Towards User-friendly Conversational Systems

Hiroaki Sugiyama, Ryuichiro Higashinaka, and Toyomi Meguro

Abstract

Remarkable progress has been made with conversational systems in recent years, and they are becoming much more common. However, many problems remain to be solved such as errors in speech recognition and the narrow range of tractable dialogue topics. In this article, we introduce our efforts to improve the dialogue quality of our dialogue systems and to prevent dialogue breakdown using multiple dialogue robots.

Keywords: conversation, dialogue robots, dialogue breakdown detection

1. Introduction

Many robots and applications have been developed recently that are designed to converse with people. A few years ago, most conversational systems were implemented on smartphone applications. Some companies working in this field, notably Pepper (SoftBank Group Corp.) and OHaNAS (TOMY Company, Ltd.), changed direction and began developing technology to enable conversations between people and robots. Robots that can communicate with people through conversation are expected to be used as a natural interface between people and information and also as a way to improve human communication skills.

Yet how fluently can such conversational systems talk with people? People who have actually had conversations with them may have been disappointed if the robots output meaningless utterances because the robots did not understand some aspects of the human voice or did not have detailed knowledge of certain dialogue topics.

Current (especially commercial) conversational systems are developed with many hand-crafted response rules assuming that correct texts are obtained from speech recognition. This approach enables us to create appropriate and interesting response rules for frequently used user utterances. Furthermore, we can reduce the cost of developing such rules by dissociating textual appropriateness from speech recognition performance. However, it is obvious that not all of the topics of user utterances are covered by hand-crafted rules, and inappropriate system utterances can be generated when speech recognition fails. In this article, we introduce our recent work to overcome these problems.

2. Automatic response generation for various topics

Rule-based utterance generation is widely used in conversational systems. In this method, we first construct a dialogue example database that consists of pattern-response utterance pairs (called *rules*). The rules are created manually or gathered from actual dialogues. Then, a system applying this approach retrieves patterns that match a user utterance and outputs responses associated with the retrieved patterns. This rule-based approach works well when the range of dialogue topics is narrow. For example, recent rule-based systems such as A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) have repeatedly won the Loebner Prize (an artificial intelligence competition for chatterbots). However, to generate utterances for conversational systems, the huge variety of topics in conversations means that substantial

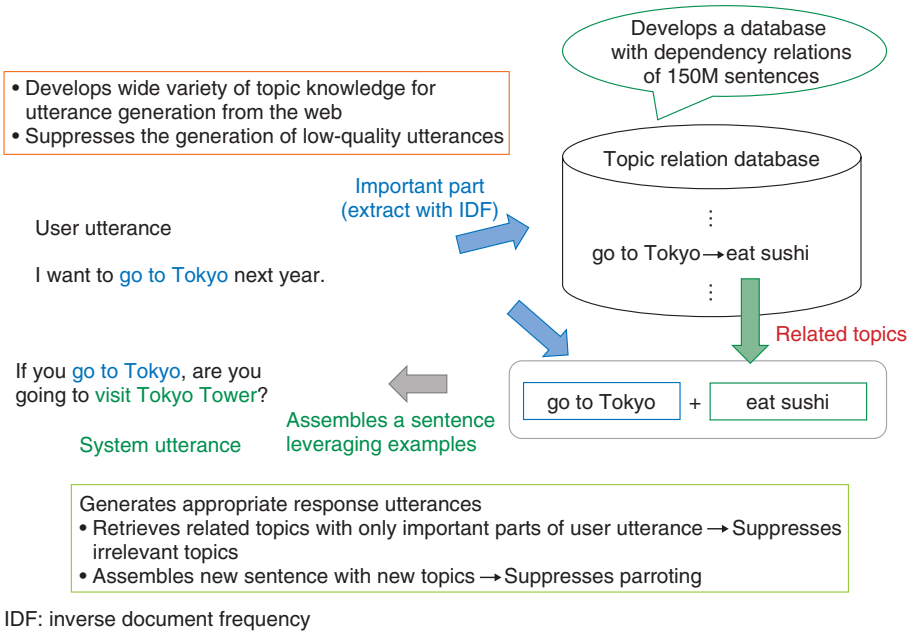


Fig. 1. New utterance generation method.

resources are required to build enough rules to cover all topics and to maintain the developed rules without contradiction.

