Avoiding Breakdown of Conversational Dialogue through Inter-Robot Coordination

Socially Interactive Agents Track

Hiroaki Sugiyama
NTT Communication Science Laboratories
Kyoto, Japan
sugiyama.hiroaki@lab.ntt.co.jp

Toyomi Meguro
NTT Communication Science Laboratories
Kyoto, Japan
meguro.toyomi@lab.ntt.co.jp

Yuichiro Yoshikawa
Osaka University
Osaka, Japan
yoshikawa@irl.sys.es.osaka-u.ac.jp

Junji Yamato
Kogakuin University
Tokyo, Japan
yamato@cc.kogakuin.ac.jp

ABSTRACT

Although conversational dialogue systems are required for continuing long conversations with users to build relationships, they sometimes make sentences that are not related to the dialogue context, causing the dialogue to easily break down. We propose a novel dialogue system framework with which two robots coordinate to create long conversations by avoiding dialogue breakdown.

KEYWORDS

Conversational dialogue systems; Multiple robots; Avoiding dialogue breakdown

ACM Reference Format:
Hiroaki Sugiyama, Toyomi Meguro, Yuichiro Yoshikawa, and Junji Yamato. 2018. Avoiding Breakdown of Conversational Dialogue through Inter-Robot Coordination. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION

Building conversational dialogue systems has recently gained much attention [6, 11, 15]. People naturally talk to each other not only for such actual benefits as information exchange or consensus building but also to establish and maintain social ties with dialogue partners [2, 3, 9, 12]. One crucial problem in the development of conversational dialogue systems is dialogue breakdown (a situation in a dialogue where users cannot proceed with the conversation) [4]. It is difficult to generate system utterances that are as appropriate as those of users because systems possess an insufficient amount of commonsense knowledge, which is obviously inferior to that of the users. Even though the detection of incoherent/discontinuous utterances in conversational dialogue systems (dialogue breakdown detection) has become a hot topic recently [4, 5], the current level of performance remains insufficient to accurately distinguish between natural and unnatural utterances [7, 13].

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

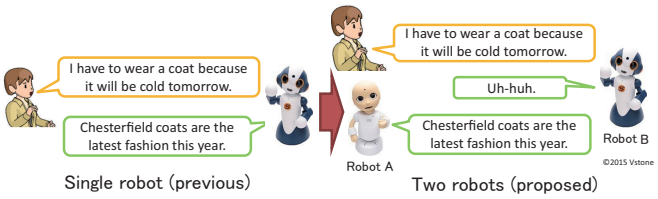

Figure 1: Example of avoiding dialogue breakdown with interruption by another robot

We propose a novel framework that develops a conversational dialogue system with two agents who coordinate to continue dialogues by avoiding breakdown. By configuring our system with two robots, the system can switch speaker robots for each system utterance (Figure 1). If one of the two robots accepts user utterances with backchannel, since the responsibility of responding to user utterances is resolved by the backchannel, the other robot can speak unnatural utterances that do not match the detailed contents of the user utterances. We investigate the effectiveness of inter-robot coordination for avoiding dialogue breakdown.

2 SYSTEM

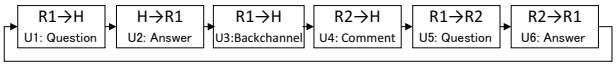

Figure 2: Dialogue flow

2.1 Dialogue flow

Figure 2 shows a dialogue flow performed by two robots. Each box indicates the behavior of one of the speakers at a certain timing, and A → B in the upper part of the box indicates that speaker A is talking to speaker B. Speaker R1 represents the first robot, R2 represents the second robot, and H represents the human user.

The dialogue begins with greetings and question U1 (e.g., *What is your favorite food?*) from R1 to H. H responds to question with answer U2 (*I like miso ramen.*), and R1 responds to H with backchannel U3 (*Oh, I see. Miso ramen.*) and R2 makes an additional comment U4 (*Miso ramen' broth is a bit strong.*), which is associated with H's answer U2. After comment U4, R1 and R2 make QA pair U5 (*Do you like ramen?*) and U6 (*Yes, I do.*), which are related to additional comment U4. Then the flow returns to question U1. This flow is based on the insight of an attentive-listening (listening-oriented)

dialogue [10] that shows the "listener" speaker first accepts the user's utterance with backchannel and self-disclosure, and then he/she asks questions in the attentive-listening dialogue.

