
Article

A Dialogue System that Models User's Opinion Based on Information Content

Yoshiki Ohira ^{1,2*}, Takahisa Uchida ^{1,2}, Takashi Minato ^{2,3} and Hiroshi Ishiguro ^{1,2}

¹ Graduate School of Engineering Science, Osaka University; ohira.yoshiki@irl.sys.es.osaka-u.ac.jp (Y.O.); takahisa.uchida@irl.sys.es.osaka-u.ac.jp (T.U.); ishiguro@irl.sys.es.osaka-u.ac.jp (H.I.)

² ATR; minato@atr.jp (T.M.)

³ RIKEN; takashi.minato@riken.jp (T.M.)

* Correspondence: ohira.yoshiki@irl.sys.es.osaka-u.ac.jp; Tel.: 81+6-6850-6360

Abstract: When designing rule-based dialogue systems, the need for elaborate design by the designer is a challenge for dialogue systems. One way to reduce the cost of creating content is to generate utterances from data collected in an objective and reproducible manner. This study focuses on rule-based dialogue systems using survey data, and more specifically on opinion dialogue where the system models the user. In opinion dialogue, there has not been much study of topic transition methods for modeling users while maintaining their motivation for dialogue. To model them, we adopted information content. Our contribution includes the design of a rule-based dialogue system that does not require elaborate design. We also reported an appropriate topic transition method based on information content. This is confirmed by the influence of the user's personality characteristics. The content of questions can give the user a sense of the system's intention to understand them. We also reported the possibility that the system's rational intention contributes to the user's motivation for dialogue with the system.

Keywords: opinion model; user modeling; information content; dialogue strategy; dialogue system

1. Introduction

Recently, systems for having dialogue with people have been widely studied. Dialogue systems can be divided into two categories: non-task oriented and task oriented. Non-task oriented dialogues contribute to continuing conversations with users [1-2] and to building human social relationships [3]. There are two main methods for developing non-task oriented dialogue systems: a machine learning approach [4-6] and a rule-based one [7-8]. The former approach generally requires large corpus data. When using existing corpora, however, it is difficult to control the utterances of the dialogue system. There is difficulty in context management which makes it unsuitable for long dialogues. On the other hand, this method is not dependent on the designer's skill or subjectivity, so it is highly reproducible and can handle many contents. Compared to this, rule-based methods have advantages in terms of control of utterance and context management and can generate high-quality utterances on a limited number of topics. However, it is necessary for designers to consider every response, and the cost of generating utterances is high. In rule-based dialogue, it is desirable to have a method that can reduce the cost of generating utterances and can produce them as intended by the designer.

Understanding (modeling) the user is also important for dialogue systems. Uchida et al. [9] have reported that a female android's willingness to understand their user correlates with their satisfaction with the dialogue and their dialogue motivation with it. Another system collects the information necessary for product recommendation through dialogue [10]. Therefore, it is desirable for dialogue systems to model users to increase their dialogue motivation, or to improve the accuracy of estimation. Additionally, there is a benefit

in reducing the cost of implementing a dialogue system by clarifying how to realize a dialogue system that understands the user based on the collected data.

In this study, we focus on non-task oriented dialogue among dialogue systems, and among them, opinion dialogue. Opinion dialogues are dialogues in which subjective opinions are exchanged, and it is said that exchanging subjective opinions is a kind of self-disclosure [11]. Therefore, it reflects the user's personal values, etc. There are several possible purposes for people to engage in opinion dialogue. One is to obtain information that they do not know or that they are interested in; Bohm et. al [12] describes dialogue as a way to collectively observe hidden values, intentions, and cultural differences. In other words, it can be interpreted that users are motivated to engage in dialogue when they become aware of values they were unaware of through dialogue. The desire to have one's opinion known (self-expression) is one of the generally recognized motivations for dialogue [13], and the function of self-expression is listed as one of the functions that activate interaction. Hiraki et al. [14] also described "assertiveness" : findings related to the desire for people to share the same opinions and to agree with one another. In order to realize this in a dialogue system, users can ask questions to the dialogue system to obtain information that they do not know or are interested in, or present their opinions. In other words, the user can achieve this through user-initiated dialogue, and the system only needs to respond to the user's questions. However, in terms of self-expression and assertiveness, it may be effective for the system to ask the user questions or express opinions in an attempt to get to know the user. In the case of a first-time dialogue, when there is no mutual understanding between speakers, it is difficult for the user to actively recognize the dialogue system as a partner for assertiveness and self-expression. Therefore, there is merit in using system-driven methods to achieve the dialogue sought by the user.

Until now, dialogue systems have mainly focused on building systems that answer users' questions. In contrast, there have been few approaches aimed at the latter. That is, there were no models for expressing opinions to explore how to collect user opinions to satisfy users. As a result, no investigation of what kind of topic transition strategy would be best had been pursued.

To address this issue, this paper builds an opinion expression model that can elicit users' opinions with various strategies, and clarifies what kind of topic transition strategy will increase user satisfaction. By doing so, we contribute to identifying dialogue strategies that increase users' willingness to interact.

In the dialogue act of a dialogue system, speech in which the dialogue system presents a general opinion or questions the user's opinion in the form of a specific opinion is considered effective. For example, "Pizza is good?" the target and opinion should be explicitly stated by the system. Since different users have different interests, they may not have specific opinions on a given topic or may not be interested in it. It is undesirable to ask such users ambiguous questions expecting that they will spontaneously give specific opinions. Furthermore, if they are repeatedly asked questions about topics in which they are not interested, they may become less willing to interact with the system.

In human-human dialogue, the dialogue changes gradually depending on the relationship. For example, without information about the user, it is not known on which topics they hold many subjective opinions. Therefore, it is conceivable to find the opinions that they hold by asking easy questions for many people to answer, or by learning about their experience first. If the user modeling is advanced, however, presenting topics on which the user is likely to hold opinions will be efficient to collect their subjective opinions. In this study, we examine the topic selection rules for what kinds of questions should be selected to promote the user to feel that the system is "trying to understand the user."

The dialogue system can consider various strategies such as whether it is sufficient to know only one "very rare" opinion of the user, whether the quality (rarity) and quantity (number of opinions) are important, or whether the number of opinions is sufficient. This corresponds to considering whether quality or quantity is more important for the dialogue system to understand the user. Through an experiment with this dialogue system and impression evaluation, we examine what kind of estimation methods make the user's

impression of the dialogue better to understand them. From this knowledge, we discuss what makes people feel that they understand others. Since these are considered related to the user's personality, we also examine the relationship between the personality and the system's willingness to estimate.

