Talk the Walk: Robotic NLP vs. Human Sublanguage Acquisition

Vladimir Kulyukin
Computer Science Assistive Technology Laboratory (CSATL)
Department of Computer Science
Utah State University
Logan, UT, 84322
vladimir.kulyukin@usu.edu

Abstract

The paper investigates the appropriateness of natural language processing (NLP) in the context of robot-assisted navigation for the visually impaired. Several assumptions of corpus-based robotics are examined. It is argued that, in the short term, NLP may be inadequate and, in the long term, it may not be necessary to enable robotic guides to solicit route directions from bystanders. Human sublanguage acquisition is presented as one feasible alternative.

Introduction

The Americans with Disabilities Act of 1990 provided incentives to remove *structural* barriers to universal access, e.g., retrofitting vehicles for wheelchair access, building ramps and bus lifts, and providing access to various devices through specialized interfaces. For the 11.4 million visually impaired people in the United States, the R&D activities induced by the adoption of the Act have mostly failed to remove the main *functional* barrier: the inability to navigate dynamic and complex environments. This barrier denies the visually impaired equal access to many private and public buildings and makes the visually impaired a group with one of the highest unemployment rates (74%) (LaPlante & Carlson 2000). Thus, there is a significant need for systems that improve the wayfinding abilities of the visually impaired, especially in dynamic and complex environments, where conventional aids, such as white canes and guide dogs, are of limited use.

Can robots assist the visually impaired with wayfinding? Several reasons suggest that this question can be answered affirmatively. First, robot-assisted wayfinding offers feasible solutions to two perennial problems in wearable navigation for the visually impaired: hardware miniaturization and portable power supply. The amount of body gear carried by the visually impaired navigator (the navigator henceforth) is significantly minimized, because most of it is mounted on the robot and powered from on-board batteries. Consequently, the physical load is reduced. Second, insomuch as the key wayfinding capabilities, such as localization and orientation, are delegated to the robot, the navigator enjoys a smaller cognitive load. Third, the robot can interact with other people in the environment, e.g., ask them to yield or receive instructions. Fourth, robots can carry useful payloads, e.g., suitcases and grocery bags.

Interaction with Bystanders

An important question that arises once it is agreed that a robot can act as a guide is whether the robot should interact with bystanders in the environment and, if so, how? Why is this question more important than, say, the question of how the navigator should interact with the robot? As we argued elsewhere (Kulyukin & Gharpure 2006), from the navigator's point of view, an environment can be represented as a directory of destinations browsable with a portable device, e.g., a cell phone or a wearable keypad. Once a destination is selected, the robot knows the navigator's intent.

So, should the robot interact with bystanders? At design time, the robot can be endowed with a topological graph of the environment whose nodes are landmarks that can be recognized by the robot at run time and whose arcs are behaviors that the robot can execute to reach the adjacent landmarks from the current landmark (Kupiers 2000). Alternatively, the robot can be endowed with a global map of the environment built through simultaneous localization and mapping (SLAM) (Fox 1998). The knowledge of landmarks can then be specified by giving sets of robot poses corresponding to landmarks. In either case, the robot's designer can rest assured that, if the target environment has a reasonable degree of dynamism and complexity, there will be a point in time when the robot's map will become out of sync with the environment. In office environments, hallways may be temporarily or permanently blocked due to repairs and in airports passages may be blocked due to passenger traffic flows or construction. The topological structure of the environment may undergo changes due to the addition of new passages and the disappearance of old ones.

The question becomes: when the robot discovers that its map is out of sync with the environment, what should the robot do? The robot's designer could consider three options: contact the designer for additional knowledge engineering, solicit route directions from bystanders, and repair the map autonomously. All three options are valid, but differ in required levels of robot autonomy. In this paper, we will consider the second option as the one most likely to be needed in socially assistive robots.

Corpus-based Robotics

Corpus-based robotics is one approach to robotic route direction following that has recently received some prominence (Bugmann, Lauria, & Kyriacou 2004; MacMahon, Stankiewicz, & Kupiers 2006). The approach consists of three steps: compilation of instructions from potential users, developing a spatial knowledge base with an inference engine, and developing a natural language processor that connects user inputs to robot actions.

This approach is based on two assumptions. First, potential users are naive in robot programming and can only use their own words to explain a task to the robot. Second, the collected corpus is strongly representative of what potential users will say to the robot.

