

Introduction method for argumentative dialogue using paired question-answering interchange about personality

Kazuki Sakai¹, Ryuichiro Higashinaka², Yuichiro Yoshikawa¹,
Hiroshi Ishiguro¹, and Junji Tomita²

¹Osaka University / JST ERATO

²NTT Media Intelligence Laboratories, NTT Corporation

¹{sakai.kazuki,yoshikawa,ishiguro}@irl.sys.es.osaka-u.ac.jp

²{higashinaka.ryuichiro,tomita.junji}@lab.ntt.co.jp

Abstract

To provide a better discussion experience in current argumentative dialogue systems, it is necessary for the user to feel motivated to participate, even if the system already responds appropriately. In this paper, we propose a method that can smoothly introduce argumentative dialogue by inserting an initial discourse, consisting of question-answer pairs concerning personality. The system can induce interest of the users prior to agreement or disagreement during the main discourse. By disclosing their interests, the users will feel familiarity and motivation to further engage in the argumentative dialogue and understand the system's intent. To verify the effectiveness of a question-answer dialogue inserted before the argument, a subjective experiment was conducted using a text chat interface. The results suggest that inserting the question-answer dialogue enhances familiarity and naturalness. Notably, the results suggest that women more than men regard the dialogue as more natural and the argument as deepened, following an exchange concerning personality.

1 Introduction

Argumentation is a process of reaching consensus through premises and rebuttals, and it is an important skill required in daily life (Scheuer et al., 2010). Through argumentation, we can not only reach decisions, but also learn what others think. Such decision-making and the interchange of views are one of the most important and advanced parts of human activities. If an artificial dialogue system can argue on certain topics with us, this can both help us to work efficiently and establish a close relationship with the system.

Recently, there have been some studies concerning argumentative dialogue systems. Higashinaka et al. developed an argumentative dialogue system that can discuss certain topics by using large-scale argumentation structures (Higashinaka et al., 2017). However, this system could not provide all users with a satisfactory discussion experience, even though it could appropriately respond to their opinions. One possible reason for this is that some users are not necessarily motivated to argue on the topics suggested by the system.

We aim to improve an argumentative dialogue system by adding a function to motivate a user to participate in an argumentative dialogue. To increase the user's motivation to participate in the argumentative dialogue, we focus on small talk. Small talk can help participants build certain relationships before they enter the main dialogue (Zhao et al., 2014). In negotiation and counseling, a close relationship between two humans can improve the performance of certain tasks (Drolet and Morris, 2000; Kang et al., 2012). Relationships between a user and system are important for reaching a consensus through dialogue (Katagiri et al., 2013). Thus, it is considered to be possible for a user to be naturally guided into an argumentative dialogue by performing small talk.

In practice, we adopted a question-answering dialogue, where users are casually asked about their personal experiences or ideas. This was implemented by using what we call a personal database (hereafter PDB), which involves pairs consisting of a personal question and a corresponding example answer, which are likely to appear in human-human conversation. When asked about personal issues, users are expected to feel interested in the system, and then be induced to feel open and close to the system. Meanwhile, the system provides its own answers to the questions

by using the PDB. From the answers of the user and the system, users are expected to gain an idea of what is common and different between them, a requirement which has been suggested to be important for humans to be motivated to understand one another (Uchida et al., 2016).

In this research, we extend the argumentative dialogue system described in (Higashinaka et al., 2017) to add a function that can smoothly introduce argumentative dialogue by inserting a question-answering dialogue using the PDB (hereinafter referred to as PDB-QA dialogue). It is considered that users of the proposed system can be expected to be motivated to partake in the argumentative dialogue, and that they can then partake in a deep discussion with the system. To verify the effectiveness of this system, we conducted a subjective experiment using a text chat interface.

The remainder of this paper is organized as follows. In Section 2, we describe related work. In Section 3, we describe our proposed method, including how to develop the question-answering dialogue and how to integrate this into an existing argumentative dialogue system. In Section 4, we describe an experiment we conducted, in which human subjects expressed their impressions of the dialogue through a text chat interface. We summarize the paper and discuss future work in Section 5.

2 Related work

Although there is little work on an automated system that can perform discussion with users, recently, there has been a great deal of work aimed at automatically extracting premises and conclusions from text; argumentation mining has been applied to various data, including legal text (Moens et al., 2007), newswire text (Bal and Saint-Dizier, 2010), opinions in discussion forums (Rosenthal and McKeown, 2012), and varied online text (Yanai et al., 2016).