2. Existing Research

Dialogue systems that model the user are a different approach from question-and-answer agents such as ELIZA [7] and machine learning-based dialogue systems such as LUNA [15]. This system is expected to be used in dialogue systems and dialogue robots to build longer-term relationships. Collaborative filtering [16] is a method to model the user. It can estimate their opinions from the data of known users based on the similarity between them and the co-occurrence of opinions. It has been reported that this method is not fully effective with a small number of data of the users [17]. To solve this problem, methods such as grouping objects [18] and utilizing knowledge [19] have been proposed. However, these do not take into account interactive methods that require additional information input from the user to maximize estimation. Uchida et al. [9] have reported a dialogue system that expresses user's opinions as tables and estimates their preferences and similarities of items. The dialogue system asks questions to the user, resulting in an efficient estimation of other user's opinions from a small number of opinion data. However, the estimation handles only likes and dislikes, and cannot handle multiple adjectives. In addition, only one third-party data set is referred to in the estimation process, and estimation is not possible by referring to models of multiple users. The ability to assume multiple third-party models and to allow for a high degree of freedom in expression would allow for more general chat dialog situations. In this paper, we propose a method for modeling user opinions interactively with less data that can be associated with the user, while allowing for multiple third-party models and highly flexible expressions of opinions.

It is shown that the dialogue system's willingness to understand the user dialogue contributes to the dialogue motivation in a certain topic [9]. However, the relationship between estimation methods and the willingness has only been investigated to a limited extent. Therefore, this paper proposes a user's opinion estimation system and investigates its effect on the user's subjective impressions.

3. Proposed Method

3.1. Opinion Model

The knowledge of the dialogue system is called an opinion model. The model holds opinion information defined by nouns, adjectives, and their pairs for each person. The system holds opinion models for multiple people, and each noun and adjective is described on a different semantic network. This has the advantage that it is easier to express the polysemy of a particular noun. For example, consider the case where we describe the relationship among three nouns: "pizza" "pizza served at restaurant A" and "restaurant A." "pizza" is a general noun that is not limited to be served at a restaurant, while "pizza served at restaurant A" is a noun that is limited to the restaurant. Here, both "pizza" and "pizza served at restaurant A" are food, and can be included in the expression "Did you eat X?". In this way, the processing differs depending on the type of relationship with other nodes and is the same regardless of whether the relationship exists or not. The semantic networks can describe the type of relationship (edge type) between nouns and can flexibly increase or decrease the number of other nouns (neighbor nodes) that are related to a particular noun. Therefore, we adopted it as a knowledge model that satisfies the requirements of both language generation and knowledge processing. In the previous study [9], it is difficult to increase the number of noun types. On the other hand, in our method using semantic networks, the number of noun types can be increased simply by creating new nodes and adding them to the existing semantic network.

When estimating other user opinions (estimated data) from a small number of obtained user opinions (input data), which method is superior, referring to the entire data or only certain data? In the case of the topic "impressions of food" which is the subject of this system, there is a relationship between "opinions about cake" and "opinions about pizza." Here, it is important whether the other is useful in estimating opinions on the other topic. However, the opinions are limited to the same topic. In the previous study [9], it is not suggested that this method is effective even in the case of mixed topics, such as "pizza" and "cake" where it is not certain that the tendencies of opinion are the same. Hence, This study adopts semantic networks to the model user's opinion.

3.2. Dialogue System

The purpose of this dialogue system is to model users maintaining their motivation for. Modeling users means estimating opinions of specific users. But the dialogue system is not allowed to ask any questions. Dialogue has rules, such as maintaining context, and any deviation from these rules leads to dialogue breakdown [20]. The information that the dialogue system is looking for to model the user does not necessarily match what questions can be asked in the current situation. Also, the semantic network is needed in the process of modeling the user. While related research has been conducted such as in [21-22], few have been developed for the purpose of being used in dialogues. Since it costs to collect, maintain, and construct such data, it is desirable to be able to construct such data from opinion data that can be collected on a large scale through questionnaires. However, the information obtained in this way is considered to be simple, so it may not be able to guarantee the complexity required for a dialogue system. Therefore, it is necessary to find a method to increase the accuracy of the dialogue system by utilizing data with a simple structure and with as little manual maintenance as possible.

There are two main types of functions required for opinion dialogues: The first is the function to correctly recognize the user's intentions from the user's utterance (intention level), which are developed in [23]. The second is the function to ask questions that will promote the user's modeling based on their opinions collected through the dialogue and the knowledge of the system (opinion level). The intention level has two additional functions: the first is to determine whether the user has the relevant answer to the topic or question and is suitable to ask it (interest or knowledge). The second function is to correctly recognize what the user said and what the user intended in response to the user's answer, and to recover if the user's answer is not recognized.

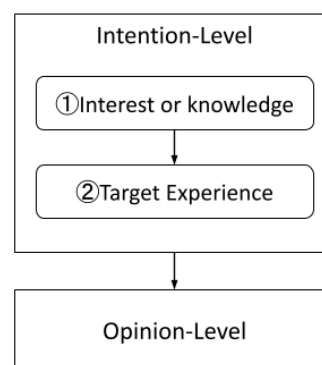


Figure 1. Process flow.

In this paper, we implemented the dialogue system shown in Figure 2.

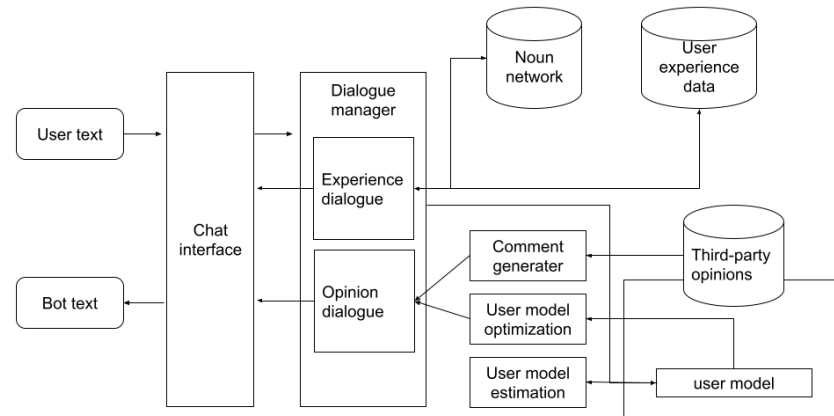


Figure 2. Dialogue System Architecture.

The intention level dialogue corresponds to the "Experience dialogue". The second is "Opinion dialogue" which corresponds to the opinion level. This is the phase in which the dialogue system asks the user for their opinion on a particular noun and obtains the user's response. Since the number of opinions that a user has for a particular noun is limited, it is necessary to change the topic to collect more opinions. Therefore, the dialogue system repeats "Experience dialogue" and "Opinion dialogue". The following is the flow from "Experience dialogue" to "Opinion dialogue". This is referred to as one cycle.

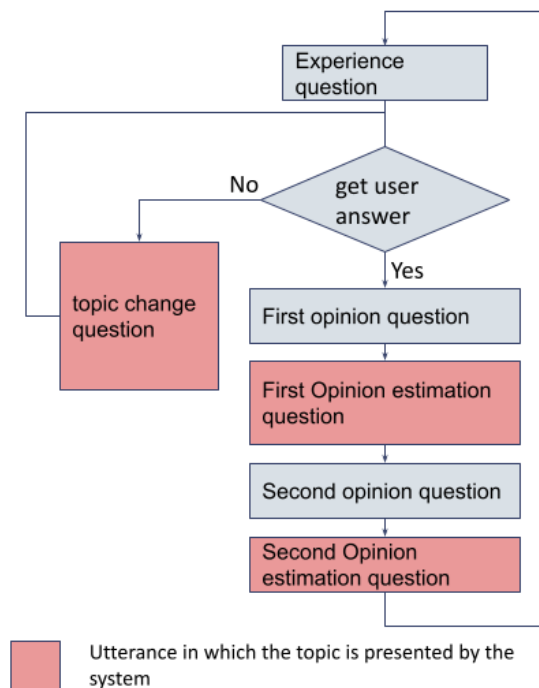


Figure 3. Flow of Dialogue.

