A Unified Approach to Dialogue Model for Situated Referential Grounding

Mohammad Fazleh Elahi, Dimitra Anastasiou, Hui Shi
Ludwig Maximilian University of Munich, Luxemburg Institute of Science and Technology, Universität Bremen
Fazleh.Elahi@anglistik.uni-muenchen.de, dimitra@d-anastasiou.com,
shi@informatik.uni-bremen.de

Abstract

This paper presents a dialogue system, which combines the Agent-Oriented Dialogue Management (AODM) model (Ross and Bateman, 2009b) and the generalized dialogue modeling approach proposed in Shi et al. (2011) for referential grounding. Benefiting from theories, the approach deals with user description and the mental state (of the participants) to enable effective dialogues to make the referential task successful. The aim of the dialogue system is to identify objects that user refers to in the environment. To demonstrate the application of this approach to human and robot interaction, the model is implemented on the backbone of DAISIE (DiaSpace's Adaptive Information State Interaction Executive) (Ross and Bateman, 2009a), an information state based dialogue modeling for situated dialogues.

1. Introduction

In situated dialogue, humans and artificial agents (e.g., robots) are co-present in an environment to achieve joint tasks. It is a form of dialogue in which the expressed meaning may refer to a physical environment. When talking with a human, the system should understand how the dialogue relates and refers to that environment. The task of the dialogue manager of a situated dialogue system is to determine which communicative actions to take (i.e. what to say), given a goal and observations on the interaction history and the physical environment. Therefore, the dialogue manager is the central component of a dialogue system. It accepts spoken input from the user, produces messages to be communicated to the user, interacts with external knowledge sources, and generally controls the dialogue flow. Referential grounding dialogues present notable challenges in dialogue modeling communities. First, human and robot have mismatched capabilities for understanding the shared physical world. The system needs to model how conversation partners mediate a shared basis when they have mismatched capabilities of understanding the shared world. Second, an adequate analysis of text is mandatory for a robot to understand utterances, such as "Go to the box" or "The box is in front of the small ball". It is necessary for robots to have an adequate dialogue model with situated language understanding. Third, human's references are often ambiguous. Therefore, how humans establish reference in dialogue (i.e. reference and grounding in situated human-robot interaction) needs to be analyzed and this adds complexity to the dialogue manager.

This paper presents a unified approach to dialogue, which combines the Agent-Oriented Dialogue Management (AODM) model (Ross and Bateman, 2009b) and the generalized dialogue modelling approach proposed in Shi et al. (2011) for referential grounding. Benefiting from theories, the approach deals with user description and the mental state (of the participants) to enable effective dialogues to make the referential task successful. The aim of the dialogue system is to identify objects that user refers to in the environment. To demonstrate the application of this approach to human and robot interaction, the model is implemented on the backbone of DAISIE (Ross and Bateman, 2009a), an information state based dialogue modelling for situated dialogues.

2. Related Work

Clark and Wilkes-Gibbs (1986) investigated how participants in a conversation collaborate in making a referring action successful. We adopt the underlying idea of this referential grounding model in our dialogue system. Liu et al. (2012) conducted experiments on human-human interaction to investigate collaborative referring with mismatched perceptual capabilities. In these experiments, two human subjects play a set of naming games. In this approach, the dialogue model (Liu et al., 2013) is designed using Clark's Contribution Model (Clark & Brennan, 1991). In their model, the participant describes the physical object of the scene in the installment. An installment is a description that is divided into several parts/installments, each of which needs to be confirmed before continuing to the rest.

3. System Architecture

The aim of the dialogue system is to identify objects that human refers to in the environment. The human user selects an object from the environment and gives a motion command (i.e. go to the box) to the system. To make the referring action easy to the system to understand, the human user can describe various features or spatial settings of the object. The system needs to identify what the human refers to in the environment. To achieve the

goal, the system asks questions or shares information with the human.

