2019-03-22

Towards Generating Spatial Referring Expressions in a Social Robot: Dynamic vs Non-Ambiguous

Wallbridge, CD

10.1109/HRI.2019.8673285

ACM/IEEE International Conference on Human-Robot Interaction

Towards Generating Spatial Referring Expressions in a Social Robot: Dynamic vs Non-Ambiguous

Christopher D. Wallbridge
University of Plymouth
Plymouth, UK
christopher.wallbridge@plymouth.ac.uk

Séverin Lemaignan
Bristol Robotics Laboratory
University of West England
Bristol, UK
severin.lemaignan@brl.ac.uk

Emmanuel Senft
University of Plymouth
Plymouth, UK
emmanuel.senft@plymouth.ac.uk

Tony Belpaeme
IDlab - imec
Ghent University
Ghent, Belgium
tony.belpaeme@ugent.be

***Abstract*—We present in this paper our work towards a new dynamic method of generating spatial referring expressions. While people are generally ambiguous in their description of locations, previous methods of artificial generation mostly considered non-ambiguous descriptions. However, to increase the naturalness of interaction and share workload in the communication, robots should be able to generate language in a more dynamic way. Our method initially produces ambiguous spatial referring expressions followed by dynamically generating repair statements. We built a classifier using data from 18 participants as they described locations to each other. We perform a preliminary analysis on this method using two further pilot studies.**

***Index Terms*—Spatial Referring Expressions; Ambiguity; Dynamic; Classifier; Social Robotics; HRI**

I. Introduction

The ability to generate Spatial Referring Expressions (SRE) is a key area to allow robots to communicate naturally with people within an environment. A typical assumption in robot development is that the best description is one that allows an object or location to be uniquely described [1]. This approach often has an issue of combinatorial explosion, which more recent algorithms attempt to resolve as efficiently [2].

Work on understanding a SRE realises that the description provided by a person is often ambiguous and steps need to be taken to disambiguate a description [3]. This process can be cumbersome for a robot, with a lot of dialogue required to narrow down a description. However much of the information can be disambiguated by the situational context [4].

Evaluation frameworks for generation algorithms are often based upon a single direction of communication [5]. However communicating the location of an object to someone else is often a two way communication [6]. Descriptions are often underspecified and a dynamic strategy of repair is used to correct these mistakes. This strategy allows for the sharing of cognitive load between a describer and a listener. In children this process can be highly dynamic with the child receiving the description making actions to prompt the child in the role of describer, or allow for a simpler description [7].

The use of a dynamic description given by a robot should be investigated for potential benefits. However, this area of research (ambiguous spatial referring) is not explored by the community. While the interactions between two people are often dynamic in a normal discussion, we want to explore if that would be beneficial in the case of interactions between a robot and a human.

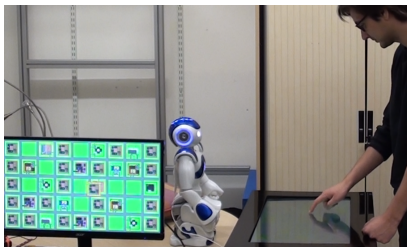

Fig. 1. A participant interacting in our study in one of the pilots. The robot describes the locations of buildings on a game board which the user then places using the touch screen. On the left a monitor is mirroring the touch screen.

II. Methodology

This paper presents three pilot studies taking place on an interactive tabletop (Fig 1) where an agent, the *describer* has to guide a second agent, the *manipulator* to move an object to a required location. The basis of the interaction was a 'city planning' game in which the manipulator has to move the picture of a building on a map to an empty location which can only be described with confusing, ambiguous or complex utterances (requiring between 1 and 4 descriptors to disambiguate their location e.g. "A residence is above a commercial district and to the right of a fire department"). Before starting the game, a tutorial screen was presented that showed each of the building types with names that they could be identified with, and information on the game was given to the participants. In each session two maps were displayed, each with 7 objects to be placed that were presented one at a time. Data (position of the target and of the moving object, completion time and video recording) was recorded on all interactions.

The first pilot study involved 18 human participants who interacted in groups of two and were asked to take the roles of manipulator and describer. The roles were swapped upon completion of the first map. The objective of this pilot was to gather data on the way the objects were moved and the strategies applied by the describer. To that extent, we used Underworlds [8] to represent the state of game and we sample

978-1-5386-8555-6/19/$31.00 ©2019 IEEE

the coordinates of the target object and the moving object at 10 Hz.

We manually coded the video to annotate three categories of repair statements used by participants to clarify ambiguous descriptions:

- *Negate* - A negative response indicating that the manipulator is heading in the wrong direction (e.g. "Not this one.").
- *Elaborate* - A response to give more information, when the manipulator appears to be hesitating (e.g. "... and also to the left of the hospital.").
- *Positive* - A positive response given to the manipulator to indicate they are heading in the right direction (e.g. "Yes.").