The dialogue system asks an experience question, and if it is an opinion that the system has knowledge of, it asks an opinion question. The opinion question involves question generation and opinion estimation, and is repeated twice in this system. The reason for repeating this process twice is that if three or more questions are asked, the user will feel that the dialogue system is asking questions in a monotonous and one-sided manner, and their motivation for dialogue will decrease. We thought that the user would be able to agree that the questions were based on the results of the estimation made by the

dialogue system in response to the first question if they were asked up to twice. If the user's intention was unclear or if their intention did not exist in the system's knowledge, an error recovery was performed.

In this study, opinions are defined as a combination of nouns and adjectives. In other words, the dialogue system aims to identify nouns and adjectives in a single cycle, each of which is conducted in "Experience dialogue" and "Opinion dialogue". In order to identify nouns in the "Experience dialogue" shown in Figure 1, it is necessary to specify the candidates by considering the noun ones presented by the user and the context of the dialogue. For this purpose, a "Noun network" a semantic network with nouns as nodes, is necessary. Furthermore, to prevent users from asking questions that they have already asked once, it is necessary to maintain "user experience data" which is a combination of polarity and nouns that express whether the user has experience.

The "Comment generator" generates general opinions from "third-party opinions" which are opinion data. This is used in "Opinion dialogue" to reduce the proportion of questions in the dialogue act. This is because a dialogue consisting only of questions directed from the dialogue system to the user will decrease in the user's motivation for dialogue. "User model optimization" estimates opinions that should be added to the current user model and generates questions. "User model estimation" estimates and extends the user model from the current user model, which is not explicitly stated in the dialogue. It is not directly involved in utterance generation, but indirectly influences it through the user model.

3.3. Data Collection

In this study, we handle opinions about food as the object of the opinion model. This is because eating and drinking are matters deeply related to people's daily lives, and it is highly likely that many people have opinions about them. The opinion model is defined in the form of "[noun] is [adjective]". For example, opinions such as "pizza is good". First, opinion data on food and beverages were collected through a crowdsourced questionnaire survey. The participants were 74 people between the ages of 20 and 60 living in Japan. The questionnaire items are as follows.

Table 1. Opinion data collection item.

Question Type	Items
Whether you have been to a particular restaurant	Marugame Udon, McDonald's, Saizeriya, Yoshinoya, Gyoza no Ousho
Impressions of specific menus	good, bad, salty, greasy, sweet, spicy, expensive, cheap, no particular impression ¹
Whether or not they have had a particular menu item	Have eaten, have never eaten

Eight adjectives were adopted to the impressions of specific menus. Restaurants were selected from restaurants that have a nationwide presence in Japan and are considered well-known throughout the country. Some other restaurants offering similar food and beverages were omitted. This was done to avoid bias in the opinion data collected. Menus were those listed on each restaurant's website as of December 2021. Alcoholic beverages and general soft drinks, which are restricted to those who can drink, were removed, and semi-solid items such as shakes were used as those classified as sweets. Opinion data were collected for an average of 31 items per menu and 367 items per person in 493 types of food and beverage items offered under the name of restaurants.

From the opinion data, a set of specific restaurants (restaurant) and their menus (menu) are extracted. Then, a genre (genre) was assigned to each menu. The criteria for this process are as follows:

1. Nouns included in multiple menu names were used as genres.
2. One genre is adopted for each menu item.
3. If there are multiple menu components, the main one is used.

4. Assign a menu genre that can be inferred from the restaurant where the menu item is served.
5. If there are abbreviations or synonyms, treat them as the same.

In this dialogue, "genre" means "What did you eat? For example (genre)?" (6) The word is used in the context of "What did you eat?". In this dialogue, genre is used in the context of "What did you eat?". For example, toppings and condiments such as cheese, sauce, and miso are additional to menus, and it is not appropriate to consider them as typical items when menus are abstracted. In some cases, (burger) is also considered a genre, but since it is synonymous with (hamburger), it is treated as the same genre. In many cases, the genre is not included in the product name of menus offered at specialty stores. For example, some menus offered at hamburger restaurants are expressed in the form of "X burger" while others are hamburgers but omit the "burger" part. These were adjusted manually. These subcategories of nouns and adjectives such as "menu", "restaurant", "genre" and "objective" which are the components of opinions, are hereinafter referred to as categories. The "noun network" in Figure 1 describes a semantic network with relations between noun categories, the relations between restaurant and menu, and between genre and menu. The genre-menu relation is called isa relation, and the restaurant-menu relation is. The genre-menu relationship is called "isa" and the restaurant-menu relationship is called "has". "User model" and "third-party opinions" are described as a pair of menus and objectives.

3.4. Dialogue to Specify Intention

In this dialogue system, the "intention level" identifies "what have you eaten at restaurants before?". In the opinion level that follows, the dialogue system asks opinion questions such as "Do you think it tastes good?" or "Do you think it is expensive?" about a specific food or drink, and the dialogue system estimates "Then I thought you would say X is tasty too". After the opinion level is completed, the next experience question is asked by determining the topics that can be presented in order to continue the dialogue. At this point, the user does not always speak in a way that is ideal for the system. In such cases, the system identifies what the cause is, and then makes an utterance to return to a dialogue that achieves the system's goal. This is called error recovery. Error recovery in the "presence or absence of interest or knowledge" section is explained in the "determining strategies according to interest or knowledge" chapter, "identifying the intended concept" is explained in the chapter of the same name, and "opinion level" is explained in the "opinion dialogue" section.

3.4.1. Determining strategies based on interest and knowledge

There are cases where users do not retain the answers intended by the system with respect to the topics presented by the dialogue system. For example, if the answer to the question "What did you eat for lunch?" was given in the morning, the user may respond "I did not eat" or "It is not noon yet". Also, the user may not remember the question. In this case, the answer to the question, such as "I don't remember" may clearly indicate that the answer to the question cannot be provided at this time. For example, in response to a question such as "Have you ever been to (a restaurant)?" the respondent may explain the reason by saying, "I went there a long time ago" indirectly indicating that they do not remember or that they do not intend to continue the topic. In such cases, if a dialogue strategy that elaborates on the topic is chosen in such cases, the user may be forced to recall memories that he or she does not remember, or the user's motivation for dialogue may be dampened by the use of unmotivating topics. In these cases, it is necessary to check the user's interest and knowledge, and if necessary, to avoid the topic. This requires rules that depend on the state of the dialogue and the type of questions asked by the dialogue system. The rules used in this dialogue system are described below.

Table 2. Error Recovery - Type.