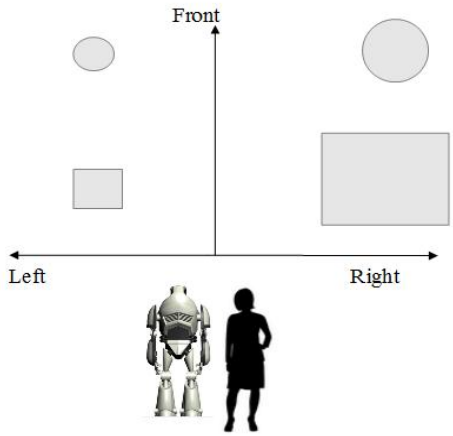

Figure 1. The environment

Figure 1 shows an example of the environment. A successful interaction between the human and the robot through our dialogue model is shown. Here, the environment consists of two boxes and balls spread over a room (the system is capable for solving for arbitrary number of physical objects). The perspective of the human and the robot are shown by an arrow. Some of examples of the dialogue system are given below:

Example 1:
System: Hello! Where do you want me to go?
User: go to the box.
System: Do you mean the large box or the box in front of the small ball?
User: i mean the large one.
System: Ok, I found the large box and I'm going there.

Example 2:
System: Hello! Where do you want me to go?
User: go to the large box in front of the large ball.
System: Ok, I found the large box in front of the large ball and I'm going there.

Example 3:
System: Hello! Where do you want me to go?
User: go to the small box.
System: you mean the box in front of the small ball, right?
User: yes
System: Ok, I found the small box in front of the small ball and I'm going there.

Example 4:
System: Hello! Where do you want me to go?
User: go to the long box.
System: I cannot see the long box.

The system works as follows: first, the user utterance is converted to text and then sent to the language analyzer module applying CCG parsing (Clark et al., 2002). The parse is then analyzed by "linguistically motivated ontology" that provides linguistic semantics for spatial expressions (Bateman et al., 2010). The system then communicates with the probabilistic model for Referring Expression Generation (REG) (Mast & Wolter, 2013) to relate the language with the physical world (Vivien et al., 2014). The dialogue management module (Ross and Bateman, 2009b; Shi et al., 2011) then plans dialogues that are then given to the KPML (The KOMET-Penman Multilingual Linguistic Resource Development Environment) language generator (Bateman, 1997) for generating system utterance.

The dialogue system is implemented on the backbone of DAISIE (Ross and Bateman, 2009a), a java implementation of information state based dialogue modeling for situated dialogues. The system uses VoCon[1] as speech recognizer, OpenCCG[2] parser as language analyzer and GUM-Space ontology, KPML[3] as language generator, and MaryTTS[4] as text to speech generator.

4. Dialogue Management

In this section a unified approach of the Agent-Oriented Dialogue Management (AODM) model and the generalized dialogue modeling approach proposed in Shi et al. (2011) is discussed with respect to the referential grounding domain.

Our dialogue structure is similar to Clark and Wilkes-Gibbs (1986) model of referential grounding. In our approach, the user (i.e. human) would describe a physical object in the scene and present an initial referring expression. The system would then judge the referring expression by one of the following options:

- Accepting it, if the object is found in the scene;
- Rejecting it, if it is not possible to find an object with the description;
- Ask clarification questions (i.e. Do you mean the large box or the box in front of the small ball?), if the previously given description (of the user) is not acceptable enough by the probabilistic reference handler (Mast & Wolter, 2013) to disambiguate an object in the scene.

The human then either refashions the referring expression, expanding it by adding further information, or replacing the original expression with a new expression. The referring expression that results from this is then judged again, and the process continues until the referring expression is acceptable enough to the participants for current purposes.

