Female-Type Android's Drive to Quickly Understand a User's Concept of Preferences Stimulates Dialogue Satisfaction: Dialogue Strategies for Modeling User's Concept of Preferences

Takahisa Uchida^{1,2} · Takashi Minato¹ · Yutaka Nakamura² · Yuichiro Yoshikawa² · Hiroshi Ishiguro^{1,2}

Accepted: 21 November 2020 / Published online: 7 January 2021
© The Author(s) 2021

Abstract

This research develops a conversational robot that stimulates users' dialogue satisfaction and motivation in non-task-oriented dialogues that include opinion and/or preference exchanges. One way to improve user satisfaction and motivation is by demonstrating the robot's ability to understand user opinions. In this paper, we explore a method that efficiently obtains the concept of user preferences: likes and dislikes. The concept is acquired by complementing a small amount of user preference data observed in dialogues. As a method for efficient collection, we propose a dialogue strategy that creates utterances with the largest expected complementation. Our experimental results with a female-type android robot suggest that the proposed strategy efficiently obtained user preferences and enhanced dialogue satisfaction. In addition, the strength of user motivation (i.e., long-term willingness to communicate with the android) is only positively correlated with the android's willingness to understand. Our results not only show the effectiveness of our proposed strategy but also suggest a design theory for dialogue robots to stimulate dialogue motivation, although the current results are derived only from a female-type android.

Keywords Preference model · Concept model · Dialogue strategy · Dialogue robot · Android

1 Introduction

The purpose of this research is to develop a conversational robot that increases users' dialogue satisfaction and motivation in non-task-oriented dialogues that exchange opinions. Non-task-oriented dialogues are designed to continue conversations with users. For example, dialog robots and systems that allow continuous dialogues with users can be applied to communication support for elderly people [3,15]. Another example is building social bridges among users (e.g., [26]). Thus, the robot can improve the quality of life (QOL) of various types of users.

Some research exists on non-task-oriented dialogue systems. Some systems generate utterances based on many rules [23,28,29], and others statistically generate utterances

from large-scale corpus data [1,8,19]. Another study identified whether dialogues are linguistically broken [7,24] to avoid dialogue breakdowns. These studies mainly focused on how to generate linguistically correct responses to user utterances.

On the other hand, linguistically correct utterances do not always promote user willingness to interact with the system. Users will only want to talk to a non-task-oriented dialogue system if they feel familiar with it [2]. To improve the user's satisfaction and motivation, we demonstrate a robot's ability to understand his/her opinions. This paper focuses on user preferences (likes and dislikes) in a certain category and explores a method that demonstrates the robot's ability to understand them.

One way to demonstrate this ability is to endow the robot with the capability to estimate the user's concept of preferences and generate a dialogue based on this idea. We define a concept as a mental representation that is concerned with whether items are near each other in a certain viewpoint. For example, if the robot recognizes that the user likes A, it also estimates that he/she likes B, which resembles A from a certain perspective, and tells him/her about it. Furthermore, the robot can generate such utterances as whether A and B are

✉ Takahisa Uchida
uchida.takahisa@atr.jp

¹ Advanced Telecommunications Research Institute International, 2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan

² Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka, Japan

similar in terms of X . By generating utterances based on an estimated concept, the robot can express the ability to understand (by making abstractions and analogous reasoning) and a willingness to understand user's preferences. We hypothesize that this will improve user satisfaction and motivation regarding dialogues with the robot.

The contributions of this study are listed below.

- We propose a method that efficiently obtains the concept of user preferences in dialogues.
- Furthermore, we propose a dialogue strategy through which we efficiently build this concept.
- From the verification of the proposed methods, we clarified that the robot's willingness to understand the user improves the dialogue satisfaction and motivation.

2 Related Work

Our goal of this study is to develop a robot that understands the concept of user preferences through a dialogue while maintaining his/her satisfaction and motivation to talk. To satisfy this requirement, the robot needs to efficiently obtain the user's concept through dialogues. In this section, we describe the novelty of this study by comparing it with previous studies.

For estimating preferences, Kobayashi et al. [13] developed a dialogue system that estimates user's preferences in dialogues by recognizing the polarity of user utterances using syntax, surface case, and deep case analysis. Other methods estimate user interests by combining linguistic and non-linguistic information [9,32]. Although estimating user preferences for the current item or topic is possible, no relation between the preferences for other items, that is, the concept of user's preferences, is estimated. Collaborative filtering [18], which is commonly used to estimate user preferences in recommendation systems, estimates the preferences of a target user based on the information of many other users. However, it does not work effectively when insufficient user information is available (the sparsity problem [6]). If it is applied in a dialogue system, many questions must be asked to get user information. This requirement decreases motivation to interact with it.

An information provision system [22] can quickly estimate the user's concept in dialogues. It is enabled by using the concept spaces of multiple people acquired in advance. However, a method has been proposed for task-oriented dialogues (i.e., information provision), and the system simply asks questions one after another about the user's preferences. Another study [11] concluded that asking too many questions must be avoided in an attentive-listening dialogue. Information must be collected about the user through dialogues while encouraging user motivation.

This paper solves these issues by efficiently obtaining the user's concept through non-task-oriented dialogues.

3 Dialogue Robot System

In this study, to stimulate user dialogue satisfaction and motivation, we propose a dialogue strategy that shows a robot's comprehension ability and willingness to understand the user's concept of preferences. The concept of preferences consists of the user's subjective similarity between items (similarity model) and his/her preferences for them (preference model). When people try to understand another person's opinion through dialogues while constructing a model of his/her thought (concept) based on the obtained information, they ask questions that are likely to update the model. Therefore, in a scenario that tries to understand user preferences, we first construct a method that updates them and the similarity models from the user information obtained through dialogues. Next with the models, we propose a method that efficiently collects the data of user's subjective similarity among preferences through dialogues.

3.1 Data Representation of Preferences and Similarities

Here we specify a user's subjective similarity between items (similarity model) and his/her preferences for them (preference model). The system needs to handle information about whether the user likes restaurant A in terms of its taste, or whether he/she regards restaurants A and B similar in terms of their menu.

Our preference model is represented by a table. Here the viewpoint of the preferences is expressed as an attribute of each item. Let M be the number of items, and let N be the number of attributes, where the matrix size is $M \times N$ (Fig. 1). The number of items and attributes increases through dialogues. The preference data are stored as 1 for like and 0 for dislike. If no data exist, the cell is left blank. The example in Fig. 1 shows that restaurant A is preferred from the perspective of its menu, but not from the perspective of its atmosphere (i.e., mood).

The similarity model is also expressed in a table. A similarity table is a diagonal matrix of $M \times M$ (Fig. 2). Similarity data are stored as 1 with the similarity between two items and 0 for no similarity. If no data exist, the cell is left blank. Similarity depends on the attribute of the items. N tables were made for each different attribute.

There are two types of data in this study: data obtained from user utterances (observed values) and those estimated from dialogues with users (estimated values). An update operation fills the table with observed data, and a comple-

Fig. 2 Similarity model **Fig. 1** Preference model

ment operation fills it with estimated data. The concept of preferences consists of these preference and similarity tables.