Rule Name	Immediately preceding question - type	Example of user's utterance	System Response
forget	any	"I forgot" "I don't remember"	Ask close questions with one level of elaboration. (1)
not interest	open topic	"I'm not interested"	"Are you interested in a meal?" and if yes, end the dialogue.

open topic is an utterance that resets the current context and presents a new topic. (1) is an act of speech that corresponds to an example, and "for example" is attached to the beginning of the sentence. For example, "Have you ever had a cheeseburger at (a restaurant)?" This is a dialogue act that falls under the example. In this case, both the user and the system can be used when the subject of the cheeseburger has not been mentioned immediately before. (The purpose of (1) is to encourage the user to recall the experience. The details are explained in the section "Intention Estimation by Elaboration". The [not interest] indicates that the user is not interested in the topic or does not feel that it is worthwhile to interact with the system, and the preconditions for a conversation to begin with are not met. Therefore, in this case, it is judged that the dialogue must be terminated.

3.4.2. Identification of Intended Concepts

The act of interpreting the user's response in this dialogue system is implemented by mapping candidate answers maintained by the system to the user's intentions (specific concepts that the user wanted to answer to the system's questions). The efficiency of this process leads to an increase in the number of user opinions obtained in relation to the number of dialogue turns, thus better achieving part of the purpose of the dialogue (modeling the user). The following two methods were implemented to identify the intended concepts.

3.4.3. Intention Estimation by Elaboration

In this dialogue system, opinions are defined as a combination of menu and option. Since each menu has its own name, it is ideal that users could directly suggest the name of the menu. However, it is not always possible for the user to provide such a name. For example, in response to the question "Please tell me what else you have eaten at (a restaurant)" some users provide specific examples such as "I have eaten at (restaurant) X" while others provide answers without specificity such as "I have eaten a variety of foods". This is assumed to be because the intent of the question is too vague, there are too many things that come to mind to narrow it down to one, or the memory is too vague to recall immediately. Therefore, in such cases, the system asks, "Have you ever eaten X?" and the system should provide specific examples. The intention is to stimulate the user's memory, and to have the user recall an experience that leads to a concrete opinion. The concrete example presented at this time should be a plausible guess. This item is explained in the section on "plausible assumptions". As described above, open questions are those that prompt the user to say "what" or "where" but do not provide specific examples.

The purpose of close questions is to encourage users to recall their experiences. Therefore, it is desirable that the candidate answers presented by the dialogue system through close questions be concepts that are close to or encompass the user's intentions. Consider the following example.

SYSTEM "What have you eaten at (a restaurant)?" (1)
 USER "Various things." (2)
 SYSTEM "For example, have you had teriyaki corn pizza?" (3)

In utterance (3), the dialogue system presents a specific menu. The user who receives this utterance considers the relevance of the utterance (3) based on the immediately preceding context (1) and (2). In this case, if there is a common understanding that (teriyaki corn pizza) is a typical menu item at (a certain restaurant), the relationship between (1)

and (3) can be inferred. However, if there is no such common understanding, the user cannot fully understand the intention of the utterance (3). In other words, in the above dialogue example, the reason why the dialogue system made such an estimation may be unclear to the user. The cause of this problem is that there is no plausibility that the user thinks the experience presented by the system is what the user is experiencing. In other words, the system asks questions that have no basis for estimation and that are considered to have a low probability of being experienced by the user. Therefore, one solution is to generate questions that have a high probability of being agreed upon, regardless of what intentions the user may have recalled. Specifically.

SYSTEM "What have you eaten (a restaurant)?"

USER "Various."

SYSTEM "Have you ever eaten pizza, for example?"

It is desirable that Pizza is a genre, a concept that encompasses multiple menus in the "noun network" of Figure 1. It is expected that the probability is higher for a "genre" that can be affirmed by having eaten any one of the menus in the "noun network" than the probability that a user retains the experience of eating each individual menu. Therefore, in this dialogue system, the following flowchart was implemented to make the "noun network" correspond to the intention-specific dialogue.

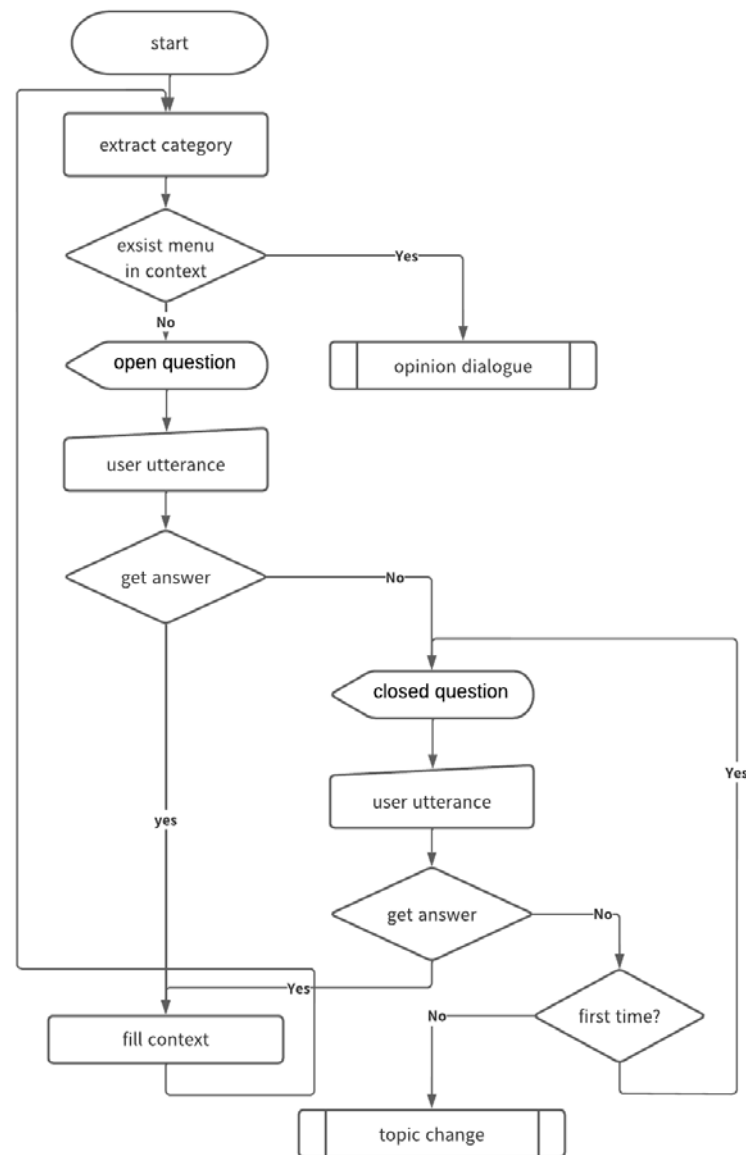


Figure 3. Dialogue system flowchart.