Let us examine these assumptions one by one. It is true that potential users may be naive in robot programming insomuch as they will not be able to re-program the robot whenever necessary. It is also reasonable to assume that natural language is one means that users may want to use to communicate with the robot. But why assume that the robot must be designed to handle arbitrary, unconstrained inputs from the user?

If the robot designer is focused on feasibility, the designer is interested in building a robot in such a way that the user and the robot arrive at the shared vocabulary as fast as possible. This can be achieved by designing a robot that maps arbitrary natural language inputs to its actions or by designing a robot that can quickly reveal its vocabulary to the user and teach the user how to use it effectively. The latter option is often met with two critiques from the symbolic AI community: 1) Where is AI in this?; and 2) This is just a temporary fix; once we have adequate NLP, we will not need it.

To answer the first critique, we point out that the robot designed to address its own limitations can be considered as artificially intelligent as the robot designed to exhibit the human level of language understanding. The robot that explains its limitations to the bystander is designed to utilize the bystander's intelligence to satisfy its goals. The second critique does not take into account the fact that even if the robot has a human-level ability to understand natural language, establishing shared vocabulary will still take time, which the bystander may not have. The bystander may not even want to converse with an artifact, to begin with. Furthermore, there is a body of evidence in cognitive psychology that humans routinely fail to understand route directions they receive from each other (Allen 2000). What then is the basis for expecting robots to do better?

The second assumption of corpus-based robotics is necessary, because if the sampled vocabulary is not strongly representative of what potential users will say to the robot, the robot will either have to infer the meaning of uknown words or engage the user in a dialogue to obtain the meaning. Both options are known to present major conceptual and practical problems to robot designers who use corpus-based robotics.

Sublanguage Acquisition

Is there an alternative to NLP in soliciting route directions from bystanders? Human sublanguage acquistion is one possibility. Sublanguages are artificially constructed subsets of natural languages. They are constructured for specific domains, such as weather reports or stock quotes (Kittredge 1986). Unlike natural languages, subset languages have restricted grammars and lexicons constructed with an explicit objective to remove ambiguity. Subset languages have been experimetally shown to be quickly learnable by people (Sidner & Forlines 2002).

As we argued elsewhere (Kulyukin 2006), speech recognition errors pose a major problem to language-based human-robot interaction. An automatic speech recognition (ASR) system may average 95 to 97 percent accuracy in dictation, where user training is available and the consequences of misrecognized words are easily absorbed. However, an assistive robot that misrecognizes 5 out of 100 commands is a definite risk. Preliminary results reported in the literature, e.g. (Sidner & Forlines 2002), indicate that sublanguages may have adequate speech recognition rates due to small size grammars.

The cognitive requirements on users of subset languages remain relatively unexplored. Two issues are critical: learning rates and sublanguage retention. We decided to check the feasibility of sublanguage acquisition in giving route directions to a robot. A sample of 8 participants has so far been selected. The recruitment of participants is still ongoing. The target sample size is 30. The ages of the selected participants ranged from 22 to 58. Three of the participants had degrees in computer science. The rest were college administrators. The sample included 3 females and 5 males. None of the participants had prior experience with robots. All of them were computer savvy. All participants were native speakers of American English.

We developed a simple user interface where the participant could type English sentences and receive feedback from the robot. We made the decision to exclude speech, because, technically speaking, speech recognition rates do not have much to do with sublanguage acquistion and retention. We selected a route in the office space near our laboratory. The route started at the entrance to our laboratory and ended at the elevators. At the beginning of the experiment, each participant was shown a video of our robotic guide for the visually impaired navigating a route in a different office environment. After the video, the participant was shown the interface on a laptop. It was explained to the participant that this interface is a prototype interface for giving route directions to the robot. The participant was told that his or her task is to give the robot route directions to get to the elevators from our laboratory. All participants were familiar with the environment. The average instruction including the video lasted two minutes.