[1] http://www.nuance.de/for-business/speech-recognition-solutions/vocon-hybrid/index.htm

[2] http://openccg.sourceforge.net/

[3] http://www.fb10.uni-bremen.de/anglistik/langpro/kpml/readme.html

[4] http://mary.dfki.de/download/

The dialogue structure is implemented at the illocutionary level using the generalized dialogue modeling (Shi et al., 2011) by Recursive Transition Networks (RTNs) (Sitter S. & Stein A, 1992). Figure 2 shows the transition diagram of Ground (user, system) initiated by the user and responded to by the system. As it can be seen from the diagram, the dialogue model is represented as the traversal of a state transition network with arcs denoting transitions and nodes denoting states (i.e. state a, state b, state c and state d). The black circle (state d) denotes final state. In generalized dialogue model, such as the one depicted in Figure 2, there are states from which more than one transition are possible. To this end, we use *conditional transitions* proposed in Shi et al. (2011). A *conditional transition* is activated only if its conditions are satisfied. Therefore, each transition of Ground (user, system) is associated with a set of conditions under which the dialogue action can be taken.

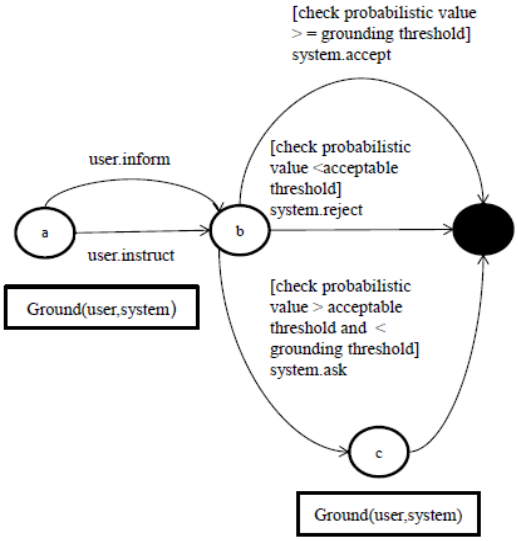

Figure 2. The generalized dialogue modelling of Ground (user, system) using RTNs.

The generalized transition diagram Ground (user, system) shown in figure 2, is initiated (i.e. state a) by one of the user's dialogue acts (*instruct* or *inform*). If the user presents an initial referring expression by the dialogue act of type *instruct* (i.e. go to the box), then the transition will be user.instruct. If the user expands it by adding further information (i.e. I mean the small one) or presents a new referring expression (i.e. It is in front of the small ball), then the transition will be user.inform. For the purpose of the *conditional transition* at state b, we assign two thresholds to the probabilistic values of referring expressions proposed in Mast and Wolter (2013). These are

- if the probabilistic value of the referring expression is greater than or equal to the *grounding threshold*, then the system accepts it and the transition will be system.accept (i.e. ok, I found it);
- if it is less than *acceptable threshold*, then the system rejects it and the transition will be system.reject (i.e. I am sorry, I can't see it);
- if it is greater *acceptable threshold* but less than *grounding threshold* then the transition will be system.ask (i.e. Do you mean the large box or the box in front of the small ball?).

Conditional transition model neither deals with context (and history) nor the mental state of the dialogue participants. To address these issues, we combine the generalized dialogue modeling with the Agent-Oriented Dialogue Management (AODM) model (Ross and Bateman, 2009b) that uses information-state update theory (Ginzburg, 1996) with a light-weight rational agency model. The information state structure and the non-dialogic mental state of our model are similar to those of the AODM model. We use the belief models (Lison et al., 2010) to form and maintain common ground, which is not presented here in detail. The belief model contains all referring expressions and their probabilistic values generated from the probabilistic Referring Expression Generation (REG) (Mast and Wolter, 2013). In our model, the conditional transitions of each state of the transition diagram are generated considering the dialogue history and mental state of the participants. For example, in the case of state b (figure 2), the intention of the system is to find a physical object from the user's referring expressions. The intention then triggers a plan that checks the probabilistic values of the user's referring expression from the belief model and generates conditions for conditional transition of the Recursive Transition Networks (RTNs).