3.2 Table Complementation Rule

This section explains the rules of complement operation. Assume that the system has a user's preference model and its own model (the system's preference table is predefined). Since people refer to their own model when estimating their partner's model (i.e., preference/similarity), this system also has its own model. There are four possible complements: from preference table to preference table, from preference table to similarity table, from similarity table to preference table, and from similarity table to similarity table. If the estimated values are affirmed by the data confirmation utterances in the dialogue, the values are treated as observed values. If the complemented data are inconsistent with the obtained data within each rule, no complementation occurs, for example, where a cell of the preference table is estimated as "like" through a preference to preference rule, but as "dislike" through a similarity to a preference rule. Each complementation rule is explained below. The numbers with $\langle \rangle$ in the figures below represent the estimated values that are complemented.

3.2.1 Complementation of Preference Table from Preference Table

This rule complements a preference table based on the similarity in preferences. Figure 3 shows an example of the rule. If the system and the user partially share identical preferences

Fig. 3 Complementation of preference table from preference table

in a certain attribute, the preference table is complemented so that both share identical preferences. If the coincidence degree of the observed preference data in a certain attribute is greater than or equal to a threshold, and if the number of known items in the preference data is greater than or equal to a threshold, the same value of the preference data is complemented as the estimated values. Let M_p be the number of known items of preference data, and let s_p be the amount of matching preference data in the two tables. The coincidence degree in attribute S_p is calculated by the following formula:

$$S_p = s_p / M_p. \quad (1)$$

In Fig. 3, the system has negative preferences for restaurants A and C and a positive preference for restaurant B about atmosphere, and it knows that the user's preferences of A and B are identical as its own. In this case, the system estimates that the user also has a negative preference for restaurant C due to its atmosphere. This rule represents speculation that if the preferences for some items in a certain attribute match between the system and the user, both will have identical preferences for other items in the attribute.

3.2.2 Complementation of Similarity Table from Preference Table

This rule complements a similarity table based on the similarity in preferences. If the preferences of two items in a certain attribute match, the rule estimates that those items have similar attributes. Figure 4 shows an example of this rule. If the two observed preferences in the attribute match, their complemented similarity data are 1; if they do not match, the value is 0. In Fig. 4, the system knows that the user likes restaurants B and C due to their menus. Then it estimates that the user believes that B and C have similar types of menus. Here, the system assumes that those who have identical preferences have similarity.

3.2.3 Complementation of Preference Table from Similarity Table

This rule complements a preference table based on the data of a similarity table. In the opposite case of Sect. 3.2.2, if two items are similar for an attribute, the rule estimates that they have the same preferences in the attribute. Figure 5 shows an example of this rule. If the observed similarity between two items of an attribute is 1 (similar) and if only one item's preference data are missing, we complement the preference of this item with the same value of the known preference data. If 0 (not similar) and if only one item's preference data are missing, we complement the preference of this item with the opposite value of the known preference data. In Fig. 5, since the system knows the user does not like restaurant A's atmosphere, it estimates he/she likes B's atmosphere since it knows he/she thinks A and B are dissimilar. It also estimates that he/she does not like C's atmosphere since it knows he/she thinks A and C are similar. Here, the system assumes that they have similarly like and dislikes.

Fig. 4 Complementation of similarity table from preference table

Fig. 5 Complementation of preference table from similarity table

Fig. 6 Complementation of similarity table from similarity table

3.2.4 Complementation of Similarity Table from Similarity Table

This rule complements a similarity table based on the data of a similarity table, which is the same operation as in Sect. 3.2.1 on similarity. Figure 6 shows an example of the rule. If the coincidence degree of the observed similarity data in a certain attribute is greater than or equal to a threshold, and if the number of known items in the similarity data is greater than or equal to a threshold, the same value of similarity data is complemented as the estimated values. Let M_s be the number of known items of the similarity data, and let s_s be the amount of matching of the similarity data in the two tables. The coincidence degree in attribute S_s is calculated by the following formula:

$$S_s = s_s / M_s. \quad (2)$$

In Fig. 6, the system believes that the menus of restaurants A and B and B and C are similar. It also knows that the user believes that A and B and B and C have similar types of menus. Then it estimates that he/she has identical similarities

Table 1 Classification and examples of utterances that fill table cells

	Data acquisition utterance	Data confirmation utterance
Preference	“Do you like A’s menu?”	“You like A’s atmosphere, don’t you?”
Similarity	“Does A’s atmosphere resemble B’s?”	“The atmosphere of A resembles B’s, doesn’t it?”

concerning the menus for the other items as well as itself. This rule represents speculation that the similarity for an attribute will match some items if the similarity for it matches other items.

3.3 Utterance Generation Rule

In this section, we describe the implementation of the complementation rules in our dialogue system. First we explain the types of utterances assumed in this system. Next we describe the utterance generation rules and the related data processing.

3.3.1 Utterance Type

Two types of system utterances are utilized to acquire the user’s concept of preferences through dialogues. One acquires the user’s data, and the other confirms whether the complemented data are correct. Examples are shown in Table 1. The system asks about the preference of an item in an attribute or the similarity between two items in an attribute.

3.3.2 Dialogue System

The target of this study is such non-task-oriented dialogues as chats as well as those where the topics and dialogue flow are prepared in advance with databases and scripts. The system has a conversation on a predetermined topic (e.g., dogs) and discusses the preferences/similarities of the topic items while learning about the user’s preferences for the items (e.g., golden retrievers or French bulldogs) related to the topic. Assume the system and the user have their own preferences for the item in each attribute (e.g., appearance and personality). Figure 7 shows the dialogue system architecture. First, the speech recognition unit identifies the user’s utterance. The content recognized in this system is the item name and the answer (agreement/disagreement) to the system’s question or confirmation. Based on the recognition result, the dialogue manager generates the system’s utterances. The parts required for utterance generation are preference and similarity models for the system and the user, a database of items and attributes for each topic, and a comment database. Preference and similarity models are constructed by updating the information obtained through dialogues or by complementing the estimated values. The

Fig. 7 System architecture

topic item and attribute database store multiple item data and attributes for each topic. The comment database stores sentences that describe each item. In this system, the dialogue follows a preset scenario. Data acquisition, data confirmation, and comment utterances are generated based on a script. Comment utterances are inserted when each item first appears. Comments prevent the dialogue from descending into a meaningless flurry of interview-type questions and to retain a form that resembles a non-task-oriented chat.

Figure 8 shows a flowchart for generating utterances as the content of the dialog manager. First, we set one topic (e.g., tourist destinations) that the robot and its user discuss. Within the set topic, the system generates utterances. It first makes an initial utterance, e.g., “Which Y do you like from the viewpoint of X?,” where X is an attribute and Y is a topic. If the user’s initial utterance contains an element in the item database, the preference data are stored in the user’s preference table. If the initial utterance is not found in the item database, the system repeats it for the randomly selected item: e.g., “Do you like Y’s X?,” where X is an attribute and Y is an item. If no comment utterance has been made for the item, a comment utterance is generated. After that, the system checks whether the complementation rules can be applied to the user’s tables. When complementation occurs, the last complemented data is selected to confirm whether its complementation is correct (data confirmation utterance). For example, when the user’s preference for A (e.g., the appearance of a golden retriever) is complemented, it says, “You like A, don’t you?” If the user agrees, the complemented data are adopted. If he/she disagrees, the complemented data

Fig. 8 Flowchart

are rejected, and the obtained data are stored in the user's preference table; if the user disagrees with "Do you like A?," then it stores the data that indicate that he/she does not like A. Data acquisition utterances are performed when complementation rules cannot be applied by selecting one cell without preference/similarity data. The data obtained by the data acquisition utterances are stored in the table; the user likes A or thinks A resembles B. After that, it checks the above complementation rules and performs the same processes.