By interposing an inter-robot QA (U5 and U6) just after such incorrect comment utterances, even if comments U4 are unnatural, we can recover from the failures and improve the chances of establishing an entire dialogue. By changing the comment-speaking robot from the robot that is making the backchannel comments, it may improve the continuity of the topics and reduce the impression that a breakdown is happening [1], since once the user utterance is resolved or received by the backchannel-speaking robot and the responsibility of responding to it is discharged, the other robot can change the dialogue topic. To should be noted that, to benefit from comment-speaker switching, backchannel utterances are required to accept user utterances.

2.2 System implementation

Questions U1 and U5 are retrieved from a large-scale QA pairs about the preferences and experiences of the speakers [14] with the highest sentence similarity to the last two non-backchannel utterances (U1: U5 and U6, U5: U2 and U4). We adopt Word Mover's Distance (WMD) to calculate the similarity among sentences [8].

Backchannels U3 are basically ambiguous responses such as "oh" or "I see." In addition, if a noun is included in the user utterance, the system repeats it as *Oh, I see.* <*NOUN*>. For obtaining additional comments U4, the system leverages an open-domain utterance-generation method [6], which utilizes predicate-argument structures and word co-occurrences in large text corpora and manually created rules to generate system utterances. As answer U6, a response sentence associated with question U5 is uttered.

As a dialogue interface, we employed two units of CommU developed by VSTONE. This robot can control gaze flexibly, which is important for humanlike social behaviors such as turn-taking.

3 EXPERIMENT

3.1 Experiment settings

We examine that coordinated behaviors of two robots effectively alleviate the feeling of dialogue breakdown. To scrutinize the effects of multiplying robots, we compared six settings (Table 1) that correspond to the variations of the coordinated behaviors. Here if we do not change the comment speaker robot, the bystanding (not speaking) robot nods while watching the speaking robot.

We employed 23 Japanese speakers as experiment participants (10 males, 13 females, whose ages ranged from 20 to 50, none of whom had experienced dialogues with multiple robots). Each speaker engaged in 12, 4-minute dialogues with the robots (six settings done twice). After each dialogue, they evaluated it with the feeling of breakdown (whether the dialogue has broken down) on a 5-point Likert scale. In this experiment, all the differences were examined using a Wilcoxon signed-rank test ($p < .10$ for significance) corrected with the Holm-Sidak method for multiple testing.

3.2 Results and analysis

We investigated the research question that both the comment speaker switching and the inter-robot QA are necessary for the improvement or not, by comparing S5 (*No inter-robot QA / 2 robots*) and S6 (*No switching / 2 robots*) with S2 (*All / 1 robot*).

Table 1: Differences among dialogue settings. B.C. denotes backchannel and Switch denotes that R2 speaks additional comments U4 instead of R1.

ID	Setting name	# of robots	B.C.	Switch	QA
S1	All / 2 robots	2	✓	✓	✓
S2	All / 1 robot	1	✓		
S3	No backchannel / 2 robots	2		✓	✓
S4	No backchannel / 1 robot	1			
S5	No QA / 2 robots	2	✓	✓	
S6	No switching / 2 robots	2	✓		✓

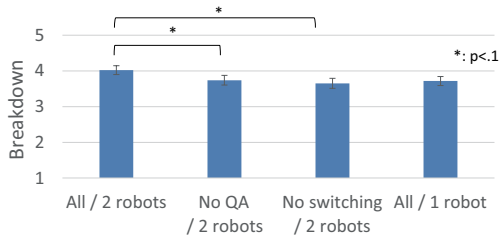

Figure 3: Breakdown comparison (higher is better)

Figure 3 shows no significant differences among S5 (*No inter-robot QA / 2 robots*), S6 (*No switching / 2 robots*) and S2 (*All / 1 robot*). In contrast, the S1(*All / 2 robots*) is significantly superior to both of the settings of two robots. This is clear evidence that both comment-speaker switching and inter-robot QA are necessary for improving the dialogue establishment.