There has been some research concerning the introduction of a dialogue. Rogers et al. showed that it became easier for two people to talk during the first meeting by using an application that can share their opinions on a display (Rogers and Brignull, 2002). Patricia et al. reported that small talk in an initial discourse improved the interaction in a business situation (Pullin, 2010). Inaguma et al. analyzed the prosodic features of shared laughter as an icebreaker in initial dialogues (Inaguma et al., 2016).

However, it is unclear how to develop an initial dialogue for smoothly introducing a discussion.

It is known that people interact with artificial constructions such as dialogue systems, virtual agents, and robots in the same manner as they interact with other humans (Reeves and Nass, 1996). Schegloff et al. showed that human conversation usually interleaves the contents of a task-oriented dialogue with social contents (Schegloff, 1968). Jiang et al. showed that 30% of all utterances of Microsoft Cortana, a well-known task-oriented dialogue system, consist of social contents (Jiang et al., 2015). It is considered that performing small talk can be natural in argumentative dialogue systems.

There have been many studies on dialogue systems that include small talk. Bechberger et al. developed a dialogue system that conveys news text and performs small talk related to the news (Bechberger et al., 2016). Kobori et al. showed that inserting small talk improved the impressions of an interview system (Kobori et al., 2016). Bickmore et al. showed that the task success rate was improved by constructing a trust relationship using small talk (Bickmore and Cassell, 2005). Tina et al. developed a dialogue system that included the function of interacting using small talk (Klüwer, 2015). We consider that argumentative dialogues may be performed deeply since small talk can improve the trust relationship.

Related to the studies dealing with multiple dialogue strategies including argumentative and social dialogues, there are several works concerning hybrid dialogue systems that integrate task-oriented and chat-oriented dialogue systems. Papaioannou et al. proposed a method to acquire dialogue strategies for hybrid systems in a robot using reinforcement learning (Papaioannou and Lemon, 2017). Yu et al. showed that multiple dialogue systems can interact using appropriate dialogue strategies learned through reinforcement learning (Yu et al., 2017). Akasaka et al. demonstrated a classification method for input utterances to select what dialogue systems are used (Akasaki and Kaji, 2017). However, in initial dialogue, it is unclear which dialogue strategies can be employed to smoothly introduce an argumentative dialogue.

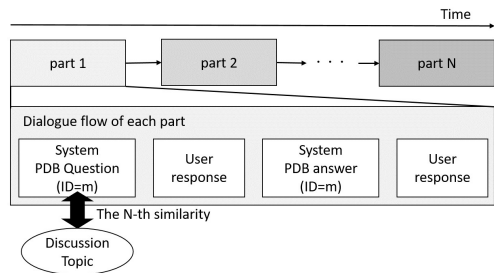


Figure 1: Flow of PDB-QA dialogue. Each part contains two system utterances and two user utterances. We used questions in an order based on the similarity between the dialogue topic and question text.

3 Proposed method

We propose a method for introducing an argumentative dialogue using the PDB-QA dialogue, which is a question-answering dialogue concerning personality. We then describe some existing argumentative dialogue systems. Next, we explain how to develop an extended argumentative dialogue system using the PDB-QA dialogue.

3.1 PDB-QA dialogue by using question-answering pair about personality

The PDB consists of personal questions and example answers and is used to ask the interlocutor for detailed information (Tidwell and Walther, 2002). Such questions may be asked even when the interlocutor is a dialogue system (Nisimura et al., 2011). In this study, we used the PDB described in (Sugiyama et al., 2014). This PDB is a large-scale database of pairs of questions and answers related to personal information. Questions included in the PDB involved various personal questions, question categories, answer examples, and topics attached to each question. Based on the degree of overlap of questions, question-answer pairs frequently encountered during conversation are extracted. The PDB includes personal questions such as “what dishes do you like?” and “which places have you visited?”