The [extract category] determines the categories for which the context is empty in the order restaurant, genre, menu. The [open question] and [close question] are determined depending on this category. If no answer is obtained after one [open question], [close question] is repeated up to two times. Since [close question] requires estimation of the dialogue system, the accuracy of the estimation depends on whether the purpose of the dialogue can be achieved. On the other hand, in [open question], there is a wide range of user responses, which can lead to errors such as ambiguous responses or failure of the dialogue system to recognize them. Therefore, by employing both types of questions, the probability of recognizing the user's experience is increased. The reason why [open question] is placed before [close question] is that [close question] is a one-way repetition of questions from the dialogue system, which is undesirable from the viewpoint of motivation for dialogue. The reason for repeating [close question] up to two times is also the same: repeating the same dialogue action multiple times may remind the user of mechanical behavior toward the dialogue system.

The concept of context was used above. This is a set of variables that describe the state of the dialogue, in the following format

Table 3. Context used in dialogue to specify intention.

Label	Description	Rule
restaurant	Name of restaurant	
genre	Name of genre of menu	Allow unknown when menu is known
menu	Name of food	
opinion	Opinion on menu	

In Table 3, the levels of abstraction are higher from the top and lower from the bottom. Question generation, such as which genre to select from a particular restaurant and which menu to select from a particular genre-restaurant combination, is explained in the next chapter. It also identifies the restaurant even if the genre is filled first, and identifies the restaurant even if the menu is filled. However, if the menu is filled, we added a rule that the restaurant is not identified if the menu is filled. The reason for this is that the relationship between restaurant and menu is has, while the relationship between genre and menu is *isa*. In the usage of what kind of experience the user retains, in the case of has, the relationship from menu to restaurant is not specified (the dialogue system cannot determine without confirming with the user that the presented genre or menu is offered by the restaurant), whereas in the case of *isa*, the relationship from menu to restaurant is not specified. In the case of *isa*, this is because of the social (not dependent on the user's experience) structure of the menu itself. For example

SYSTEM "What have you eaten at (a restaurant)?"
 USER "I have eaten Margherita pizza."
 SYSTEM "Is Margherita pizza a pizza?"

An example would be the above utterance. Asking this question to the user is an act that emphasizes the lack of common sense of the dialogue system, and may decrease in the user's motivation for dialogue, and is also useless in efficiently identifying the user's intention. For this reason, we have added the rule that if a sub-concept of "*isa*" exists, the blank space in the upper-level concept is acceptable. The dialogue labels to fill in the above context are listed in Table 4.

Table 4. Type of utterance.

Label	Description
open-genre	Ask a question about GENRE in a WHAT.
close-genre	Ask a question about genre in a YesNo.
open-menu	Ask a question about menu in a WHAT.
close-menu	Ask a question about menu in a YesNo.
open-restaurant	Ask a question about restaurant in a WHERE.
close-restaurant	Ask a question about restaurant in a YesNo.
open-topic	While changing the topic, ask the experience question, "What have you eaten before?"
close-opinion	Ask questions about the user's experience in YesNo while presenting adjectives
talk-opinion	Give feedback on the system in response to user responses

For each of the categories listed in Table 3, open and close are implemented, plus open-topic. However, open-opinion was not implemented; two question types, open and close, were provided for each column. This is the minimum structure that was needed to implement the following hypotheses.

1. Priority is given to topics presented by the user rather than topics presented by the system.

Since the purpose of dialogue is the user's experience, it is desirable for the user to present specific experiences to the dialogue system. When the system uses estimation to ask questions to the user, the probability of achieving the dialogue

objective depends on the accuracy of the estimation. Therefore, in order to obtain the user's experience with a small number of dialogue acts, the dialogue system should give priority to utterances that elicit topic suggestions from the user as much as possible. The reason why a small number of dialogues is desirable is that it is burdensome for the user to continue a monotonous dialogue for a long period of time. In other words, user satisfaction can be improved by modeling the user with a smaller number of conversations.

2. Encourage users to recall their experiences.

In some cases, users may not be able to recall their own experiences clearly in response to system questions. In a chat dialog, a highly abstract question such as "What did you do yesterday?" in a chat dialog, a question with a high level of abstraction may not allow the user to recall concrete details immediately. Therefore, if the user does not give a clear answer, it is necessary to provide an example from the dialogue system.

3. Do not repeat the same dialogue act.

If the user does not give a meaningful answer to a particular question to the system, the system should attempt to resolve the issue by other means. For example, if the question "What did you eat?" does not return an intention that the system can interpret, it is undesirable to repeatedly ask "What did you eat?" Repeated mechanical responses will decrease in the user's motivation for dialogue. "What did you eat at that time at X?" The latter is a solution to some extent by changing the expression to "What did you eat at that time?" However, this method cannot solve the problem when the question intention is accurately conveyed to the user and no answer is obtained. In addition, "Did you eat X?" followed by "Then did you eat X?" Then did you eat X? If the user repeatedly says the same thing over and over again, but only the words are changed, the probability of being able to infer the user's intention is low. However, if there are only two or three candidates for estimation, it is acceptable to say, "Then, did you eat X?" in the sense of "If it is not X, then it is only X." is acceptable. It is also possible that (a) "I thought it was X, but it could be X" could be accepted as the intention. However, if the same question type is not answered twice in a row, the same question is not asked a third time. The second failure is acceptable to the user because the user feels that the system is generating a different question than the first one, based on the results of the first failure. However, since the third question is not different from the second one, repeating the failed method may have a negative effect on the user. Therefore, the number of repetitions of close questions in one cycle is limited to two.

While maintaining the opportunity for the user to speak freely, (1) and (2) can be satisfied by giving an example from the system if no answer is obtained. If no answer is obtained after two open and close questions, the question type cannot be repeated due to the condition (3). Therefore, it is necessary to change the topic by using open-topic. Talk-opinion and close-opinion are described in the chapter "Opinion Dialogue".

3.4.4. Intention Estimation Handling Ambiguity

Dialogue systems need to identify user responses from user statements. In [Intention Estimation by Elaboration], we described an intention estimation method based on slot-filling. This is called intention estimation by slot filling. As another method to identify user responses, we also implemented a method that uses analogy from words that indicate the characteristics of branches (menus). This is called intention inference from notation. In [Intention Estimation by Elaboration], specific intentions are identified one by one from ambiguous intentions. However, if the user directly states a word that indicates the item (menu) to be inferred, the menu can be identified without the above means. Intention estimation by slot filling as in [Intention Estimation by Elaboration] has the advantage that the dialogue proceeds under the initiative of the dialogue system, but on the other hand,

the number of dialogue turns until an opinion is obtained is large. On the other hand, intention estimation by notation requires users to spontaneously say the words as intended by the system, but the number of dialogue turns is small. However, not many users have the opportunity to apply the intention estimation by notation. However, if the user provides a clear answer, the dialogue act of slot-filling intention estimation, which reveals the answer little by little, will lead to a decrease in the user's motivation for dialogue. Because the system asks questions about words that the user has already said, the user may distrust the recognition ability of the dialogue system, leading to a sense of frustration and a decrease in the user's motivation for dialogue. Therefore, it is desirable to use the two intention estimation methods together.