We obtained from these interactions 4701 datapoints assigning one of these three types of utterance to a game situation (distance to target, change in distance to target, magnitude of motion and change in angle from previous sample of motion). These points were divided in a training set (80%) and a testing set to train a classifier (a SVC with an RBF kernel) assigning a type of statement to a game situation. This classifier was then used to create a robot using ambiguous statements and repairs which was evaluated in the third pilot.

The second and third pilots were interactions between a human and an autonomous robot, where the robot took the role of the describer and had to guide the participants in the manipulation task. The second pilot (n=8) evaluated the control condition, the robot used a non-ambiguous strategy for the two maps, and in the last pilot (n=9), the conditions were alternated: the robot used our new classifier with the dynamic description on the first game board, while reverting to non-ambiguous description for the second.

III. Preliminary Results

The classifier managed to achieve an 89% success rate with our testing data. While this success rate seems encouraging the confusion matrix (Table I) reveals a 33% success rate classification of negation. This issue with negation is likely due to a lack of data compared to the other two classifications, only making up 7% of the overall data.

We found that it was not possible to get consistent timing on when and how often to provide feedback from the data we gathered on two people describing. We believe that this is due to the person in the role of describer trying to process an with which environment they were unfamiliar, and making their own mistakes without realising. For the subsequent pilot we decided to use a manually coded timing mechanism. Feedback was based on an average result from the classifier over a period of time. A higher weighting towards negation was added in an attempt to offset the current deficiencies of the classifier while more data was gathered. Future work may emphasise establishing natural timing and amount of feedback.

The first follow-up study had 8 participants. We saw on especially complex descriptions that it was necessary for people to hear at least one repetition of the description before being able to disambiguate the location.

		Prediction		
		Negate	Elaborate	Positive
Actual	Negate	20	41	0
	Elaborate	3	390	18
	Positive	0	39	436

TABLE I
CONFUSION MATRIX OF THE CLASSIFIER MADE WITH THE RESULTS OF THE FIRST STUDY.

In the final pilot the dynamic condition had mixed results over 9 participants. For some users, who moved objects confidently whether correct or not, the dynamic condition worked well, and they said the dynamic condition felt more natural. For those who were more hesitant, in both when they moved and how they moved, they often found the robot elaborating too much, and described it as overwhelming.

IV. Discussion and Future Work

The issues we found in the dynamic description appeared to be caused by two problems in our current system. Firstly, that the robot would often elaborate when they were moving in the wrong direction rather than negating. This is caused by the issue in the classifier that we identified in the confusion matrix. We intend to take the data from the pilots to improve the classifier further. Secondly the robot's elaboration was sometimes not enough to indicate that a selected target was incorrect. We need to focus on making sure that whatever feedback is given helps to disambiguate the current location from the correct location.

Upon correction of these issues we intend to run a full study. This study will compare a robot providing a dynamic description, to a robot that provides a non-ambiguous description.

V. Acknowledgements

This work was supported by the EU H2020 L2TOR project (grant 688014).

References

[1] R. Dale and E. Reiter, "Computational interpretations of the Gricean maxims in the generation of referring expressions," *Cognitive science*, vol. 19, no. 2, pp. 233–263, 1995.

[2] J. D. Kelleher and G.-J. M. Kruijff, "Incremental generation of spatial referring expressions in situated dialog," in *Proceedings of the 21st COLING and the 44th annual meeting of the ACL*. Association for Computational Linguistics, 2006, pp. 1041–1048.

[3] M. Shridhar and D. Hsu, "Interactive visual grounding of referring expressions for human-robot interaction," *arXiv preprint arXiv:1806.03831*, 2018.

[4] A. Magassouba, K. Sugiura, and H. Kawai, "A multimodal classifier generative adversarial network for carry and place tasks from ambiguous language instructions," *arXiv preprint arXiv:1806.03847*, 2018.

[5] T. Williams and M. Scheutz, "Referring expression generation under uncertainty: Algorithm and evaluation framework," in *Proceedings of the 10th International Conference on Natural Language Generation*, 2017, pp. 75–84.

[6] H. H. Clark and D. Wilkes-Gibbs, "Referring as a collaborative process," *Cognition*, vol. 22, no. 1, pp. 1–39, 1986.

[7] C. D. Wallbridge, S. Lemaignan, E. Senft, C. Edmunds, and T. Belpaeme, "Spatial referring expressions in child-robot interaction: Let's be ambiguous!" in *4th Workshop on Robots for Learning (R4L) - Inclusive Learning @HRI 2018*, 2018.

[8] S. Lemaignan, Y. Sallami, C. Wallbridge, A. Clodic, T. Belpaeme, and R. Alami, "Underworlds: Cascading situation assessment for robots," 2018.