We explain the procedure for generating a PDB-QA dialogue using this PDB. As shown in Figure 1, the PDB-QA dialogue consists of several parts. Each part consists of four utterances: the system’s question using the PDB, the user’s response, the system’s answer, and the user’s response to this. To determine the order in which

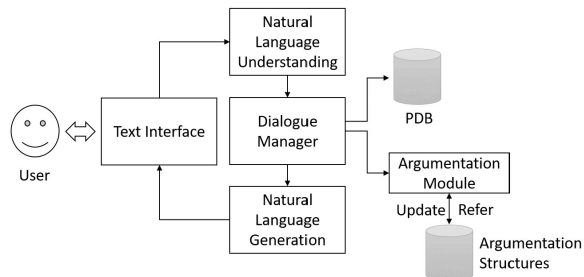


Figure 2: Architecture of developed dialogue system.

to ask multiple questions, we used the similarity between the topic of argument and the question text, calculated by Word2vec (Mikolov et al., 2013). From parts 1 to N, we used questions in an order starting from the highest similarity, i.e., part 1 uses a question with the N-th highest similarity and part N uses another question that has the highest similarity. This is because it is considered that approaching the topic gradually is natural as a dialogue structure. Through this process, we can perform N parts of the PDB-QA dialogue.

3.2 Argumentative dialogue system

We used the argumentative dialogue system described in (Higashinaka et al., 2017). This system can generate appropriate argumentative dialogue text based on large-scale knowledge structures, called argumentation structures, which are constructed manually. An argumentation structure is represented by a graph structure, composed of nodes that represent premises and edges representing support or nonsupport relationships, based on an extended version of Walton’s model (Walton, 2013).

A user utterance is input into two modules: dialogue act estimation and proposition identification. The dialogue act estimation module estimates four dialogue-act types: assertion, question, concession, and retraction. The proposition identification module determines the argumentation node that contains the content closest to the input user utterance. The discussion manager updates the argumentation structure on the basis of the understanding result, which checks whether the corresponding node is already mentioned. Then the dialogue manager retrieves premises that can be used for support or rebuttal based on traversing along with argumentation structures. The system outputs a supportive or nonsupportive response to

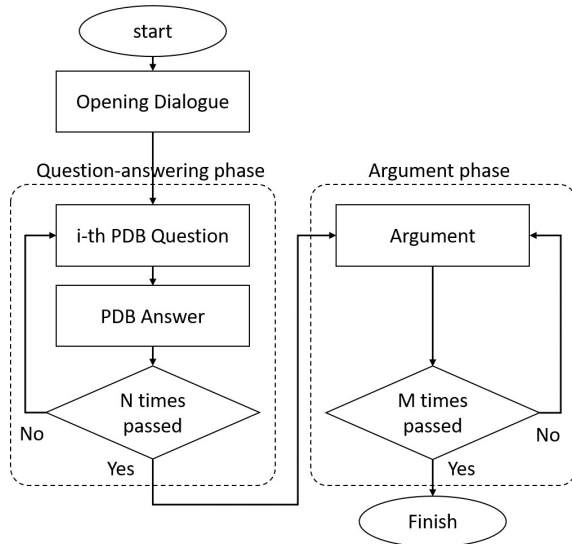


Figure 3: Flow in the dialogue manager.

the user’s utterance.

3.3 Integration of argumentation dialogue system and PDB-QA dialogue

Figure 2 illustrates the architecture of our argumentative dialogue system. The user interacts with the system through the text chat interface on the browser. The natural language understanding module has two modules related by the argumentative dialogue system. Note that this module is only used in the argument phase described as follows. The dialogue manager manages two dialogue states. One is the question-answering phase, and the other is the argument phase. Figure 3 illustrates the flow of dialogue managed by the dialogue manager. First, the dialogue manager initiates the opening dialogue, such as by asking the user her name. Then, it begins the question-answering phase. In this phase, the PDB-QA dialogue is performed, as described in Section 3.1. The PDB-QA dialogue is a predefined question-answering dialogue, regardless of user utterances. The answer of the system for a PDB question is prepared by the experimenter in advance. The natural language generation module changes the system response such as adding conjunctions and changing the end of the sentences using a dialogue act. Later, the dialogue manager begins the argument phase. In the argument phase, the utterances of the system are premises that can be used for support or rebuttal, and they consist of the results of the argumentative dialogue system.

Line	Speaker	Utterance
1	S	Do you like trips?
2	U	Yes, I do.
3	S	I see. I like them too. I want to go on a trip during the weekend.
4	U	I also want to go during the weekend.
5	S	Where do you want to go for a trip?
6	U	I want to go abroad.
7	S	Me too. I want to go to France if I have the time.
8	U	French dishes seem great.
9	S	Which is the better place to travel in Japan, Hokkaido or Okinawa?
10	U	I want to go to Okinawa.
11	S	You can enjoy unique dishes in Okinawa.
12	U	I think goya-champuru (a traditional Okinawa dish) is delicious.
13	S	But Hokkaido also has delicious dishes, such as Genghis Khan.
14	U	I see. Hokkaido has more delicious food.