Even when the system has narrowed down to one applicable intention for a user's utterance, it may not be appropriate to determine that it is the user's intention. In other words, the user may hold a concept that is unknown to the system. For example, in response to the opinion "I ate pizza" it is undesirable for the dialogue system to proceed on the assumption that the user ate Margherita pizza because "this is the only menu item that includes pizza". If we use the method of [Intention Estimation by Elaboration], we would make statements such as, "Do you think Margherita pizza is good?" However, the user feels that the conversation is interpreted in a way different from their intention, making it difficult to continue the dialogue. The solution is to change the response depending on the level of confidence. This can be done with unknown words, as described below, or with confirmatory dialogue such as, "Are you talking about Margherita Pizza?" and the confirmation dialogue "Are you talking about Margherita Pizza?" If the intention presented by the system in the confirmation dialogue is far from the user's intention, it is inevitable that the user will feel distrust in the recognition ability of the dialogue system because they cannot intuitively understand the reason for the estimation of the dialogue system, and the user's motivation for dialogue will be dampened. However, we believe that the dialogue system itself will be more likely to be accepted by the user by showing an attitude of being unsure of its interpretation. On the other hand, when the estimation accuracy is high, confirmation questions should not be asked. For example, when someone says, "I ate a nugget" it is redundant to ask back, "Did you eat a chicken nugget?" is redundant. In this way, the user may not be able to grasp the intention of the system's question when the notations are almost identical or clearly indicate something. Therefore, it is useful to have a method in which the criterion by which the confirmation utterance should be implemented depends on the ambiguity of the user's utterance, and the level of certainty for each concept should be determined from the user's utterance.

For example, assume that the system has a menu item named "Double Cheese Hamburger" and the user assumes a similar menu item. In such a case, the user may not be able to correctly word the formal menu name word for word. For example, the user may use phrases such as "cheeseburger" or "double cheeseburger". In such cases, the system must evaluate the proximity between the user's utterance and the candidate's answers that the dialogue system keeps in its internal memory. Also, abbreviations may be used, or a higher level concept may be implicitly indicated by presenting it. Therefore, it is necessary to have multiple notations for one concept.

Each of the notations can also be broken down into words, resulting in the composition shown in Figure 5. When a user makes an utterance, it is broken down into morphemes, and whether or not the words in Table 3 are included is determined. KNP [36] was used for morphological analysis. In this example, it is judged that the user's utterance "hamburger store" contains the words "hamburger" and "shop". Therefore, since all the words that make up the notation "hamburger shop" were included in the user utterance, the notation "hamburger shop" was judged to be a 100% match. The percentage of words of this designation included in the user utterance is hereafter referred to as the percentage of agreement.

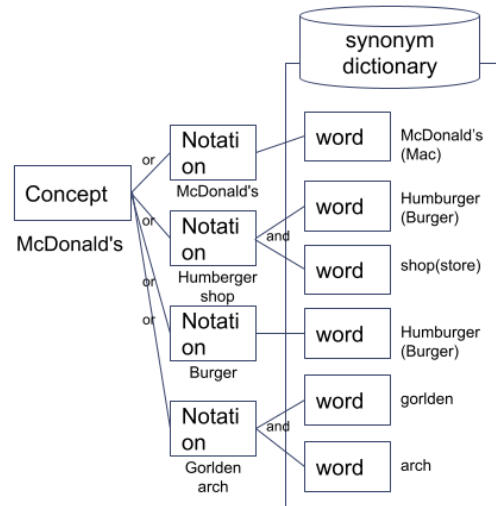


Figure 5. Notation blur of hamburger restaurant.

Notation is acceptable if at least one of the conditions is met. A synonym dictionary is also used at the word level. In this case, "shop" and "eatery" are registered in the dictionary as words that have the same meaning in the context of this dialogue. The above method can be used to determine the percentage of agreement for each of the designations. We considered this percentage as representing the confidence level of the system's interpretation of the user's utterance. In other words, since only one of the maximum matches is meaningful, we calculate the maximum percentage of matches for each invocation. The closer this value is to 1, the more likely the user accurately referred to the name held by the system, and the lower the value, the more likely the user referred to a different concept. In this dialogue system, the rule is that if the maximum match ratio is 1, no confirmation question is asked; otherwise, a confirmation question is asked. The reason for this was to take into account the possibility that a word that was not included in the user's utterance was an important word if the maximum agreement ratio was not 1.

(1) Nominalizing dictionaries and noun dictionaries corresponding to the superordinate categories and categories

Each element of a superordinate category itself may be mentioned in a dialogue. For example, "What kind of [place] is that?" or "The [restaurant] was nice" can be read as referring to a restaurant. When a specific opinion is presented by a user, it is necessary to have a word or words that refer to a superordinate category or category itself in order to identify the object to which it refers. For this reason, we have nouns such as "place" that have the meaning of location, and words such as [genre] and [menu] that indicate the elements of the category itself.

(2) Notation fluctuation dictionary of instances

In order to distinguish them from higher-level categories and categories, specific menu names, genre names, restaurant names, adjectives, etc. that belong to the category are called instances. We used the notation warping dictionary when there are multiple expressions for each instance. When a restaurant uses the expression "ordinary" for a standard menu in a specific genre, the combination of ordinary + menu name, etc. are registered.

(3) Dictionary of word distortion

While [Notation fluctuation dictionary of instances] registers instance-specific word variants, this dictionary registers variants that are not instance-specific. Only those that are judged to have synonyms that will not change while the context of this dialogue is maintained are registered. Specifically, such pairs as katakana, hiragana, kanji, and Japanese-English, and abbreviations such as "hamburger" and "burger" are registered.

3.5. Common Sense Candidate Reasoning

When asking a close question, a specific noun should be presented from the context. Specifically, after the user states that he has been to restaurant A, the system asks, "Have you ever eaten at X?" and generates a question when the user asks the question. For example, consider the following context

```
{ restaurant : (restaurant A) , genre : hamburger , menu : unknown }
```

After this, the context is expressed by keyed brackets. In this case, one of all menus that have a relationship with restaurant and an isa relationship with genre will be presented. For example

```
{ restaurant : (restaurant A) , genre : hamburger , menu : hamburger }
{ restaurant : (restaurant A) , genre : hamburger , menu : cheeseburger }
```

are the candidates. The following two algorithms were considered when selecting one of the menus.

- Priority is given to the one with the highest frequency in the total data.
- Estimating a third party close to the user and estimating the user's opinion from the third party's opinion data

The overall data is represented in Table 1 as a table with columns for people, adjectives, and nouns. Taking the adjectives and nouns as inputs, we can obtain from the overall data a list of Boolean values for the number of third parties stored by the system. At this stage, the context of the dialogue does not include adjectives. In other words, in order to execute the above algorithm, it is necessary to fix the adjectives to one. Here, the adjective is fixed to "good". Frequency in 1 means the percentage of true values in the list of Boolean values for the number of third parties. The second method refers to the method 1 and the second method refers to the method 2. Method 2 differs from method 1 in terms of the third parties referenced: while method 1 refers to all persons, method 2 refers only to the data of a subset of persons who are judged to be close to the user. The details of method 2 are described in the section [Opinion Dialogue].