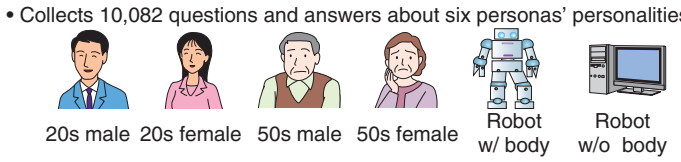
To make it feasible to automatically generate system utterances that are relevant to such a wide variety of topics of user utterances, a retrieval-based approach has been proposed. This approach retrieves sentences from the web or microblogs as system utterances by word matching with user utterances. This approach can generate responses relevant to user utterances by leveraging a wide variety of topics of web articles. However, since the retrieved sentences include the inherent contexts of the document in which the sentences originally appeared, the retrieved sentences may contain information that is irrelevant to user utterances.

To automatically define the relevancy between topics, we utilize dependency relations that express more specific relationships than normal co-occurrence. We propose an utterance generation method that combines two strongly related semantic units (phrase pairs with dependency relations that represent the topics of utterances) to create a system utterance; here, the first semantic unit is the one found in the user utterance, and the second semantic unit is the one that has a dependency relation with the first one in a large text corpus (**Fig. 1**) [1]. Our method generates utterances that have new information relevant to the current topics, which makes it easier for users to continue talking about the topic than with conventional methods.

3. Design of system *personality*

Using the method explained above, we can automatically obtain system utterances that relate to user utterances. However, is the development of conversational systems all that is required? Actually, in conversations, people often ask questions related to the specific personality or characteristics of the person with whom they are talking, for example, questions about their favorite foods or their experience playing sports. Such personality questions have reportedly appeared in conversations with task-oriented dialogue systems [2]; therefore, it is necessary to respond to such questions to achieve conversational systems. However, these questions cannot be answered with the prior utterance generation method. Moreover, if we develop question-answering systems based on information on the web, it will be difficult to maintain consistency among answers.

Therefore, we developed a question-answering system for questions that ask about an agent's specific personality, using manually created large-scale question-answer pairs. We first developed a Person Database (PDB) with large-scale personality question-answer



Question	Question category	Topics	Persona	Answer
What color do you like?	Favorite color	Favorite color	20s female	Pink
What's your favorite color?			50s male	Green
Where do you go on trips?	Where you want to travel to	Places you want to go	50s female	Bhutan
What countries do you go to?			50s male	France
Where do you go on domestic trips?	Where you go on domestic trips		20s female	Azumino

Fig. 2. Development of Person Database (PDB).

pairs for six personas gathered from many questioners and a few answerers and categorized the questions manually (**Fig. 2**) [3]. Our question-answering system responds to about 60% of personality questions and improves user satisfaction of dialogues.

4. Implementation of dialogue systems in actual robots

We have so far developed conversational systems for the text-chat format, but it is becoming more popular to use robots with speaking capabilities as a dialogue interface. To examine how our system can talk with people naturally, we collaborated with Professor Hiroshi Ishiguro at Osaka University and implemented our conversational systems in Geminoids, which are robots with human-like appearance. The architecture of this system is as follows. The system first captures user voices with a microphone and converts the voices to text using speech recognition technology. Our conversational system generates response texts for the user utterances. Finally, the text-to-speech system converts the response texts to system voices. We demonstrated this robot system at a well-known event called South by South West (SXSW) [4], and on a TV program titled "Matsuko x Matsuko."