3.3.3 Dialogue Strategy for Effective Data Acquisition Utterance

It is desirable for our system to efficiently complete the preference and similarity tables through dialogues while asking questions that are likely to contribute to model updating. Therefore, we consider a method that efficiently completes the tables from dialogues with users.

The items whose cells have not been filled with data are candidates for data acquisition utterances. When the system obtains the user's preference for an item or the similarity between two items after the data acquisition utterance, 1 (like or similar) or 0 (unlike or dissimilar) is stored in the cell. Let N^0 be the number of cells that are complemented by the complementation rules after 0 is assumed to be stored

in the target cell, and let N^1 be the number of cells that are complemented by the complementation rules after 1 is assumed to be stored. Expected complementation amount (ECA) E_{ij} for the cell in the i th row of the j th column is calculated by the following formula:

$$E_{ij} = (N^0 + N^1)/2. \quad (3)$$

ECA is the expected amount of complementation after the system asks about the user's preference for the i th item in the j th viewpoint or about the user's opinion about the similarity between the i th and j th items in the k th viewpoint. ECA is calculated for every empty cell in the preference or similarity tables, and the cell with the maximum ECA is adopted as the reference cell for the data acquisition utterance. For example, when the cell of the m th item and the n th viewpoint in the preference table are adopted, the following utterance is generated: "do you like item M from the viewpoint of N?" For another example, with the cell of the m th and n th items in the l th similarity table, the following utterance is generated: "do you think items M and N are similar from the viewpoint of L?" If no complementation rule is performed in any cell due to the data acquisition utterance (i.e., ECA is 0 for every cell), a cell without data is randomly selected and the data acquisition utterance is performed. By using this method, we can

Fig. 9 ERICA

effectively complete the table and acquire the user's concept of preferences.

3.3.4 Robot

This section describes the dialogue robot used in our experiment. We adopted ERICA [5] (Fig. 9), a female-type android with a very human-like appearance. A previous study (Ishiguro 2006) concluded that people anthropomorphize communication targets, and humanoids or androids can have an ideal interface for humans. Therefore, compared with machine-like robots, since an android has a human-like appearance, its users can more easily talk with it. If a robot behaves like a human, we can understand a part of its human-likeness from a constructive point of view. In this sense, we believe that using an android that has a human-like appearance in a dialogue system is meaningful.

Its voice is synthesized by a text-to-speech system of VOICE TEXT ERICA.¹ When ERICA is speaking, it moves its lips, head, and torso in synchrony with its voice's prosodic features. The lips, head, and torso movements are automatically generated from its voice using previously developed systems [10,16]. It randomly blinks its eyes.

4 Experiment

The hypotheses of this study are summarized below.

- The generation of data acquisition utterances based on the expected complementation amount (ECA) conveys to users the system's comprehension ability and willingness to understand.

- The generation of data acquisition utterances based on ECA enhances user satisfaction and motivation for dialogues.

Below we describe our experiment that verified the above hypotheses.

4.1 Design

In this section, we describe the experimental design. To verify the hypotheses, we prepared two conditions: generating data acquisition utterances based on ECA (ECA condition) and generating them randomly by selecting a cell (control condition). In both conditions, the same complementation rules were applied.

We used the Wizard of Oz (WoZ) method [4] for the speech recognition part. We chose the WoZ method, which is widely used in the field of human-robot interaction (HRI), to eliminate the possibility that the misrecognition of user utterances would affect the evaluation of dialogue strategies.

We adopted a within-participant design. The dialogue's topic is determined depending on the participant's initial interest, as described below. Each subject participated in both conditions. We counterbalanced the order of the conditions among the participants to eliminate the possibility that the dialogue evaluation between them depends on the dialogue topic.

4.2 Dialogue Data

We prepared the following six topics and three attributes for the item data: pro soccer players [appearance, personality, play style], dogs [appearance, personality, size], J-POP singers [appearance, voice, performance], Pokemon characters [appearance, personality, strength], fashion brands [price, design, name], and tourist destinations in Japan [history, landscape, atmosphere]. For each topic, we gathered 30 items. For the comment data, we referred to Wikipedia² and created comments in the "I heard XX is YY" format. For the tourist destinations, the following is an example comment on Todaiji Temple: "I heard that Todaiji Temple in Nara is a UNESCO World Heritage Site."

To prepare the preference and similarity tables of the system (robot), just one person created all of the data for one topic. A man did the soccer players, woman A did dogs and J-POP singers, woman B did the Pokemon characters, woman C did the fashion brands, and woman D did tourist destinations in Japan. They did not participate in the dialogue evaluation experiment described later.

The threshold for the degree of coincidence in the complement rules described in Sects. 3.2.1 to 3.2.4 was set to 0.75,

¹ <http://voicetext.jp/>.

² <https://www.wikipedia.org/>

and the threshold for the amount of known preference and similarity data was set to 2. Generally, the number of items exceeds that of the attributes (three in this experiment). If the similarity table size is greater than the preference table size, the similarity's ECA will tend to always be larger than that for the preference, and thus a similarity utterance will be always generated. To balance them, we applied the complementation of the similarity table from the preference table to the first three items in this experiment. We set to ten the number of turns with which the participant interacted with the robot. When the conversation's turn limits are reached, the robot says, "I think today's conversation was productive. Thank you" and ends it.

4.3 Questionnaire

We used three types of questionnaires in our experiment.

4.3.1 Questionnaire 1: Subject Preferences and Similarity for Items

The first addresses the preferences and similarities of the participants. We collected them from the participants themselves and compared them with the data estimated by the system through dialogues to evaluate the effectiveness of the complementation rules.

4.3.2 Questionnaire 2: Impressions

The second questionnaire evaluated their impressions of the android. Its items included the android's comprehension ability, its willingness to understand, the suitability of its comments, satisfaction with its dialogues, and motivation to talk with it. We presented the following sentences to the subjects:

- Comprehension ability: "This android understood you."
- Willingness to understand: "This android was trying to understand you."
- Dialogue satisfaction: "I was satisfied with my conversation with this android."
- Dialogue motivation: "I want to talk to this android again."
- Comment suitability: "This android's comments were helpful."

The comprehension ability evaluates the effectiveness of the proposed complementation rules, and the willingness to understand investigates the contribution to dialogue satisfaction and motivation. In this dialogue system, comment utterances were inserted when each item first appeared to avoid a flurry of questions. The suitability of the comments investigated the effect of the comments between conditions.

4.3.3 Questionnaire 3: Dialogue Motivation for Each Topic

The third questionnaire provides a more detailed evaluation of the motivation to talk with the android. The evaluation of the dialogue motivation might depend on the participant's interest in the topic. To study the relation between topic types and participant motivation, we investigated the participant's dialogue motivation with the android on various topics. The topics were created based on previous research [31] that sorted Japanese topics into 16 categories shown in appendix A. The participant's interest in each topic affects the dialogue motivation. To eliminate the effect of individual-dependent interest, the participants answered the same questionnaire before and after the first condition dialogue, and we compared the amount of change in the dialogue motivation for each topic. All question items were evaluated on a seven-point Likert scale: 1: strongly disagree, 7: strongly agree.