Figure 4: An example of the dialogue. The topic is that of which place the better to travel to in Japan: Hokkaido or Okinawa. Lines 1 ~ 8 are part of the PDB question dialogue, and lines 9 ~ 14 are part of the argumentative dialogue. Speaker S and U represent the system and user, respectively.

Figure 4 shows an example of the dialogue we performed. The topic is as follows: which is the better place to travel to in Japan: Hokkaido or Okinawa? Lines 1 ~ 8 are a part of the PDB question dialogue, and lines 9 ~ 14 are a part of the argumentative dialogue. Speakers S and U represent the system and the user, respectively.

4 Experiment

In this section, we describe a subjective experiment to verify the effect of inserting the PDB-QA dialogue. We compared the subjects’ evaluations and behavior for two types of dialogue: one with PDB-QA and the other without it. The hypothesis is that by inserting the PDB-QA dialogue in advance, users are motivated to partake in the argumentative dialogue and can then discuss deeply with the system. To verify this hypothesis, subjects communicated with the argumentative dialogue system through a text chat interface on a browser, and then recorded their impressions in a questionnaire. We quantitatively evaluated the average number of words per utterance of the user in the argument phase. It is expected that the number of words per user’s utterance in our argumentative dialogue system should be relatively lower than that in the previous system, because when a

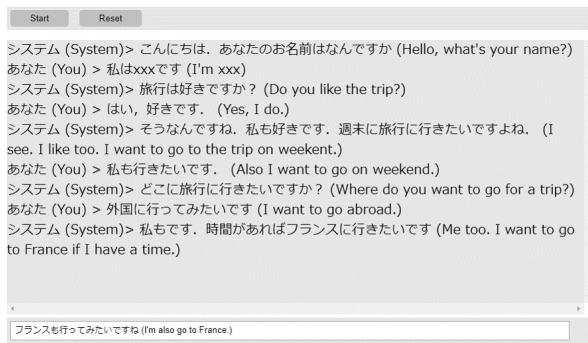


Figure 5: Screenshot of the text chat interface.

user builds a relationship with the system, the user expresses own ideas with fewer words.

4.1 Method

4.1.1 Subjects

Thirty-two Japanese adults (16 males and 16 females, with an average age of 20.3 years) participated as subjects. Half of the subjects participated with the PDB condition, and the other half without it. The ratio of males to females in each condition was the same. One male with the PDB condition and two males without were excluded because of system failures, and the utterances of the remaining 29 people were analyzed.

4.1.2 Apparatus

The experiment was conducted in a space separated by curtains. A laptop PC was placed on the table, and the PC displayed a web browser to show the text chat interface, as shown in Figure 5. Note that the dialogue in the experiment was performed in Japanese. The dialogue text of the interaction between the system and the subject was displayed in the middle part of the browser, and a text box for the subject to input his/her own utterances was displayed at the lower part of the browser. Note that we call the sentence displayed in the interface an “utterance.” In other words, sentences produced by the system and input by the user with a keyboard are called the system’s and user’s utterances, respectively.

4.1.3 Stimuli

In this experiment, we compared two conditions: with and without the PDB. The condition with PDB included two phases of dialogue: a question-answering phase and an argument phase. The condition without PDB included one phase of dialogue: the argument phase. In this experiment, the

subject and the system alternately provided utterances. Each pair of such utterances is referred to as one turn. Both conditions included two turns of opening dialogue, such as asking the subject’s name and a greeting. The question-answering phase consisted of three parts, each of which included two turns of dialogue. In total, six turns of dialogue were performed. The argument phase contained six turns of dialogue. We prepared five discussion topics and assigned any of these to the subject at random: (1) the pros and cons of driving automobiles, (2) benefits of living in the countryside vs. living in the city, (3) which is the better place to travel to in Japan between Hokkaido and Okinawa, (4) which is the better breakfast between bread and rice, and (5) which is the better theme park between Tokyo Disney Resort and Universal Studios Japan.

4.1.4 Procedure

This experiment was conducted according to the following procedure. First, the experimenter gave a subject the instructions for the experiment. The contents of the instructions were that the subject interacts with the system through the text chat interface on the browser, interacts only once, and answers the questionnaire after the dialogue. Next, the experimenter asked the subject to read the questionnaire in advance. After that, interaction was started. After completing the dialogue, the experimenter asked the subject to answer the questionnaire.