The participants were instructed to type English sentences as if they were giving directions to a human. The sublanguage processing component was built using the direct memory access parsing algorithm (Riesbeck & Schank 1989). Our DMAP parser had 38 memory organization packages (MOPs) referenced by 140 phrases. Each MOP corresponded to a location in the environment. The sublanguage processor was built in such a way that the robot could process instructions of the type *Go to location X* or *Move to point X*, where X is a location number. For example, the processor would be able to process the following route description: *Go to location 1. Go to location 2. Go to location 4. Go to location 5.*

Our objective was to determine how quickly the participants could learn this language and how well they retained it. When the sublanguage processor was unable to parse the input, the processor would display a bitmap with the map of the environment where each location known to the subprocessor was marked by a circle with a number. The processor would then give a user an example of a successful route description: *If you want me to go from location 1 to location 4, you can say: Go to location 2. Go to location 3. Go to location 4. You can also say: Go to point 2. Go to point 3. Go to point 4.* The description given in the example was for a route different from the one the participants were asked to describe. The subprocessor would give this feedback both with a textual message and through synthetic speech. The participant was then asked by the processor to retype his or her input.

All participants first typed route descriptions in unconstrained English. On seeing the map and the example, all of them immediately switched to the sublanguage to accurately direct the robot to the destination. For each participant, we timed how long it took to give the robot correct route directions. The timer was started when the participant started typing and was stopped when the processor would say that the route directions were successfully processed. The average completion time was 3 minutes and 30 seconds.

Each participant was asked to come back to repeat the experiment in two days (48 hours). When the participant came back, the participant was asked to instruct the robot on how to navigate a different route. The interface now contained a help button, which the participant was instructed to press if he or she needed help from the system. All partcipants successfully completed the task without pressing the help button. The average completion time was 55 seconds. The t-test suggests that the difference in the task completion times is significant at $\alpha = 0.05$.

Conclusions

Since our experiment is still ongoing, the conclusions are preliminary. Our findings suggest that the participants in the sample quickly acquire the sublanguage and, more importantly, retain it over a period of two days. The current design of the experiment has an important limitation: the input is confined to typing.

We agrued that, in the short term, NLP may be inadequate due to problems with unconstrained vocabulary and, in the long term, it may not be necessary to enable robotic guides to solicit route directions from bystanders. Human sublanguage acquisition is one possible alternative that has the potential to succeed. The robot should not be designed to walk the bystander's talk. Instead, it should be designed to have the bystander talk the walk in the language that the robot can process.

Our conclusion should not be construed as suggesting that human sublanguage acquistion is the only alternative to NLP with respect to soliciting route directions from bystanders. Touch-screen graphics is another possibility that, in our opinion, should be investigated in greater depth.

The paper can be extended to full length. Twenty two more participants are expected to complete the experiment within the next month. After 30 participants complete the experiments, they will be asked to repeat the experiment with a commercial speech recognition engine. The paper will be extended with a detailed analysis of the results. It will also be extended with screen shots of our interface.

References

Allen, G. 2000. Principles and practices for communicating route knowledge. *Applied Cognitive Psychology* 14:333–359.

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

Fox, D. 1998. *Markov Localization: A Probabilistic Framework for Mobile Robot Localization and Navigation*. Ph.D. Dissertation, University of Bonn, Germany.

Kittredge, R. 1986. *Variation and Homogeneity of Sublanguages: Studies of Language in Restricted Domains, R. Kittredge and J. Lehrberger (Eds.)*. New York: Walter de Gruyter.

Kulyukin, V., and Gharpure, C. 2006. Robot-assisted wayfinding for the visually impaired in structured indoor environments. *Autonomous Robots* 21:29–41.

Kulyukin, V. 2006. On natural language dialogue with assistive robots. In *Proceedings of the 2006 ACM Conference on Human-Robot Interaction*. Salt Lake City, Utah: ACM.

Kupiers, B. 2000. The spatial semantic hierarchy. *Artificial Intelligence* 119:191–233.

LaPlante, M. P., and Carlson, D. 2000. Disability in the united states: Prevalence and causes. In *U.S. Department of Education, National Institute of Disability and Rehabilitation Research*.

MacMahon, M.; Stankiewicz, B.; and Kupiers, B. 2006. Walk the talk: Connecting language, knowledge, and action in route instructions. In *Proceedings of AAAI*.

Riesbeck, C., and Schank, R. 1989. *Inside Case-based Reasoning*. New Jersey: Lawrence Erlbaum.

Sidner, C., and Forlines, C. 2002. Subset languages for conversing with collaborative interface agents. In *Proceedings of International Conference on Spoken Language Processing*.