4.4 Procedure

Figure 10 shows a flowchart of our experimental procedure. First, the experimenter led the participant to the dialogue room where the android is located and introduced it. At this time, the android and the participant did not talk. The participants were told that they are going to have a conversation with the android. Next they move to the questionnaire answer room to evaluate their motivation to talk for each topic in advance using Questionnaire 3. The experimenter introduces the android to the participants in advance to provide a clear image of it before they answer the questionnaire.

The participants rates the above six topics that assess the level of their interest in them. Next they extract the items that they know from the 30 items that are related to the topic with the highest interest. If the number of known items exceeds ten, the topic is used in the conversation with the android. Otherwise, we follow the same procedure for the topic of the next highest interest. This procedure is repeated until the topic used in the experiment is determined. We excluded participants whose topic of greatest interest was not adopted in the later analysis to avoid a floor effect in the evaluation of motivation. The participants are unlikely to be motivated to talk with the android (regardless whether the generation of data acquisition utterances was effective) if they have little interest in the topic.

After the topic is determined, they answered Questionnaire 1 about their preferences and similarities for each item, moved to the dialogue room, and engaged in the first condition dialogue with the android. After this dialogue, they moved to the questionnaire answer room and answered Questionnaire 3 again about their motivation to talk about each topic and Questionnaire 2 about their impression evaluations.

After answering, they returned to the dialogue room for a second condition dialogue with the android. At this time,

Fig. 10 Experimental procedure

they were instructed to engage in a dialogue, evaluate it, and assume that the android is different from the first conversation. After the dialogue, they moved to the questionnaire answer room and answered Questionnaire 2 about impression evaluations.

4.5 Result

4.5.1 Subject

Sixteen Japanese people (6 men, 10 women, average age, 20.3 (SD 1.72)) participated in the experiment. None had any history of neurological or psychiatric illness. All provided written informed consent prior to the start of the study, which was approved by the Ethics Committee of Advanced Telecommunication Research Institute International, Japan.

As our first investigation of this study, we chose young people as its participants because they may more easily accept state-of-the-art technology. A previous study (Ogawa et al. 2011) reported that elderly people failed to pay attention to the android's utterances even though they evaluated it highly. Since our study mainly focuses on the generation of android's utterances, we selected young people first.

4.5.2 Conversational Data

The data related to the dialogues are shown below. Concerning topic selection, one participant selected pro soccer players, no one chose dogs, five chose J-POP singers, six chose Pokemon characters, one chose fashion brands, and five chose Japanese tourist spots. The average number of data acquisition utterances was 4.63 (SD 0.927) in the proposed condition and 5.88 (SD 0.857) in the control condition. The average number of random data acquisition utterances in the proposed conditions was 1.38 (standard deviation 0.927) (occurrence rate was 28.1%). Under the proposed conditions,

since the occurrence rate of the random data acquisition utterances did not exceed 50% for any participant, the proposed condition is different from the control condition. The dialogue examples with a participant are shown below. Table 2 is an example of the proposed conditions, and Table 3 is the control condition.

4.5.3 Complementation Data

Table 4 shows the result for the complementation data. We evaluated the amount of data and the correct rate estimated by the system. For the amount of estimated data, we adopted the complementation rate as a scale. It is the ratio of the number of complemented cells in the preference/similarity table to the number of cells in the preference/similarity table (averaged viewpoints mentioned in the dialogue). The correct rate is the ratio of the number of complemented cells that match the participant's preference/similarity (answered before the experiment) to the number of complemented cells in the preference/similarity table. Table 4 shows the averaged data among the participants.

The complementation rate of the proposed condition is higher than that of the control condition. Therefore, the proposed conditions for making the data acquisition utterances with the largest expected complementation amount (ECA) performed more estimations than the control condition that randomly made data acquisition utterances. In addition, the correct rate exceeded 60% overall in both conditions. The correct rates are almost identical in both conditions because the same complementation rules are applied in both conditions.

4.5.4 Impression Evaluation

Next we describe the results of the impression evaluations. We compared the proposed and control conditions for comprehension ability, willingness to understand, dialogue satisfaction, dialogue motivation, and suitability of comments. A Wilcoxon signed-rank-sum test revealed a significant tendency ($p = 0.0708$) in the willingness to understand and a significant difference ($p = 0.0295$) in dialogue satisfaction. We clarified that satisfaction with the proposed condition is higher than the control condition.

We also tested the difference from the score of four points, which is the intermediate value (neither agree nor disagree) of each impression evaluation item. In the comprehension ability, we identified a significant tendency ($p = 0.0927$) in the control condition. In the willingness to understand, we identified a significant difference in both conditions: proposed: $p = 0.000657$ and control: $p = 0.00217$. We also found a significant difference of the dialogue satisfaction in the proposed condition ($p = 0.0371$), a significant difference of the dialogue motivation in both conditions (proposed:

Table 2 Dialogue example in ECA condition

System 1	You're interested in sightseeing spots in Japan, right? What's the atmosphere of your favorite spot in Japan?	System 13	As expected. I heard that Miyajima in Hiroshima is also called Itsukushima
User 1	Himeji Castle in Hyogo.	User 13	Yes
System 2	You like the atmosphere of Himeji Castle in Hyogo, right?	System 14	Well, do you like the history of Todaiji Temple in Nara?
User 2	Yes.	User 14	Yes
System 3	I heard that Himeji Castle in Hyogo is registered in the UNESCO World Heritage List and has been selected as one of Japan's 100 most famous castles.	System 15	You like it, right?
User 3	That's right.	User 15	Yes
System 4	Well, do you like the atmosphere of Fushimi Inari Taisha Shrine in Kyoto?	System 16	I heard Todaiji Temple in Nara is registered as a World Heritage Site by UNESCO
User 4	I like it.	User 16	Yes
System 5	You like it.	System 17	Well, do you like the history of Fushimi Inari Shrine in Kyoto?
User 5	Yes.	User 17	Yes
System 6	I heard that Fushimi Inari Taisha Shrine in Kyoto is the general headquarters of the Inari Shrine, which has about 30,000 members nationwide.	System 18	You like it, right?
User 6	Yes.	User 18	Yes
System 7	Well, do you think the atmosphere of Hiroshima Peace Memorial Museum in Hiroshima resembles that of Himeji Castle in Hyogo?	System 19	Well, do you think the history of Miyajima in Hiroshima is similar to Todaiji Temple in Nara?
User 7	No, not really.	User 19	No, not really
System 8	You don't think they are similar?	System 20	You don't think they are similar?
User 8	No.	User 20	No
System 9	Maybe you don't like the atmosphere of Hiroshima Peace Memorial Museum in Hiroshima?	System 21	Maybe you like the history of Mt. Koya in Wakayama?
User 9	I can't really say.	User 21	I like it
System 10	You don't really have an opinion, right?	System 22	That's what I thought. I heard Mt. Koya in Wakayama is one of the most sacred places in Japanese Buddhism
User 10	Right.	User 22	Yes
System 11	I heard that the Hiroshima Peace Memorial Museum in Hiroshima is also called the Atomic Bomb Museum.	System 23	Well, do you like the scenery of Kinkakuji Temple in Kyoto?
User 11	Yes.	User 23	I like it
System 12	Perhaps you think the atmosphere of Fushimi Inari Shrine in Kyoto resembles Miyajima in Hiroshima?	System 24	I see. I think today's conversation was informative. Thank you for talking with me today
User 12	That's right.	User 24	Thank you very much

$p = 0.00601$, control: $p = 0.0462$), and a significant difference of the suitability of the comments in both conditions: proposed: $p = 0.0281$ and control: $p = 0.0384$. We clarified that the proposed condition has a significantly higher evaluation value than the intermediate value for dialogue satisfaction and dialogue motivation.

