; and Kuipers, B. 2006. Walk the talk: Connecting language, knowledge, and action in route instructions. *Proceedings of the National Conference on Artificial Intelligence* 1475—1482.

Matuszek, C.; Fox, D.; and Koscher, K. 2010. Following directions using statistical machine translation. In *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction*, 251–258. ACM.

Ng, A. Y., and Jordan, M. I. 2001. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes.

Ng, A.; Jordan, M.; and Weiss, Y. 2001. On spectral clustering: Analysis and an algorithm. In *Proceedings of Advances in Neural Information Processing Systems*, 849–856.

Shimizu, N., and Haas, A. 2009. Learning to follow navigational route instructions. In *IJCAI'09: Proceedings of the 21st international jont conference on Artifical intelligence*, 1488–1493. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Skubic, M.; Perzanowski, D.; Blisard, S.; Schultz, A.; Adams, W.; Bugajska, M.; and Brock, D. 2004. Spatial language for human-robot dialogs. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 34(2):154–167.

Sutton, C., and McCallum, A. 2007. An Introduction to Conditional Random Fields for Relational Learning. *Introduction to statistical relational learning* 93.

Tversky, B., and Lee, P. U. 1998. How space structures language. In *Spatial Cognition, An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, 157–176. London, UK: Springer-Verlag.

Vanetti, E., and Allen, G. 1988. Communicating environmental knowledge: The impact of verbal and spatial abilities on the production and comprehension of route directions. *Environment and Behavior* 20(6):667.

Vogel, A., and Jurafsky, D. 2010. Learning to follow navigational directions. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 806–814. Uppsala, Sweden: Association for Computational Linguistics.

Wei, Y.; Brunskill, E.; Kollar, T.; and Roy, N. 2009. Where to go: Interpreting natural directions using global inference. In *IEEE International Conference on Robotics and Automation*.