In the [Intention Estimation by Elaboration], [Inference of Common Sense Candidates] corresponds to close-restaurant, close-menu, and close-genre. These are always performed after open, as shown in Fig. 4, to explore the possibility of advancing the dialogue with fewer dialogue turns by asking open questions. When that fails, the dialogue system presents a plausible candidate based on the current context and the overall data of the dialogue system to encourage the user to recall the experience. The "plausibility" criterion can be thought of in two ways: first, by presenting opinions that are generally accepted, and second, by presenting opinions that are appropriate to the user model. The difference between the two is described in [Selection Algorithms Based on Information Content].

3.6. Opinion Dialogue

After [Identification of Intended Concepts], [Opinion Dialogue] is executed. The purpose of [Identification of Intended Concepts] is to recall the user's experience and identify nouns. In [Opinion Dialogue], the user model is further expanded by identifying adjectives. The following two methods were employed to extend the user model.

1. Directly asking
2. Infer other opinions from the user model

In case 1, for example, "Do you think the cheeseburger (sold at restaurant A) is greasy?" and the user directly returns Yes or No. While this is highly accurate as a user model, it requires N turns of conversation to obtain N user models. 2 is the presumption

that other opinion data not explicitly affirmed by the user is also considered as a user model. This is an efficient modeling method because it allows the user modeling to proceed without any dialogue with the user. On the other hand, accuracy (the percentage of users who agree with the opinions estimated by the dialogue system) is an issue. Furthermore, rules are needed to determine on what basis the system judges the opinions to be plausible. However, this dialogue system is superior to method 1 in terms of modeling because it aims to advance more modeling with fewer dialogues.

Also, if random question generation is used, the user may wonder, "Why did the dialogue system ask these questions?" If random question generation is used, the user may wonder "Why did the dialogue system ask this question?". Continuing dialogue for a long period of time with unclear intentions is considered to be a burden on the user. The final state that the dialogue system wants to know about the user and the process leading up to that state are referred to here as "question intention". For example, if the dialogue system repeatedly asks questions that anyone can agree on, it can be interpreted as "trying to ask questions that are easy to answer for users who are meeting for the first time and have little understanding of each other". For example, if the dialogue system asks a question about "greasy" and the user agrees to the question, it can be interpreted as "the dialogue system judges that the adjective greasy is important for understanding my opinion". The user's motivation for dialogue is likely to change depending on whether or not the user can sense "question intent" from the dialogue system's questions and what type of question intent the user has.

In [Opinion Dialogue], users are modeled in two ways: in method 1, the type of questions generated is important for modeling and motivation for dialogue. In method 2, the questions asked in method 1 are used as input for estimation. In other words, the consideration of the best way to model users using methods 1 and 2 comes down to the rules of question generation in opinion dialogue. In the section [Estimation Model], we will discuss question generation.

3.6.1. Estimated Opinion Model

The purpose of this dialogue system is to model users more efficiently. Therefore, the system needs to estimate the opinions of users and the rules for generating questions to make such estimation more efficient. Here, we consider a method for estimating a larger number of user models from a small amount of data obtained through user interaction. The method employed in this paper is based on third-party data. The dialogue system stores the opinion data of many people (overall data) in advance, and judges "who the user is close to" from a small number of user models. This is based on the assumption that "people who have the same opinions have a high probability of also having similar opinion trends". The following three factors were taken into account when making this estimation.

1. Evaluation of proximity between opinions

One of the studies on classification of dialogue breakdowns [20] is unrelated topics. This means an utterance that deviates from the previous topic, and is one of the causes of dialogue breakdown. This dialogue system repeatedly asks questions to the user multiple times. In such cases, if the immediately following question is not related to the previous question, it may fall under the category of irrelevant topic. To prevent this, the system evaluates the closeness of the opinion obtained by the immediately preceding question to the opinion obtained by the next question. Since opinions are composed of noun-adjective pairs, we have the freedom to refer, for example, to the distance between nouns, or to the distance between adjectives. In this paper, only perfect agreement was used for adjectives and distance was used for nouns. This is because the number of adjective types is small (eight) and cannot be a distance measure that can be classified at a high granularity. However, it is possible to consider the association between adjectives for opinions about food, such as "I find

healthy food tasty" or "I do not like sweet food" depending on the person. This suggests that the association between adjectives may be useful for estimation when considering the association of opinions by different users. Application of this concept could make user modeling even more efficient, but this is a topic for future work.

A method for evaluating proximity between nouns is the semantic network "distance"[24], and its application to word similarity has been suggested with certain contributions in [25] and others. Also, the semantic networks used include wordnet [22] and word2vec [26-27]. However, for the reasons explained in the Data Collection section, it was necessary to use data on real restaurants and the names of the menus they sell. The data was partially created manually based on wordnet.

The above sets the restriction that only opinions whose distance in the semantic network is within a certain value as the proximity between the immediately preceding opinion and the next opinion to be presented are presented. This prevents the appearance of unrelated topics [11]. This threshold is hereafter referred to as the "similarity topic threshold".

2. Evaluation of proximity of opinion tendencies among people

The few opinions obtained in a dialogue are called "input opinions" and the opinions that can be inferred from the data are called "inferred opinions". In [opinion dialogue], it is necessary to implement a method of estimating other opinions from user models. We utilize some third-party data as a method for this, as opposed to the method of Uchida et al. [9], where opinions are estimated by a statistical evaluation that takes all data into account. This paper employs a method that first evaluates the proximity between the user and a specific third party, and then estimates the user's opinion model. This differs from the conventional method in that it does not refer to "the data of others who are not judged to be close. This is a device to make the method use only local data in order to make it applicable even if the opinion expression space or the overall data is huge. Since it is not possible to use a large amount of data, the method evaluates the closeness to a specific third party by utilizing a simple criterion: whether or not there are opinions that coincide with each other. This method is suitable for use in dialogue systems, where the emphasis is not on the accuracy of opinion estimation but on the clarity of the explanatory nature of the estimation.

To determine whether there is a tendency toward the same opinion, the "opinion agreement rate on nearby topics" is employed. Specifically, the number of nouns at a certain distance from the noun of the opinion obtained in the dialogue is used as the denominator, and the number of the same opinion held by both of the two users judging proximity among the nouns in the denominator is used as the numerator. The reason for imposing the restriction of "nouns at a certain distance from the noun of the opinion obtained in the dialogue" in the denominator for calculating the agreement ratio of opinions is based on the hypothesis that nouns that are far apart in semantic distance are less related to opinion trends. Specifically, it is not common to judge that opinions about "pizza" are also similar based on the fact that they are similar with respect to opinions about "cake. The "third party with opinions similar to those of the user" is defined as a person whose "agreement rate of opinions on neighboring topics" exceeds a threshold value. This threshold is called the "similar third party threshold.

3. Evaluation of the likelihood of the estimation

If a third party is sought who is judged to have preferences similar to the user's, the opinion data of that user should be considered as the user's estimated opinion. However, the third party's opinion data contains noise, and not all of the opinion data should be considered as estimated opinion data. In this case, the likelihood of the estimation is determined for each of the third-party opinion data to determine whether or not to consider them as estimated opinion data. Specifically, for each third-party opinion data, we calculate the percentage of third parties who hold that

opinion. Only those opinions that exceed the threshold are considered to be user opinions. This threshold is referred to as the "opinion estimation threshold."