Next we conducted a correlation analysis to investigate whether comprehension ability and willingness to understand

are related to dialogue satisfaction and motivation. Table 5 shows the Pearson correlation coefficients. We identified a correlation between dialogue satisfaction and comprehension ability ($p = 0.000464$), dialogue satisfaction and willingness to understand ($p = 0.000141$), and dialogue motivation and willingness to understand ($p = 0.0139$). Dialogue satisfaction was positively correlated with comprehension ability and willingness to understand, but dialogue

Table 3 Dialogue example in control condition

System 1	You are interested in sightseeing spots in Japan, right? What's your favorite spot ?	System 13	You like it, right?
User 1	Miyajima in Hiroshima.	User 13	Yes
System 2	You like the scenery of Miyajima in Hiroshima, right?	System 14	I heard that the actual name of Kinkakuji Temple in Kyoto is Rokuon-ji Temple
User 2	Yes.	User 14	Yes
System 3	I heard that Miyajima in Hiroshima is also called Itsukushima.	System 15	Well, do you think the history of the Shirakawago Gassho-style village in Gifu resembles Himeji Castle in Hyogo?
User 3	That's right.	User 15	I don't think they are very similar
System 4	Well, do you think the atmosphere of Mt. Koya in Wakayama resembles the village of Shirakawago Gassho-style in Gifu?	System 16	You don't think they are similar?
User 4	Yes, I think they are.	User 16	No
System 5	You think they are similar. I heard the Shirakawago Gassho-style village in Gifu is a UNESCO World Heritage Site.	System 17	I heard Himeji Castle in Hyogo is a UNESCO World Heritage List and was selected as one of Japan's 100 most famous castles
User 5	Certainly.	User 17	Yes
System 6	Well, do you think the atmosphere of the Hiroshima Peace Memorial Museum in Hiroshima is similar to the Metropolitan Government Building in Tokyo?	System 18	Well, do you like the scenery of Himeji Castle in Hyogo?
User 6	No, I don't think they are.	User 18	Yes
System 7	You don't think they are similar, right?	System 19	You like it
User 7	Right.	User 19	Yes
System 8	I heard the Tokyo Metropolitan Government Building is also called the New Tokyo Metropolitan Government Building or the Tokyo Metropolitan Government Main Building.	System 20	Perhaps you like the scenery of Fushimi Inari Taisha Shrine in Kyoto?
User 8	Exactly.	User 20	It's not bad
System 9	Hmm, do you think the atmosphere of Miyajima in Hiroshima is similar to the Fushimi Inari Taisha Shrine in Kyoto?	System 21	Like I expected. Well, do you like the atmosphere of Mt. Koya in Wakayama?
User 9	I think they are somewhat similar.	User 21	I like it
System 10	You think they are similar.	System 22	You like it. I heard that Mt. Koya in Wakayama is a sacred place in Japanese Buddhism
User 10	Yes.	User 22	Yes
System 11	I heard that the Fushimi Inari Taisha Shrine in Kyoto is the headquarters of the Inari Shrine, which has about 30,000 nationwide members.	System 23	Don't you like the atmosphere of Fushimi Inari Taisha Shrine in Kyoto?
User 11	Yes.	User 23	Well, I like it
System 12	Well, do you like the atmosphere of Kinkakuji temple in Kyoto?	System 24	I see. I think today's conversation was informative. Thank you for talking with me today
User 12	I like it.	User 24	Thank you very much

motivation was only positively correlated with willingness to understand. In other words, dialogue satisfaction must feel both the robot's understanding and willingness to understand, and dialogue motivation must feel the robot's willingness to understand.

Finally, we analyzed Questionnaire 3 regarding the motivation to talk for each topic. We calculated the amount of change in the dialogue motivation after the first condition dialogue for each topic. The results are shown in Table 7. We calculated the amount of change to eliminate such subjective factors as individual interests. Table 6 shows the

Table 4 Complementation data

	Condition	Complementation rate (%)	Correct rate (%)
Preference	Proposed	56.4	77.7
	Control	33.1	79.8
Similarity	Proposed	41.6	62.1
	Control	19.1	48.9
Overall	Proposed	44.3	69.9
	Control	21.7	65.9

Fig. 11 Impression evaluation results

average score and the standard deviation for each condition. A Wilcoxon rank-sum test revealed a significant difference ($p = 0.0269$) between the conditions.

In this experiment, our participants chose their discussion topic among soccer players, dogs, J-POP singers, Pokemon characters, and fashion brands. Perhaps the topics that they wanted to discuss with the android again depended on the topic that they actually did discuss with it. Based on existing research [31] that categorizes topics in Japanese, we classified the topics and categories as follows: soccer players as culture, dogs as humans/creatures, J-POP singers and Pokemon characters as art/hobby, and fashion brand and domestic tourist destination as culture. Since no participant chose dogs in this experiment, all of the topics were classified as either culture or art/hobby. Here we refer to the participants who discussed culture as the culture group, and those who discussed art/hobby as the art/hobby group. The culture group had seven participants; the art/hobby group had nine. We compared the average motivation changes between conditions. Wilcoxon's rank-sum test revealed a sig-

nificant difference ($p = 0.000385$) in the art/hobby group. In addition, significant differences in the items of art/hobby ($p = 0.0384$) and communication ($p = 0.0481$) were identified in it. Regarding the culture group, significant differences in the items of economy/consumption ($p = 0.0438$) and humans/creatures ($p = 0.0388$) were identified. In the total score of the art/hobby group, the dialogue motivation in the proposed condition increased significantly more after talking with the android compared with the control condition.

5 Discussion

We explored a method for efficiently obtaining the concept of user preferences: likes and dislikes. Our experimental results showed the effectiveness of a dialogue strategy based on the expected complementation amount (ECA). In this sense, our proposed method can create conversations with more engagements with its users. On the other hand, the dialogue robot cannot generate its own preferences for exchanging opinions with our proposed methods. We should develop a function that can do this in the future. In addition, we clarified that the female-type android's willingness to understand is positively correlated with dialogue motivation and satisfaction. Although many studies have focused on improving a system's ability to understand users, few have implemented willingness to understand in dialogue robots. In some cases, improving the system's ability to understand users is difficult due to their individual dependence. A result where willingness to understand is correlated to dialogue satisfaction and motivation suggests another way of increasing a user's motivation to talk. Therefore, our result is promising for designing dialogue robots. Future work will explore other methods that convey to users the system's willingness to understand.

If a machine can behave more like a human, we can understand a part of its human-like ways to comprehend/recognize

Table 5 Correlation coefficient (** : $p < .01$)

	Dialogue satisfaction	Dialogue motivation
Comprehension ability	0.583**	0.254
Willingness to understand	0.623**	0.430**

Table 6 Dialogue motivation
(* : $p < .05$)

	Proposed mean (SD)	Control mean (SD)
Change in motivation for dialogue*	0.237 (0.951)	0.0313 (0.980)

a conversation from a constructive point of view. In this sense, we are convinced that using an android that has a human-like appearance with a dialogue system is meaningful. Our experimental results also suggest that our proposed method can express the android's willingness to understand user preferences. This might indicate that the proposed modeling method can provide a human-like way of understanding conversations. Although such a cognitive scientific point is not the main target of this paper, we will discuss it further in the future.