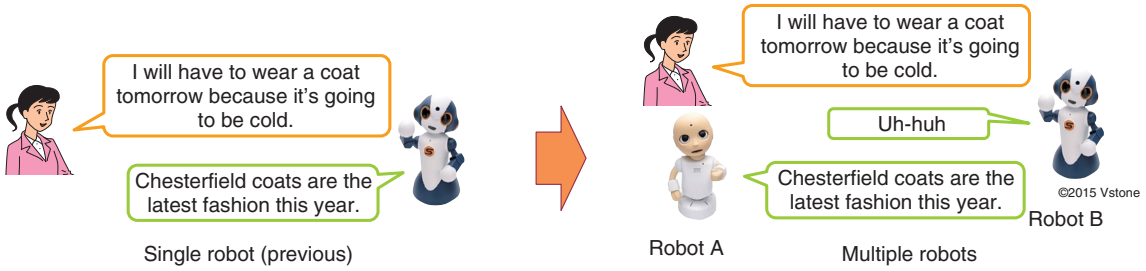
When we talked with the robot using only voice, the robot sometimes gave inappropriate utterances because of speech recognition errors, which was as we expected. Moreover, if a user said multiple sentences to the robot in rapid succession, the robot was unable to keep up with the user utterances. These difficulties were also expected. In contrast, though, some problems were unexpectedly resolved through voice conversation. For example, users that talked with the robot using voice only were more insensitive to breakdowns in dialogue logic than when text chats were used. Additionally, when a robot generated inconsistent utterances, users tended to continue the dialogue if they were facing an actual robot. This behavior was totally opposite to that observed in text chats. It has been reported that people tend to maintain relationships with dialogue partners, and we assume that this effect exists even with robots when they have a human-like appearance [5]. We are currently investigating these advantages and trying to incorporate them as fundamental techniques of dialogue systems.

5. Dialogue with multiple robots

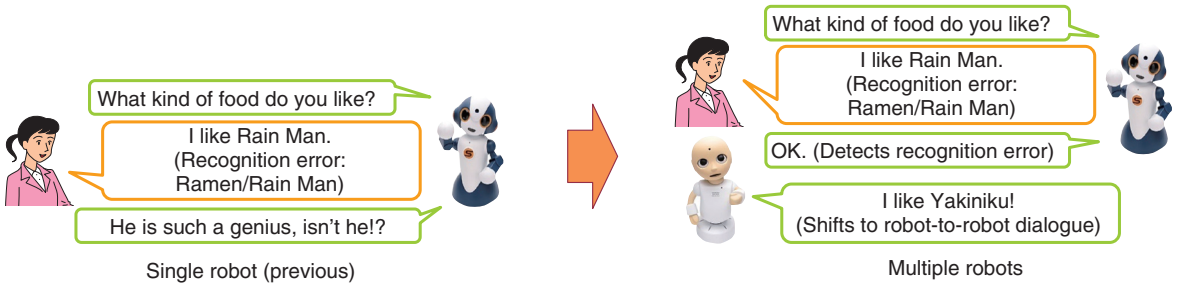
Multiple robots or computer-generated agents were reported to be effective for maintaining active and natural dialogues in system-initiative dialogue systems such as those for museum audio tour guide systems [6]. We therefore developed techniques to avoid dialogue breakdown through collaborative interaction between robots when utterance generation or speech recognition returns the wrong results (**Fig. 3**).

5.1 Utterance generation errors

An example is given in Fig. 3(a) in which the robot understands only part of the user utterance (*coat*) and generates system utterances with slightly wrong



(a) Avoiding dialogue breakdown in case of utterance generation error



(b) Avoiding dialogue breakdown in case of speech recognition error

Fig. 3. Conversation with multiple robots.

dialogue topics (*The trend this year is Chesterfield coats*). Robots that talk with users have a duty to respond appropriately to user utterances. However, in this case, since the robot cannot generate an appropriate response utterance, the user is disappointed with the robot response.

In contrast, when there are two robots, if robot A responds to a user utterance with fillers, the duty is partially fulfilled. At that time, if robot B generates the earlier utterance, this utterance can be construed as a new dialogue topic that is introduced based on the previous dialogue topic that is reacted to by robot A; therefore, the user does not sense the slight inappropriateness of the utterances and easily continues talking.