5. Conclusion

We have discussed a dialogue model for referential grounding that considers probabilistic value of the user expression (i.e. how good a referring expression matches a physical object in the scene) and the mental state of the participants.

The unified approach has several advantages over other state-of-art approach (Liu et al., 2013) of referential grounding. First, our approach is capable of enabling clarification dialogues based on object description and spatial relations. Second, it takes into consideration probabilistic value of the user expression (i.e. how good a referring expression matches a physical object in the scene). Third, our dialogue management model that fuses information-state update theory with a light-weight rational agency model (AODM) provides mechanisms to model dialogue context and history the mental state of the participants. Finally, it is less rigid and repetitive than the graph-matching based approach proposed in Liu et al. (2013). As for future prospects, we are currently extending the mental model of the AODM model, so that we can incorporate more dialogues considering the internal state of the dialogue participant.

Acknowledgements

This research was supported by the SFB/TR 8 Spatial Cognition. We would also like to thank Daniel Couto Vale for language analyzer and Vivien Mast for the probabilistic reference handler (REG).

Reference

C. Liu, R. Fang, and J. Y Chai. 2012. Towards mediating shared perceptual basis in situated dialogue. *In*

Proceedings of the 13th Annual SIGdial Meeting on Discourse and Dialogue.

C. Liu, R. Fang, L. She, and J. Y. Chai. 2013. Modeling Collaborative Referring for Situated Referential Grounding. *In Proceedings of 14th annual SIGdial Meeting on Discourse and Dialogue*.

H. H. Clark and S. E. Brennan. 1991. Grounding in communication referring expressions. *Perspectives on socially shared cognition*, 13(1991):127–149.

H. Shi, C. Jian, and C. Rachuy. 2011. Evaluation of a unified dialogue model for human-computer interaction. *In International Journal of Computational Linguistics and Applications (IJCLA)*, Vol 2, S. 155-173. Bahri.

H. H Clark and D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

J. A. Bateman, J. Hois, R. Ross, and T. Tenbrink. 2010. A linguistic ontology of space for natural language processing. *Articial Intelligence*, 174(14):1027{1071, September 2010.

J. A. Bateman. 1997. Enabling technology for multilingual natural language generation: the KPML development environment. *Journal of Natural Language Engineering*, 3(1)(1997), 15—55

J. Ginzburg. 1996. Interrogatives: Questions, Facts and Dialogue. *In The Handbook of Contemporary Semantic Theory*.

P. Lison, C. Ehrler and G. M. Kruijff. 2010. Belief Modelling for Situation Awareness in Human-Robot Interaction. *The 19th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2010)*.

R. Ross and J. A. Bateman. 2009a. DAISIE: Information state dialogues for situated systems.*12th International Conference, TSD 2009*, September 13-17, 2009.

R. Ross and J. A. Bateman and. 2009b. Agency & Information State in Situated Dialogues: Analysis & Computational Modelling. *Proceedings of DiaHolmia 2009 Workshop on the Semantics and Pragmatics of Dialogue*.

S. Clark, J. Hockenmaier and M. Steedman. 2002. Building Deep Dependency Structures using a Wide-Coverage CCG Parser. *In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*.

S. Sitter and A. Stein. 1992. Modelling the illocutionary aspects of information-seeking dialogues. *Journal of Information Processing and Management*, 28, 1992.

V. Mast and D. Wolter. 2013. A Probabilistic Framework for Object Descriptions in Indoor Route Instructions. *Proceedings of COSIT 2013 Conference on Spatial Information Theory*.

V. Mast, D. Couto Vale, Z. Falomir, and M. F. Elahi. 2014. Referential Grounding for Situated Human-Robot Communication. *The 18th Workshop on the Semantics and Pragmatics of Dialogue*. Edinburgh, September 1-3, 2014.