4.1.5 Measurement

The items of the questionnaire regarding impressions were the same for both conditions, and there were eleven items in total. These included questions related to the overall impression of the dialogue system, the argumentative dialogue, and the user’s motivation for conversing with the dialogue system. The items concerning the impression of the dialogue consisted of the following five:

- Q1 The utterances of the system are correct in Japanese,
- Q2 The dialogue with the system is easy to understand,
- Q3 The dialogue with the system is familiar,
- Q4 The dialogue with the system has a lot of content, and
- Q5 The dialogue with the system is natural.

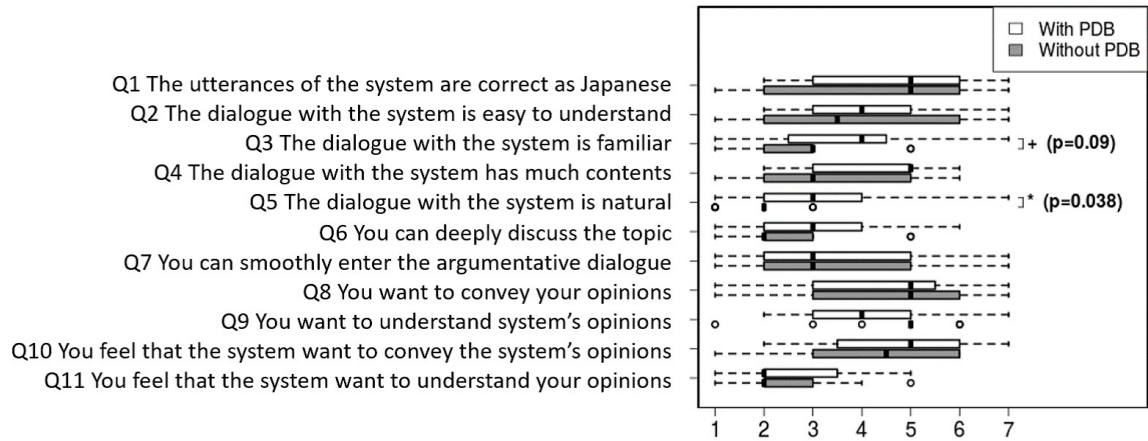


Figure 6: Box plots of the results of the questionnaire.

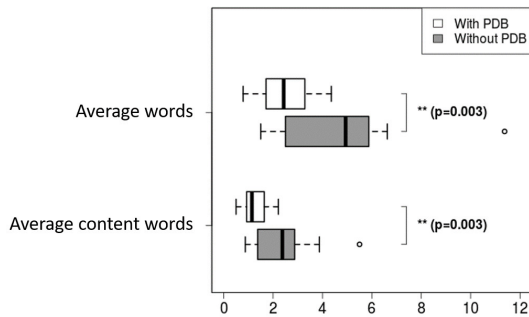


Figure 7: Box plots of results for the average numbers of words and content words (nouns, verbs, adjectives, conjunctions, and interjections) per utterance in the argument phase. We used MeCab to tokenize Japanese words.

The items concerning the impression of the argument dialogue were the following two:

- Q6** You can deeply discuss the topic of X, and
- Q7** You can smoothly enter the argumentative dialogue about X,

where X is the actual topic (e.g., which is the better place to travel to in Japan between Hokkaido and Okinawa). The items related to motivation for the dialogue were the following four:

- Q8** You want to convey your opinions,
- Q9** You want to understand the system’s opinions,
- Q10** You feel that the system wants to convey its opinions, and
- Q11** You feel that the system wants to understand your opinions.

A Likert scale was used to elicit the subjects’ impressions. We used a seven-point scale

that ranged from a value of 1, corresponding to “strongly disagree,” to 7, corresponding to “strongly agree.” The midpoint value of 4 corresponded to “undecided.”

We also counted the average number of words per user utterance and the average number of content words (nouns, verbs, adjectives, conjunctions, and interjections) in the argument phase. We used MeCab to tokenize the words and label the Japanese parts of speech.

4.2 Result

Figure 6 presents the box plots of the answers to the questionnaire. A Mann-Whitney U test was used to compare the scores on the Likert scale. For Q3, namely “the dialogue with the system is familiar,” the median score for the condition with PDB was found to be marginally significantly higher than that for the condition without PDB ($W = 143, p < 0.1$). For Q5, namely “the dialogue with the system is natural,” the median score for the condition with PDB was found to be significantly higher than that for the condition without PDB ($W = 149.5, p < 0.05$). For other questions, no significant differences between the two conditions were detected.

