2014

Clarification Dialogues for Perception-based Errors in Situated Human-Computer Dialogues

Niels Schütte
Technological University Dublin, niels.schutte@gmail.com

John D. Kelleher
Technological University Dublin, john.d.kelleher@tudublin.ie

Brian Mac Namee
University College Dublin, Ireland, brian.macnamee@ucd.ie

Recommended Citation

Schutte, N., Kelleher, J. & Mac Namee, B. (2014). Clarification Dialogues for Perception-based Errors in Situated Human-Computer Dialogues. *Workshop on Multimodal, Multi-Party, Real-World Human-Robot Interaction*, Turkey 2014. doi:10.1145/2666499.2666506

Clarification Dialogues for Perception-based Errors in Situated Human-Computer Dialogues

[Extended Abstract]

Niels Schütte
Dublin Institute of Technology
niels.schutte@student.dit.ie

John Kelleher
Dublin Institute of Technology
john.d.kelleher@dit.ie

Brian Mac Namee
Dublin Institute of Technology
brian.macnamee@dit.ie

ABSTRACT

We present an experiment about situated human-computer interaction. Participants interacted with a simulated robot system to complete a series of tasks in a situated environment. Errors were introduced into the robot's perception to produce misunderstandings. We recorded the interactions and attempt to identify strategies the participants used to solve the arising problems.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Natural Language; I.2.1 [**Applications and Expert Systems**]: Natural language interfaces

General Terms

Experimentation

Keywords

Spoken Language Dialogue Systems, Sensor Error, Clarification

1. INTRODUCTION

We describe an initial analysis of clarification dialogues from an experiment about situated human-computer interactions we are currently performing. In the experiment, human participants interacted through a text based dialogue interface with a simulated robot system. Participants were asked to complete a series of small tasks in which they had to instruct a simulated robot arm to re-arrange simple geometric objects in a simulated environment into a given target configuration. To do so, they issued instructions such as "Pick up the green ball" and "Put it in front of the red box". If the system was unable to perform a request, it would attempt to explain its problem (e.g. "Sorry, I don't see any red boxes"). Figure 1 shows the simulated world as it was presented to the participants. In some parts of the experiment, errors were introduced into the robot's perception. There were two types of errors. In the first error condition, the robot completely failed to detect the object affected (the **missing object** condition). This meant that the robot would not be able to resolve references to it, and it also would not be able to interpret it as a landmark in a referring expression. In the second error condition, the robot did perceive the affected object, but misclassified one of its properties (e.g. it mistook a green box for a red box or a green ball (the **colour misclassification** and **type misclassification** condition). As in the first condition, the robot was also unable to resolve references to the object correctly, but participants could recover from this problem by coming up with alternative descriptions.

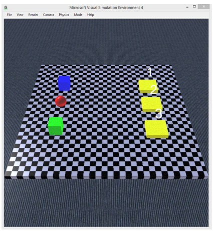

Figure 1: The simulated world.

The tasks were designed so that participants were lead to use objects that were affected by errors in their attempt to solve the task. The experiment is described in more detail in [3]. We see this work in the context of research into misunderstandings in human-computer interaction [4] and communication under conditions of uncertain shared context [1]. We believe that the results from this work may eventually inform the design of situated dialogue systems that are capable of recognizing errors in their own perception based on the behaviour of the human dialogue partners and take steps towards improving their perception e.g. by retraining classifiers.

2. EXPERIMENT

We have completed two runs of the experiment so far. In the first run, participants solved a series of tasks without errors introduced into the robot's perception. This run served to establish a baseline difficulty for the task. In the second run, errors were introduced. In each run, contributions by the participants were stored and annotated with their interpretation by the system. A number of performance related measurements such as task completion rate or the number of invalid references were also recorded.

MMRWHRI'14, November 16, 2014, Istanbul, Turkey.
ACM 978-1-4503-0551-8/14/11.
http://dx.doi.org/10.1145/2666499.2666506.

3. DATA PREPARATION

We took the data from the second run and extracted all instances where a participant made a reference that involved an object that was affected by a perception error. We collected all actions that followed after the initial problematic reference until the participant issued an instruction that resulted in the robot successfully picking up the object originally targeted. This provided us with excerpts from the dialogues in which the participants successfully managed to resolve misunderstandings. At a later stage we will contrast them with unsuccessful resolution attempts and attempt to identify distinguishing factors.