Since the objective of this experiment was to verify the effectiveness of the ECA method, we set two experimental conditions: with/without ECA. We prepared two additional conditions: a robot that only asks data acquisition questions with ECA, and a robot that only asks data acquisition utterances with ECA and data confirmation utterances. The first is expected to show whether data confirmation utterances are important for the robot's comprehension and willingness to understand. The second is expected to show whether adding informative comments is indeed useful for promoting the user's satisfaction and motivation. A previous study [11] argued that excessive questions should be avoided in an attentive-listening dialogue. Another study [21] reported that additional comment utterances improve the insufficiency of robot questions. Furthermore, another system makes comments in conversations [17]. Although these studies indicate that too many questions should be avoided and informative comment utterances work well in dialogue systems, future work needs to investigate the effect of each component in detail.

In the experimental setting, the estimation of the similarity table from the preference table considered only the first three items. Though the experiment in this paper was conducted under the above setting to balance the data acquisition utterances, the complementation rule can handle not only the first three but all of the items. Moreover, a more principled solution would normalize the total number of estimated values for the similarity tables by the total number of similarity entries and the same for the preference table. Finally, we calculated N^0 and N^1 by those normalized values. Future development of the system can adopt this idea.

The proposed system contains attributes in the preference and similarity tables. When the robot estimates values in the preference table from the observed values in the similarity table, various reasons might justify the user's preferences. For instance, if the user does not like restaurant A's atmosphere and thinks that A and B are dissimilar, then the robot

estimates that he/she likes restaurant B's atmosphere. On the other hand, the user might also dislike restaurant B's atmosphere, although for a different reason why he/she dislikes restaurant A's; A is too loud, whereas B is quiet but its seating is too cramped. While various subordinate attributes like loudness or crampedness might exist in a restaurant's atmosphere, we did not verify whether the attribute settings are suitable or adequate. Future work should include this verification and explore a method for handling attributes with a subordinate structure.

This paper examined a method for acquiring the concept of user preferences. Our proposed method shows that our system understands what the user likes about each viewpoint. Although we assumed that the system's preference and similarity tables were all filled in advance for the conversations in this experiment, the complementation rules can be applied for updating the system's preferences and similarity. In this case, the system can acquire preference concepts not only for known items but also for new items through dialogues. The following dialogue is an example.

System 1: What's your favorite restaurant?
 User 1: Restaurant A. [Restaurant A is a new item.]
 System 2: Is restaurant A's menu similar to restaurant B's?
 User 2: No, not really.
 System 3: So you don't like restaurant B's menu? [Complementation of preference from similarity: acquiring the concept of preference for existing items]
 User 3: I can't really say.
 System 4: I see. If the menu is different from that of restaurant B [Complementation similarity from similarity], then I might like it too! [Complementation of preference from similarity: acquiring the concept of preference for new items]
 User4: Really?
 System 5: Restaurant A's menu resembles restaurant C's, right? [Complementation of preference from preference]

In the utterances of System 4, the system acquires a new preference through complementation rules. A system that acquires new concepts is expected to solve the current communication problem with robots where users become bored too quickly by the interactions [27]. In this example, however, another complementation rule had to be added for such unknown items as copying the user's data for the data of the items that the system didn't know. Future work will identify complementation rules to acquire new concepts.

This study only considered data on preferences and similarities for systems and users. On the other hand, by introducing a third person's data, the system can acquire a

Table 7 Amount of change in dialogue motivation for each topic
(* : $p < .05$, ** : $p < 0.01$)

	Proposed mean (SD)	Control mean (SD)
Art/Hobby group		
Science	0.100 (0.539)	0.500 (0.866)
Humans/Creatures	0.000 (0.894)	− 0.125 (0.599)
Media	0.600 (0.917)	0.125 (1.27)
School/Study	0.400 (1.02)	− 0.625 (0.696)
Economy/Consumption	0.300 (0.458)	− 0.250 (0.661)
Art/Hobby*	0.300 (1.10)	− 0.750 (0.661)
Industry	− 0.200 (0.600)	0.250 (0.433)
Nature	0.200 (0.748)	0.375 (0.484)
Society	− 0.400 (0.663)	0.000 (0.500)
Religion/Festivals	0.600 (1.20)	0.000 (1.00)
Relationship	0.800 (0.980)	0.125 (0.781)
Life	0.400 (0.663)	− 0.250 (0.661)
Politics	0.000 (1.10)	− 0.125 (0.599)
Communication*	0.600 (0.663)	− 0.125 (0.599)
Culture	0.700 (1.00)	0.125 (0.781)
History	0.600 (0.490)	− 0.125 (1.17)
Total**	0.313 (0.910)	− 0.0547 (0.832)
Culture group		
Science	0.250 (0.829)	− 0.100 (0.632)
Humans/Creatures*	0.500 (0.500)	0.000 (0.922)
Media	− 0.250 (0.433)	0.000 (0.748)
School/Study	0.500 (0.500)	0.100 (1.04)
Economy/Consumption*	− 1.00 (0.707)	− 0.100 (0.872)
Art/Hobby	1.00 (1.00)	0.000 (1.45)
Industry	− 0.750 (0.829)	0.100 (1.14)
Nature	0.500 (1.12)	0.000 (0.775)
Society	− 0.500 (0.866)	− 0.100 (1.00)
Religion/Festivals	0.500 (1.50)	− 0.200 (0.917)
Relationship	0.00 (0.707)	0.000 (1.37)
Life	0.500 (0.500)	0.300 (1.27)
Politics	0.500 (0.500)	0.400 (0.632)
Communication	0.250 (0.433)	0.100 (1.11)
Culture	− 0.750 (0.433)	− 0.100 (1.10)
History	0.500 (1.50)	− 0.100 (1.12)
Total	0.0469 (1.02)	0.100 (1.08)

more accurate user's preference concept. In such cases, for example, collaborative filtering [18] can be applied to estimate which user is closer to the target user based on much user information. However, collaborative filtering requires a large amount of information for user profiles [6]. Gaussian process regression [30] can also be applied to complement preference tables from similarity tables, and an infinite relation model [12] can complement similarity tables to similarity tables. Future work will investigate the relationship between the data amount and applicable methods.

In this experiment, one participant created the system's preference and similarity tables. The degree of the system's ability to understand and the correct rate may depend on the created tables. In the future, based on the prepared data, we must investigate how the degree of comprehension ability and the correct rate change and how the degree of dialogue satisfaction and willingness are changed accordingly. The system might also prepare a multiple set of preference and similarity tables from many people and estimate the target user's preferences and similarities based on them. In this case, consistency might be lost in the robot's preference and

similarities, which reduces its agency. Hence, we must identify the balance of the degree of understanding and the robot's agency when we use the data of many people. These ideas are also future work.

Next we discuss the limitations of our experiment. Since its participants were only young people, generalizing the results to all users may be inappropriate. A previous study [14] assumed that experimental results with participants in their twenties can be applied to the evaluation of a dialogue robot for seniors. Their study developed a dialogue robot that its users would like to talk with. Though that desire is consistent with our goal, we should carefully investigate the difference between the elderly and the young. Regarding the analysis of the correlations, the coefficients were not high. The reason might be that dialogue satisfaction and motivation are controlled by many factors, and the willingness or ability to understand is just one of them. Future work needs to clarify and consider these limitations with a larger sample size.