As shown in Figure 6, we did not directly confirm an improvement concerning the smooth introduction to the argumentative dialogue by inserting the PDB-QA dialogue. However, this figure suggests that it is possible for the user to feel that the dialogue is familiar and more natural when the PDB-QA dialogue is inserted. This result may be because the system performs in the manner in which a human usually does, and a certain rela-

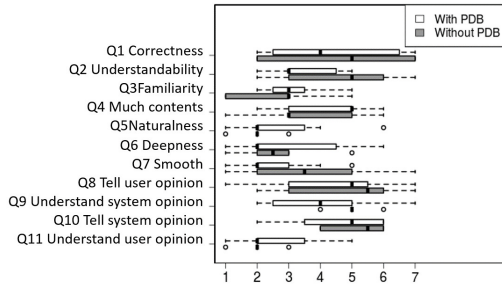


Figure 8: Box plots of results of the questionnaire for male users.

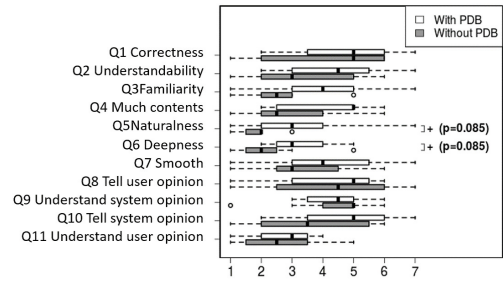


Figure 9: Box plots of female results of the questionnaire for female users.

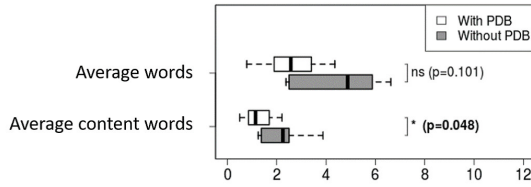


Figure 10: Box plots of results for male users for the average number of words and content words per utterance in the argumentation phase.

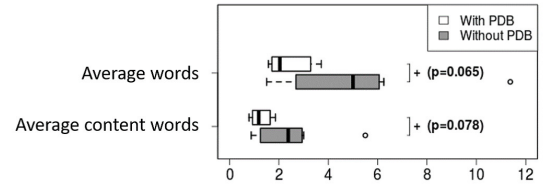


Figure 11: Box plots of results for female users for the average number of words and content words per utterance in the argumentation phase.

relationship is built between the user and the system. Thus, it is considered that inserting the PDB-QA dialogue improves the naturalness of the dialogue and relationships.

Figure 7 presents the box plots of the average numbers of words and content words per user utterance. For the average number of words, the median score for the condition with PDB was found to be significantly less than that for the condition without PDB ($W = 40, p < 0.01$). Concerning the average number of content words, the median score for the condition with PDB was also found to be significantly less than that for the condition without PDB ($W = 40, p < 0.01$).

As shown in Figure 7, it was found that the average numbers of words and content words in the condition with PDB were significantly less than those in the condition without PDB. These results suggest that when the relationship between the user and the system is not close, the users may express their opinions using a larger number of words, to correctly convey their own message; on the other hand, when the relationship is close, the users may express their opinions using fewer words.

In general, it is known that there are some differences in purposes of conversation owing to gender differences (Tannen, 2001). In this study, we

suppose that the different purposes of conversation resulting from gender differences may affect our results. Therefore, we analyzed the effects of gender. We divided the data by gender, and then plotted each result. In the result for male users, shown in Figure 8, no significant differences between the two conditions were detected. On the other hand, in the result for female users, shown in Figure 9, we observe some significant differences between the two conditions. According to this figure, for Q5, namely “the dialogue with the system is natural,” the median score for the condition with PDB was found to be marginally significantly higher than that for the condition without PDB ($W = 49, p < 0.1$). For Q6, namely “you can deeply discuss the topic,” the median score for the condition with PDB was also found to be marginally significantly higher than that for the condition without PDB ($W = 49, p < 0.1$). In addition, we compared males’ and females’ data under the conditions with and without PDB. As a result, for Q7, namely “you can smoothly enter the argumentative dialogue,” the median score with PDB for females was found to be marginally significantly higher than that with PDB for males ($W = 13.5, p < 0.1$). These results suggest that it is possible that females may feel that the PDB-QA dialogue inserted before the argumentative dialogue is more natural, and this

may lead to the result that females feel the argumentative dialogue is deepened more. Thus, it is suggested that inserting the PDB-QA dialogue in our proposed method may be more effective for females.