4. ANNOTATION

Our goal was to observe what strategies the speaker used to resolve the misunderstandings resulting from the resolution problems. Our initial hypothesis was that participants would incrementally modify their initial referring expression until they reached a successful referring expression. We therefore extracted from our original annotation a higher level annotation that summarized the expression used by the participant in terms of what information was contained in it. Based on this information we created a second level of annotation which summarized the changes in referring expressions between actions. For example, if a participant issued the instruction "Pick up the ball" and in the subsequent instruction said " Pick up the red ball" this would result in an event that represented the fact that the participant added the attribute *colour* to their description. We were specifically interested in the following types of events that described when participants

- added or removed attributes
- added or removed of spatial descriptions
- changed the value of an attribute
- changed a landmark used in a spatial description

Spatial descriptions understood by the system either took the form of a landmark reference (e.g. "the ball near the red box") or a direction based description (e.g. "the green box on the left.")

5. OBSERVATIONS

We extracted a set of 74 sequences that fit our target criteria. They have an average length of 4.6 actions (where each action consists of an utterance by the participant and a response by the system). 32.4% of the sequences were of length 2 and 45.9% were of length 3 or shorter.

Sequences of length 2 tended to occur in situations where the participant used a description containing a landmark that was affected by an error. If an **alternative landmark** was available, the participant simply used it instead and solved the problem easily. Other short sequences occurred when objects were affected by attribute errors. Some participants tried out **alternative attribute values**, e.g. if the expression "the blue ball" failed, participants tried the expression "the blue box". If they guessed correctly how the object was perceived by the system, this strategy lead to a quick resolution.

Another successful strategy was to use **more general expressions**. Since the participants knew that the system was likely to mistake colours, they sometimes used terms that did not contain a colour attribute ("the box") or a neutral term for the type of the objects ("the blue object"). This was often combined with the use of **directional expressions** such as "the ball on the left". Compared to landmark-based spatial expressions, these expressions were more robust because they did not require the participant to describe a landmark which could be affected by perception errors.

An inspection of the longer the sequences indicates that in these cases the participants, after an initial unsuccessful attempt to pick up the object in question, turned their attention to other objects and eventually returned to the original target object. In fact, some participants described in exit interviews that they often, when faced with an object they could not figure out how to describe, would **"construct a scene"** that made it easier to identify the object, e.g. by isolating it on one side of the scene.

Another interesting observation is related to the use of landmarks. Some participants, when they noticed that an expression using a landmark based spatial expression could not be resolved, instructed the robot to pick up the object they had used as the landmark event though this was not necessary to solve the task. A possible explanation for this would be that they attempted to determine whether the system perceived the landmark object in the same way as they did. This can be understood as an attempt to **"query the robot's model"** of the world.

6. DISCUSSION AND FUTURE WORK

The current work represents a first analysis step. The part of data that was analysed forms only a small part of the total available data. In the future we aim to broaden our analysis and provide a more thorough and quantitative evaluation. In particular we are also interested in relating this work to earlier work in which we analysed human-human dialogues in an instruction giving corpus [2]. Some parallels, such as the use of alternative landmarks and direction based descriptions as fall-back strategies do appear to exist after a first analysis.

We are currently performing a third iteration of the experiment in which the participants are able to access the robot's understanding either through a visual interface or language based descriptions and possibly direct querying.

7. REFERENCES

[1] C. Liu, R. Fang, and J. Y. Chai. Towards mediating shared perceptual basis in situated dialogue. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 140–149. ACL, 2012.

[2] N. Schütte, J. Kelleher, and B. Mac Namee. A corpus based dialogue model for grounding in situated dialogue. In *Proceedings of the 1st Workshop on Machine Learning for Interactive Systems.(MLIS-2012).*, Montpellier, France, Aug. 2012.

[3] N. Schütte, J. Kelleher, and B. Mac Namee. The effect of sensor errors in situated human-computer dialogue. In *Proceedings of the The 3rd Workshop on Vision and Language (VL'14).*, page to appear, Dublin, Ireland, Aug. 2014.

[4] J. Shin, S. S. Narayanan, L. Gerber, A. Kazemzadeh, D. Byrd, and others. Analysis of user behavior under error conditions in spoken dialogs. In *INTERSPEECH*, 2002.