Another limitation reflects the android's gender expression and appearance. First, gender can affect the result of an experiment. A previous study [20] reported that people tend to evaluate an opposite-gender robot as more credible, more trustworthy, and more engaging under a donating setting. To the best of our knowledge, however, in non-task-oriented conversations, no studies have investigated the effect of a robot's gender. Regarding the effect of appearance, perhaps a human-like appearance affects the evaluation for dialogue motivation. A previous study [25] concluded that a person's dialogue motivation is reduced if the android comments on topics on which the person did not expect it to have an opinion. Hence, our dialogue motivation result may change depending on which robot is used for the conversation. Still, however it remains our speculation, so the future study should also investigate the effect of the robot's gender and appearance. From these discussions, our experimental result cannot be generalized for all conversational robots, and future work must investigate the gender and appearance effects using other kinds of robots.

As the third limitation, we chose young people as the participants of our experiment's first investigation. As stated in Sect. 4.5.1, a previous study (Ogawa et al. 2011) concluded that elderly people failed to pay attention to the android's utterances, even though they positively evaluated it. Since our study focuses on the generation of an android's utterances, we adopted young people. On the other hand, if elderly people tend to have good impressions about androids, perhaps our proposed method might work more effectively for them. We plan to investigate the effectiveness of our proposed system with seniors in the future.

Next we discuss the evaluation of the dialogue motivation for each topic. Our experimental result showed that the increase of the dialogue motivation in the proposed method

was significantly higher than the control condition for the art/hobby group; the culture group showed no significant difference. Why did we only observe a significant difference in the art/hobby group? One possible reason is the participant's belief in the android's capability. Uchida et al. [25] reported that a person's dialogue motivation is reduced if the android comments on topics on which the person does not expect it to have an opinion. Since participants might struggle to attribute an opinion to the android on a topic from the culture group, no significant difference was observed in it. In the future, we must verify whether the topics spoken in the experiment can be realistically attributed to an android. In addition, we did not ask our participants whether they wanted to talk about the same topic again with the android in the experiment. We only asked whether they had interest in the topic and were familiar with some of its items. Dialogue motivation must be investigated for the topics that were actually discussed in the conversations.

In this study, we prepared preference/similarity tables for both the system and users. However, we failed to clarify what is minimally necessary to stimulate the participants' dialogue satisfaction and motivation. We did not clarify whether the system's preference and similarity tables are necessary. The relationship has not been clarified among the data representation method and the amounts of complementation, the accuracy rates, and the subjective evaluations. Verification of the data representation methods and their effectiveness is more future work. In this experiment, we prepared such topics as soccer players and J-POP singers. It is also possible to talk about soccer and J-POP. When people discuss their preferences and similarities about a topic of interest, narrowing down the topic to a more specific concept is more natural. In the future, we must investigate what concept level is effective for our proposed dialogue system. In addition, we used an android that closely resembles a woman. It remains unclear whether similar results can be obtained with other communication robots. In the future, we must verify the effects of a combination of robot types and our proposed system.

6 Conclusion

We developed a conversational robot that stimulates user dialogue satisfaction and motivation in non-task-oriented dialogues involving opinion exchanges. This paper explored a method for efficiently obtaining the concept of user preferences: likes and dislikes. To efficiently collect such preference data through dialogues, we proposed a dialogue strategy that creates utterances with the largest expected value of the amount of data complementation. We verified our hypothesis that the generation of data acquisition utterances, based on the expected complementation amount, conveys to users the system's comprehension ability and the willing-

ness to understand and increases their dialogue satisfaction and motivation. We experimentally verified this hypothesis using a female-type android that closely resembles a woman. Our result suggests that the proposed method efficiently obtains user preference models and enhances user dialogue satisfaction. In addition, the degree of satisfaction is positively correlated with both the android's ability to understand and its willingness to understand. On the other hand, the degree of motivation (i.e., the long-term willingness to communicate with the android) is positively correlated only with the android's willingness to understand. Our experimental results not only show the effectiveness of the proposed method but also suggest a design theory for dialogue robots to stimulate dialogue motivation. Since we derived our experimental result from the use of a female-type android, it cannot be generalized for all conversational robots. Future work needs to investigate the effect of gender and appearance using other kinds of robots.

Acknowledgements This work was supported by JST ERATO Grant Number JPMJER1401, Japan and by JSPS KAKENHI Grant Number JP19J13256.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Topic List (Dialogue Motivation)

Science

- Beauty of science
- Complexity of arithmetic and mathematics

Humans/Creatures

- Want of plants
- Difficulty of animals

Media

- Banality of the media
- Baseness of show business

School/Study

- Fun of examinations
- Expensiveness of learning

Economy/Consumption

- Joy of international economy and finance
- Length of job hunting

Art/Hobby

- Stubbornness of handicrafts
- Famousness of art

Industry

- Concern about construction and civil engineering
- Badness of industry

Nature

- Beauty of nature and terrain
- Freshness of the universe

Society

- Dislike of incidents and accidents
- Waste of social security and welfare

Religion/Festivals

- Loneliness of festivals
- Danger of religion

Relationship

- Dislike of manners and customs
- Liveliness of appearance

Life

- Fun of hometown
- Length of parties

Politics

- Joy of politics
- Variation of laws

Communication

- Fear of communication
- Smartness of computer

Culture

- Fear of sports
- Convenience of clothes

History

- Fear of history
- Waste of history

References

1. Bessho F, Harada T, Kuniyoshi Y (2012) Dialog system using real-time crowdsourcing and twitter large-scale corpus. In: Proceedings of the 13th annual meeting of the special interest group on discourse and dialogue, association for computational linguistics, pp 227–231
2. Bickmore TW, Picard RW (2005) Establishing and maintaining long-term human–computer relationships. *ACM Trans Comput Hum Interact (TOCHI)* 12(2):293–327
3. Dominey PF, Paléologue V, Pandey AK, Ventre-Dominey J (2017) Improving quality of life with a narrative companion. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN), IEEE, pp 127–134