5.2 Speech recognition errors

When a critical error occurs in speech recognition, it is expected that the system and user utterances will be completely inconsistent (Fig. 3(b)). We developed a technique to avoid speech recognition errors using dialogue breakdown detection technology that identifies inconsistent utterances. When a speech recognition error is detected, our robots have a conversation according to the dialogue topic that contained the previous dialogue history. In this case, although the user feels that the robots are ignoring the user, since the dialogue topics are consistent and the dialogue itself is continuing, it is more natural than when the robots generate utterances using the wrong results of the speech recognition. We found that this technique significantly improved user satisfaction compared to the case when a single robot tried to avoid this type of dialogue breakdown with the same approach.

6. Conclusion

In this article, we introduced our text-chat based conversational systems and its implementation with one or more actual robots. We are also tackling other issues such as voice synthesis that expresses utterance intentions, automatic evaluation of conversational systems, and improvements in turn-taking to achieve user-friendly conversational robots.

Acknowledgments

This article describes collaborative work done with

Professor Hiroshi Ishiguro at Osaka University and with NTT Media Intelligence Laboratories.

References

[1] H. Sugiyama, T. Meguro, R. Higashinaka, and Y. Minami, "Open-domain Utterance Generation Using Phrase Pairs based on Dependency Relations," Proc. of the 2014 IEEE Workshop on Spoken Language Technology, South Lake Tahoe, Nevada, USA, Dec. 2014.

[2] L. C. Tidwell and J. B. Walther, "Computer-mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations: Getting to Know One Another a Bit at a Time," Human Commun. Res., Vol. 28, No. 3, pp. 317–348, 2002.

[3] H. Sugiyama, T. Meguro, R. Higashinaka, and Y. Minami, "Large-scale Collection and Analysis of Personal Question-answer Pairs for Conversational Agents," Proc. of IVA 2014 (the 14th International Conference on Intelligent Virtual Agent), pp. 420–433, Boston, MA, USA, Aug. 2014.

[4] L. Ulanoff, "Eerie Geminoid Robot Can Now Carry on a Conversation," Mashable, Mar. 2016. http://mashable.com/2016/03/13/geminoid-robot-conversation/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Mashable+%28Mashable%29#IT9WelTlkkqR

[5] A. Baylor and S. J. Ebbers, "Evidence that Multiple Agents Facilitate Greater Learning," Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies, pp. 377–379, 2003.

[6] S. Takeuchi, T. Cincarek, H. Kawanami, H. Saruwatari, and K. Shikano, "Construction and Optimization of a Question and Answer Database for a Real-environment Speech-oriented Guidance System," Proc. of the 9th International Committee for the Co-ordination an Standardization of Speech Databases and Assessment Techniques, pp. 149–154, Penang, Malaysia, Dec. 2007.

Hiroaki Sugiyama

Researcher, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

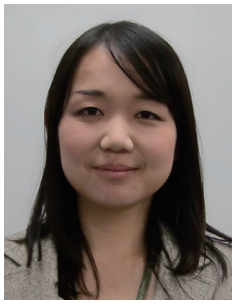
He received a B.E. and M.E. in information science and technology from the University of Tokyo in 2007 and 2009, and a Ph.D. in engineering from the Nara Institute of Science and Technology. He joined NTT Communication Science Laboratories in 2009 and studied chat-oriented dialogue systems and language development of human infants. He is a member of the Japanese Society for Artificial Intelligence (JSAI).

Ryuichiro Higashinaka

Senior Research Scientist, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

Senior Research Scientist, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.A. in environmental information, a Master of Media and Governance, and a Ph.D. from Keio University, Kanagawa, in 1999, 2001, and 2008. He joined NTT in 2001. His research interests include building question-answering systems and spoken dialogue systems. From November 2004 to March 2006, he was a visiting researcher at the University of Sheffield in the UK. He received the Maejima Hisoka Award from the Tsushinbunka Association in 2014. He is a member of the Institute of Electronics, Information and Communication Engineers, JSAI, the Information Processing Society of Japan, and the Association for Natural Language Processing.



Toyomi Meguro

Senior Research Scientist, NTT Communication Science Laboratories.

She received a B.E. and M.E. in electrical engineering from Tohoku University, Miyagi, in 2006 and 2008. She joined NTT in 2008. Her research interests are building listening agents and recommendation systems.