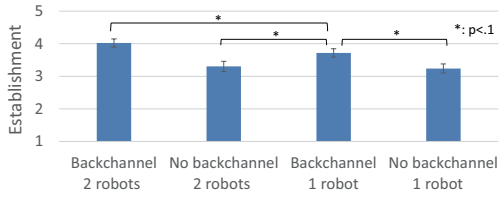

Figure 4: Comparison among the settings that vary the condition of backchannel and the number of robots

Figure 4 shows that the comparisons among the settings that vary the condition of backchannel and the number of robots. Since the backchannel with one robot is significantly better than no backchannel settings, the backchannel itself is effective for improving the feeling of dialogue breakdown. In addition, the backchannel with two robots is significantly superior to the backchannel with one robot. The backchannel is necessary to produce the positive effect of inter-robot interaction.

4 CONCLUSION

We proposed a novel framework for developing a conversational dialogue system with multiple robots and verified the effectiveness for improving the impression of dialogue breakdown. Our experiment shows that the coordinated behaviors of two robots that appropriately switch the speaker robot and create an inter-robot question-answering situation significantly improves the user impression of dialogue breakdown.

The main limitation of this work is that we fixed the robots' utterance order. The impressions of users might change if we allowed turn-taking to be done freely. In the future we intend to improve our experimental protocols and achieve unrestricted turn-taking where users can interrupt the robot utterances.

REFERENCES

[1] Tsunehiro Arimoto, Yuichiro Yoshikawa, Hiroaki Sugiyama, Toyomi Meguro, Junji Yamato, and Hiroshi Ishiguro. 2016. Generation of Between-Robots Conversation for Concealing Discontinuity in Dialog Focuses. In *Proceedings of Dialogue System Symposium*. 51–52.

[2] Christine Cheepen. 1988. *The predictability of informal conversation*. London: Pinter.

[3] Suzanne Eggins and Diana Slade. 1997. *Analysing casual conversation*. London: Cassell publishing.

[4] Ryuichiro Higashinaka, Kotaro Funakoshi, Michimasa Inaba, Yuiko Tsunomori, Tetsuro Takahashi, and Nobuhiro Kaji. 2017. Overview of Dialogue Breakdown Detection Challenge 3. In *Proc. Dialogue system technology challenge 6*.

[5] Ryuichiro Higashinaka, Kotaro Funakoshi, Yuka Kobayashi., and Michimasa Inaba. 2016. The dialogue breakdown detection challenge: Task description, datasets, and evaluation metrics. In *Proceedings of 10th edition of the Language Resources and Evaluation Conference*. 3146–3150.

[6] Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. 2014. Towards an open-domain conversational system fully based on natural language processing. In *Proceedings of the 25th International Conference on Computational Linguistics*. 928–939.

[7] Ryuichiro Higashinaka, Toyomi Meguro, Kenji Imamura, Hiroaki Sugiyama, Toshiro Makino, and Yoshihiro Matsuo. 2014. Evaluating coherence in open domain conversational systems. In *Proceedings of Interspeech*. 130–134.

[8] Matt J Kusner, Yu Sun, Nicholas I Kolkin, and Kilian Q Weinberger. 2015. From Word Embeddings To Document Distances. In *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37. 957–966.

[9] John Laver. 1975. Communicative functions of phatic communion. In *Organization of behavior in face-to-face interaction*. 215–238.

[10] Toyomi Meguro, Ryuichiro Higashinaka, Yasuhiro Minami, and Kohji Dohsaka. 2010. Controlling Listening-oriented Dialogue using Partially Observable Markov Decision Processes. In *Proceedings of the 23rd International Conference on Computational Linguistics*. 761–769.

[11] Alan Ritter, Colin Cherry, and William B. Dolan. 2011. Data-Driven Response Generation in Social Media. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 583–593.

[12] Diana Slade Scott Thornbury. 2006. *Conversation: From Description to Pedagogy*. Cambridge University Press. 377 pages.

[13] Hiroaki Sugiyama. 2016. Generation, Chat-oriented Dialogue Breakdown Detection based on the Analysis of Error Patterns in Utterance. In *Proceedings of Dialogue System Symposium*. 81–84.

[14] Hiroaki Sugiyama, Toyomi Meguro, and Ryuichiro Higashinaka. 2014. Large-scale Collection and Analysis of Personal Question-answer Pairs for Conversational Agents. In *Proceedings of International Conference on Intelligent Virtual Agents*. 420–433.

[15] Wilson Wong, Lawrence Cavedon, John Thangarajah, and Lin Padgham. 2012. Strategies for Mixed-Initiative Conversation Management using Question-Answer Pairs. In *Proceedings of the 24th International Conference on Computational Linguistics*. 2821–2834.