In addition, Figures 10 and 11 show the results for male and female users for words and content words, respectively. As shown in Figure 10, for male users, the average number of content words for the condition with PDB was found to be significantly less than that without PDB ($W = 7$, $p < 0.05$). This result may be because of their degree of motivation, but the actual reason is unknown. On the other hand, as shown in Figure 11, for female users, the average numbers of words and content words with PDB were found to be marginally significantly less than those without PDB ($W = 14$, $p < 0.1$, $W = 15$, $p < 0.1$, respectively). These results suggest that females may use fewer words when they feel familiarity with the interlocutor.

5 Summary and future work

We proposed a PDB-QA dialogue method to smoothly introduce an argumentative dialogue. We conducted an evaluation experiment to verify the effectiveness of inserting the PDB-QA dialogue. The results suggest that the impressions of the dialogue, such as familiarity and naturalness, may be improved by inserting the PDB-QA dialogue. Specifically, we found that females may perceive a PDB-QA dialogue inserted before an argumentative dialogue as more natural, and this may lead to the result that the argumentative dialogue can be deepened. We also found that when the relationship between the user and the system is not close, the users may express their opinions using a larger number of words, whereas when the relationship is close, the users may express their opinions with fewer words.

We can improve the performance of the dialogue system by adjusting several parameters of PDB dialogue, which were fixed in the experiment for the sake of control. For example, we can adjust how questions are chosen (the degree of similarity of questions to be selected), the order of questions, the number of questions, and the amount of information to be presented in an answer to a question. It may be possible to improve the performance if we select better parameters depending on a user's preferences or the context of a conversation.

For further improvement, we can consider animacy, which is another element that may be important. Animacy describes the characteristic of being like a living being, in other words, the characteristic of whether a human can relate to mind and will in an object. We suppose that in a dialogue, it is important for the user to feel animacy toward the interlocutor, because it is important for the user to recognize the dialogue system as a special target with which they can form a certain relationship. As a preliminary experiment, we measured the psychological indicators for mind perception (Gray et al., 2011). This scale can measure how much agency (capacity for self-control, planning, and memory) and experience (capacity for pleasure, fear, and hunger) the subject feels the target has. Analyzing how impressions of agency and experience might affect the answers to the questionnaire or the behavior of users will be an important aspect of future work.

In this paper, we compared the conditions with and without PDB. Comparing the two conditions, we surmise that at least three factors exist that affect the results: whether utterances are in the form of a question, whether they contain personal content, and whether they are related to the topic of the argumentative dialogue. For the first factor, we suppose that a question form can explicitly reveal common and differing sentiments in the answer to the question. It is considered that this makes it easy for the user to become interested. For the second aspect, we suppose that asking a question concerning personality can make it possible to construct a certain relationship more easily. As regards the final point, we feel this prevents a sudden change of topic. We suppose that this makes it possible for the user to enter the argumentative dialogue more smoothly. Investigating the kinds of factors that affect a natural introduction into the argumentative dialogue will be a topic of future work.

Acknowledgments

This work was supported by JST ERATO Grant Number JPMJER1401, Japan.

References

- S. Akasaki and N. Kaji. 2017. Chat detection in an intelligent assistant: Combining task-oriented and non-task-oriented spoken dialogue systems. In *ACL*.