4. Fraser NM, Gilbert GN (1991) Simulating speech systems. *Comput Speech Lang* 5(1):81–99
5. Glas DF, Minato T, Ishi CT, Kawahara T, Ishiguro H (2016) Erica: The erato intelligent conversational android. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN), IEEE, pp 22–29
6. Good N, Schafer JB, Konstan JA, Borchers A, Sarwar B, Herlocker J, Riedl J et al (1999) Combining collaborative filtering with personal agents for better recommendations. *AAAI/IAAI* 439
7. Higashinaka R, Funakoshi K, Yuka K, Inaba M (2016a) The dialogue breakdown detection challenge: task description, datasets, and evaluation metrics. In: International conference on language resources and evaluation
8. Higashinaka R, Kobayashi N, Hirano T, Miyazaki C, Meguro T, Makino T, Matsuo Y (2016b) Syntactic filtering and content-based retrieval of twitter sentences for the generation of system utterances in dialogue systems. In: *Situated dialog in speech-based human-computer interaction*, Springer, New York, pp 15–26
9. Inoue K, Lala D, Takanashi K, Kawahara T (2018) Latent character model for engagement recognition based on multimodal behaviors. In: *International workshop spoken dialogue systems*
10. Ishi CT, Liu C, Ishiguro H, Hagita N (2012) Evaluation of formant-based lip motion generation in tele-operated humanoid robots. In: *Proceeding of the IEEE/RSJ international conference on intelligent robots and systems*, pp 2377–2382
11. Ishida M, Inoue K, Shizuka N, Katsuya T, Tatsuya K (2017) Generation of various listener responses including self-talking in an attentive listening dialogue system, vol 2017, pp 239–240 (in Japanese)
12. Kemp C, Tenenbaum JB, Griffiths TL, Yamada T, Ueda N (2006) Learning systems of concepts with an infinite relational model. In: *AAAI*, vol 3, p 5
13. Kobayashi S, Hagiwara M (2016) Non-task-oriented dialogue system considering user's preference and human relations. *Trans Jpn Soc Artif Intell AI* 31(1):DSF-A_1 (in Japanese)
14. Minami H, Kawanami H, Kanbara M, Hagita N (2015) Machine response for continuing dialogue with elderly people and dialogue robots using social media. *IPSJ SIG technical reports Entertainment Computing (EC) 2015-EC-36(1):1–8* (in Japanese)
15. Sabelli AM, Kanda T, Hagita N (2011) A conversational robot in an elderly care center: an ethnographic study. In: 2011 6th ACM/IEEE international conference on human-robot interaction (HRI), IEEE, pp 37–44
16. Sakai K, Minato T, Ishi CT, Ishiguro H (2016) Speech driven trunk motion generating system based on physical constraint. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN), IEEE, pp 232–239
17. Sakai K, Nakamura Y, Yoshikawa Y, Kano S, Ishiguro H (2018) Dialogal robot that estimates user's preferences by using subjective similarity. In: *IROS 2018 workshop Fr-WS7 autonomous dialogue technologies in symbiotic human-robot interaction*
18. Shardanand U, Maes P (1995) Social information filtering: algorithms for automating “word of mouth”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press/Addison-Wesley Publishing Co, pp 210–217
19. Shibata M, Nishiguchi T, Tomiura Y (2009) Dialog system for open-ended conversation using web documents. *Informatica* 33(3):277–284
20. Siegel M, Breazeal C, Norton MI (2009) Persuasive robotics: the influence of robot gender on human behavior. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, IEEE, pp 2563–2568
21. Sugiyama H, Meguro T, Yoshikawa Y, Yamato J (2017) Avoiding dialogue breakdown through the coordination of multiple robots. In: *The 31st annual conference of the Japanese society for artificial intelligence*, 2017, The Japanese Society for artificial intelligence, pp 1B2OS25b2–1B2OS25b2
22. Sumi K, Sumi Y, Mase K, Nakasuka SI, Hori K (2000) Information presentation by inferring user's interests based on individual conceptual spaces. *Syst Comput Jpn* 31(10):41–55
23. Takeuchi S, Cincarek T, Kawanami H, Saruwatari H, Shikano K (2008) Question and answer database optimization using speech recognition results. In: *Ninth annual conference of the international speech communication association*
24. Tsunomori Y, Higashinaka R, Takahashi T, Inaba M (2018) Evaluating dialogue breakdown detection in chat-oriented dialogue systems. In: *SEMDIAL*
25. Uchida T, Minato T, Ishiguro H (2019) The relationship between dialogue motivation and attribution of subjective opinions to conversational androids. *Trans Jpn Soc Artif Intell AI* 34(1):B-I62_1 (in Japanese)
26. Uchida T, Ishiguro H, Dominey PF (2020) Improving quality of life with a narrative robot companion: Ii—creating group cohesion via shared narrative experience. In: 2020 29th IEEE international symposium on robot and human interactive communication (RO-MAN), IEEE
27. Uchida Y, Araki K (2007) A system for acquisition of noun concepts from utterances for images using the label acquisition rules. In: *Australasian joint conference on artificial intelligence*. Springer, Berlin, pp 798–802
28. Wallace RS (2004) The anatomy of ALICE. *ALICE Artificial Intelligence Foundation, Inc*
29. Weizenbaum J (1966) ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM* 9(1):36–45
30. Williams CK, Rasmussen CE (2006) *Gaussian processes for machine learning*, vol 2(3). MIT Press, Cambridge, p 4
31. Yamauchi H, Hashimoto N, Kaneniwa K, Tajiri Y (2013) *Practical standards for Japanese language education* (in Japanese). Hituzi Shobo, Tokyo
32. Yu Z, Ramanarayanan V, Lange PL, Suendermann-Oeft D (2017) An open-source dialog system with real-time engagement tracking for job interview training applications. In: *IWSDS*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Takahisa Uchida received the B.E. degrees in Engineering Science, from Osaka University in 2016. And he received the M.E. degrees in Graduate School of Engineering Science, Dep. of Systems Innovation, from Osaka University in 2018. He is now a Ph.D. student of Osaka University. He is also an internship student of the Hiroshi Ishiguro Laboratories at Advanced Telecommunications Research Institute International (ATR). His research interest includes dialogue robot and human-robot interaction. He received the School of Engineering Science Award by the School of Engineering Science, Osaka University in 2016. In 2018, he received the Graduate School of Engineering Science Award by the Graduate School of Engineering Science, Osaka University.

Takashi Minato obtained a Ph.D. degree in Engineering from Osaka University in 2004. In December 2001, he was a researcher at CREST, JST and worked as an assistant professor of the Department of Adaptive Machine Systems, Osaka University from September 2002 until 2006. In June 2006, he began to work as a researcher for ERATO, JST. In January 2011, he was a researcher in the ATR Hiroshi Ishiguro

Laboratories. Since May 2020, he has been working as a team leader in Robotics Project, RIKEN Baton Zone Program. He is also a visiting researcher of the ATR. His research interests include human-robot interaction and android science.

Yutaka Nakamura is a team leader at Robotics Project, Baton Zone Program, Institute of Physical and Chemical Research (RIKEN), where he leads the Behavior Learning Research team. He is also a Guest Professor at Osaka University. He received D. Eng. from Nara Institute of Science and Technology in 2004. After he worked as a postdoctoral researcher at Nara Institute of Science and Technology for two years, he worked at Osaka University until 2020. His research interest includes intelligent robotics and machine learning and he currently focuses on studies for modeling human behavior to realize symbiotic interaction between humans and robots.

Yuichiro Yoshikawa received the Ph.D. degree in engineering from Osaka University, Japan, in 2005. From 2005, he was a Researcher at Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International. From 2006, he

has been a Researcher at Asada Synergistic Intelligence Project, ERATO, Japan Science and Technology Agency. From 2010, He has been an Associate Professor in the Graduate School of Engineering Science, Osaka University. He is a member of Japanese Society of Robotics, Japanese Society of Cognitive Science, the Virtual Reality Society of Japan, Japanese Society for Child and Adolescent Psychiatry, and Japanese Society of Pediatric Psychiatry and Neurology.

Hiroshi Ishiguro received a D. Eng. in systems engineering from the Osaka University, Japan in 1991. He is currently Professor of Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University (2009-) and Distinguished Professor of Osaka University (2017-). He is also visiting Director (2014-) (group leader: 2002-2013) of Hiroshi Ishiguro Laboratories at the Advanced Telecommunications Research Institute and an ATR fellow. His research interests include sensor networks, interactive robotics, and android science. He received the Osaka Cultural Award in 2011. In 2015, he received the Prize for Science and Technology (Research Category) by the Minister of Education, Culture, Sports, Science and Technology (MEXT).