- B. K. Bal and P. Saint-Dizier. 2010. Towards building annotated resources for analyzing opinions and argumentation in news editorials. In *Proceedings of the language resources and evaluation conference*.
- L. Bechberger, M. Schmidt, A. Waibel, and M. Federico. 2016. Personalized news event retrieval for small talk in social dialog systems. In *Proceedings of Speech Communication; 12. ITG Symposium*.
- T. Bickmore and J. Cassell. 2005. *Social Dialogue with Embodied Conversational Agents*. Advances in Natural Multimodal Dialogue Systems.
- A. L. Drolet and M. W. Morris. 2000. Rapport in conflict resolution: Accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Journal of Experimental Social Psychology*, 36(1):26–50.
- K. Gray, A. C. Jenkins, A. S. Heberlein, and D. M. Wegner. 2011. Distortions of mind perception in psychopathology. In *Proceedings of the National Academy of Sciences of the United States of America*, pages 477–479.
- R. Higashinaka, K. Sakai, H. Sugiyama, H. Narimatsu, T. Arimoto, T. Fukutomi, K. Matsui, Y. Ijima, H. Ito, S. Araki, Y. Yoshikawa, H. Ishiguro, and Y. Matsuo. 2017. Argumentative dialogue system based on argumentation structures. In *Proceedings of the 21st Workshop on the Semantics and Pragmatics of Dialogue*, pages 154–155.
- H. Inaguma, K. Inoue, S. Nakamura, K. Takanashi, and T. Kawahara. 2016. Prediction of ice-breaking between participants using prosodic features in the first meeting dialogue. In *Proceedings of the 2Nd Workshop on Advancements in Social Signal Processing for Multimodal Interaction*, pages 11–15.
- J. Jiang, A. H. Awadallah, R. Jones, U. Ozertem, I. Zitouni, R. G. Kulkarni, and O. Z. Khan. 2015. Automatic online evaluation of intelligent assistants. In *Proceedings of the 24th International Conference on World Wide Web*, pages 506–516.
- S. Kang, J. Gratch, C. Sidner, R. Artstein, L. Huang, and L. Morency. 2012. Towards building a virtual counselor: Modeling nonverbal behavior during intimate self-disclosure. In *Proceedings of 11th International Conference on Autonomous Agents and Multiagent Systems*.
- Y. Katagiri, K. Takanashi, M. Ishizaki, Y. Den, and M. Enomoto. 2013. Concern alignment and trust in consensus-building dialogues. In *Proceedings of the 9th International Conference on Cognitive Science*, pages 422–428.
- T. Klüwer. 2015. *Social talk capabilities for dialogue systems*. universaar.
- T. Kobori, M. Nakano, and T. Nakamura. 2016. Small talk improves user impressions of interview dialogue systems. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 370–380.
- T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing System*, pages 3111–3119.
- M. Moens, E. Boiy, R. M. Palau, and C. Reed. 2007. Automatic detection of arguments in legal texts. In *Proceedings of the 11th international conference on Artificial intelligence and law*, pages 225–230.
- R. Nisimura, Y. Nishihara, R. Tsurumi, A. Lee, H. Saruwatari, and K. Shikano. 2011. Takemaru-kun: Speech-oriented information system for real world research platform. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 583–593.
- I. Papaioannou and O. Lemon. 2017. Combining chat and task-based multimodal dialogue for more engaging hri: A scalable method using reinforcement learning. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 365–366.
- P. Pullin. 2010. Small talk, rapport, and international communicative competence: Lessons to learn from belf. *The Journal of Business Communication*, 47(4):455–476.
- B. Reeves and C. Nass. 1996. *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press.
- Y. Rogers and H. Brignull. 2002. Subtle ice-breaking: encouraging socializing and interaction around a large public display. In *In Workshop on Public, Community, and Situated Displays*.
- S. Rosenthal and K. McKeown. 2012. Detecting opinionated claims in online discussions. In *Proceedings of the 6th IEEE International Conference on Semantic Computing*, pages 30–37.
- E. A. Schegloff. 1968. Sequencing in conversational openings. *American anthropologist*, 70(6):1075–1095.
- O. Scheuer, F. Loll, N. Pinkwart, and B. M. McLaren. 2010. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1):43–102.
- H. Sugiyama, T. Meguro, and R. Higashinaka. 2014. Large-scale collection and analysis of personal question-answer pairs for conversational agents. In *Proceedings of Intelligent Virtual Agents*, pages 420–433.
- D. Tannen. 2001. *You just don't understand: women and men in conversation*. New York: HarperCollins.

- L. C. Tidwell and J. B. Walther. 2002. Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations. *Human Communication Research*, 28(3):317–348.
- T. Uchida, T. Minato, and H. Ishiguro. 2016. Does a conversational robot need to have its own values? a study of dialogue strategy to enhance people’s motivation to use autonomous conversational robots. In *Proceedings of the 4th annual International Conference on Human-Agent Interaction*, pages 187–192.
- D. Walton. 2013. *Methods of argumentation*. Cambridge University Press.
- K. Yanai, Y. Kobayashi, M. Sato, T. Yanase, Miyoshi T, Y. Niwa, and H. Ikeda. 2016. Debating artificial intelligence. *Hitachi Review*, 65(6):151.
- Z. Yu, A. W. Black, and A. I. Rudnicky. 2017. Learning conversational systems that interleave task and non-task content. In *Proceedings of International Joint Conference on Artificial Intelligence*.
- R. Zhao, A. Papangelis, and J. Cassell. 2014. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *Proceedings of International Conference on Intelligent Virtual Agents*, pages 514–527.