By imposing the constraints of a in the topic presentation of the dialogue system, we can avoid dialogue breakdowns; by b , we can estimate third parties with similar opinion tendencies from the user's statements; and by c , we can extract opinions that are common to all of them. By going through the above process, the user's model can be better estimated with less dialogue. The threshold values used in this experiment are as follows.

Table 5. Opinion dialogue system thresholds.

Type	Threshold
Similar Topic Threshold	2
Similar Third Party Threshold	0.2 (20%)
Opinion Estimation Threshold	0.2 (20%)

Similarity Topic Threshold: The semantic network of this system consists of three layers: restaurant, genre, and menu. Since the target to verify the similarity of nodes is menu, if this threshold is 2, all other menus belonging to the genre to which a particular menu belongs are targeted. Since the opinion consists of menu + adjective, the threshold value must be even. If the threshold is 4, there will be two nodes between the source menu and the target menu. If the number of nodes to be relayed is more than one, the question "What is the relationship of the chosen node to the previous context?" means that the question "What is the relationship of the selected node to the previous context? In order not to make the user feel that it is an irrelevant topic, we decided that the value of 2 is appropriate. The other threshold values were related to the number of third parties and menus maintained by the system, and were adjusted manually, referring to actual dialogues and estimated opinions.

3.6.2. Selection Algorithm Based on Information Content

The previous chapter described the extension of the user model through dialogue. That is, it shows the relationship between the opinions that the dialogue system asks questions about and the user model that may be obtained in the end. Uchida et al. [9] suggested that the dialogue system's attempts to understand the user's opinions can lead to an increase in the user's satisfaction with the dialogue. At this time, the relationship between the opinion that the dialog system "understands the user better" and the level of dialog satisfaction is not clear. For example, hypotheses that contribute to the dialogue system's "attitude of trying to understand the user" include

- Maximizing the number of user opinions.
- That the opinions obtained are "rare opinions"
- Both of them

The following are possible: Knowing many opinions contributes to "user understanding. However, even if the dialogue system is able to know many opinions that are not rare, if there is no knowledge that differentiates the user from the majority, the dialogue system may not be able to say that it has fully understood the user. The method to be employed as the "evaluation index of the final user model" is denoted as "dialogue strategy" below.

If the method presented in [Estimation Model] is an acceptable estimation method for the user, the dialogue strategy is considered to be communicated to the user through the questions presented by the dialogue system. Specifically, if priority is given to "number of opinions" the questions will be those with which many people can agree, or questions about opinions of high interest in which many people hold opinions. On the other hand, if priority is given to "rarity of opinions" questions that subdivide the user's preferences based on the content of the immediately preceding question will be expressed. Thus,

we can expect to manipulate the understanding attitude of the dialogue system by such indicators as the number and frequency of user information. In this paper, we employed the quantity of information as an indicator that reflects the number and frequency of information.

$$I(E) = -\log(E) \quad (1)$$

$P(E)$ is the probability that event E occurs, where E is the event that the user holds a particular opinion. This probability is the probability that a third party holds that opinion. For example, if the dialog system remembers 50 user models and 20 users hold the opinion "Margherita pizza is good" then $P(E(\text{Margherita pizza, good})) = 20/50 = 0.4$. $I(E)$ becomes larger the more rare the opinion is.

In addition, this dialogue system needs to select the best question q from the candidate opinions O that can be obtained through the dialogue. Let Q be the set of candidate questions.

$$O = \text{est}(q), q \in Q \quad (2)$$

$$\text{next nodes} = \{q \in Q \mid \max(f(\{I(o) \mid o \in O\}))\} \quad (3)$$

est is a function that estimates the user model when a question q is selected. In (3), $f(\{I(o) \mid o \in O\})$ is calculated for the set of candidate opinions O that can be obtained through dialogue, and q that maximizes this value is used for question generation. The function f is a function that expresses "what aspects of the opinion candidates that can be estimated by the next opinion presentation are evaluated. The function $f(X)$ used in this study is shown below.

$$\text{count} := |X| \quad (4)$$

$$\text{max} := \max(X) \quad (5)$$

$$\text{sum} := \sum_{x \in X} (x) \quad (6)$$

X is the set of informational quantities, and count is the number of such quantities, not the amount of information. max is the maximum informational quantity in X , and sum is the sum of the informational quantities. We investigate the number of opinions to be estimated and how the impression of the dialogue changes depending on the amount of information.

4. Experiment

In [Selection Algorithm Based on Information Content], we presented three types of indicators to evaluate the final user model: count, max, and sum. count does not include information content in the equation, but the estimated number of pieces affects the result. In other words, the dialogue system does not focus on individual opinions, but rather on the strategy of collecting knowledge about many users. The fact that opinions agreed by many users can be obtained through dialogue means that the model to be estimated tends to have more opinions agreed by many users. max includes the amount of information in the equation, but the number of pieces estimated does not affect the result. In other words, the dialogue system focuses only on the rarity of the opinions and does not consider how many can be estimated. This type of high information content and low number of relevant opinions is a common tendency among uncommon user types. sum includes the amount of information in the equation, and the number of pieces estimated also affects the results. In other words, the dialogue system focuses on both the rarity of the opinions themselves and the number of estimates. Instead of presenting extremely rare or conventional opinions, this method maximizes the acquisition of user models in terms of the amount of

information. The impression of the dialogue is also considered to be influenced by the user's personality [28], so the system is also analyzed with these data.

The main experiment will be conducted after a preliminary survey. In the preliminary survey, third-party opinion data for 40 people were collected online. The subjects are between 20 and 59 years old, residing in Japan, and are collected in the form of questionnaire data to be answered on the Web. The questions consisted of three items: basic information (gender, age, occupation, prefecture of residence, hobbies), user personality characteristics (Japanese version of TIPI [29]), and opinion data (Table 1). This dialogue system was constructed based on the data obtained from the preliminary data collection. The main experiment was then conducted. The conditions for collecting subjects were the same as for the preliminary data collection. In this experiment, 75 subjects responded. Dummy questions were obtained for both the pre-survey and this experiment to screen the data. The following procedure was used for this experiment

1. Conduct a conversation with the dialogue system
2. Respond to the question, "What is your impression of the conversation?"
3. Repeat 1 and 2 three times
4. Respond to a questionnaire for basic information, personality characteristics [27], and opinion data

The conditions are (Formula 1) (Formula 2) (Formula 3), and the order is randomized for each subject. In the "Impressions about the conversation" section, a five-point scale from "does not apply" to "applies well" was set for the following items.

- Did the person try to understand you?
- Did they understand you?
- Would you like to talk to them again?
- Satisfaction with the conversation
- Is the presumption natural or not?

4.1. Interface

チャットボットと会話しよう

こちらのサイトではロボットと少し雑談をしていただきます。
実験のため、ロボットは基本的に質問にはお答え出来ないのでご了承ください。
回答が終わりましたら、こちらのIDをGoogle Formに記入して下さい、ただし2回目以降は記入の必要はありません。
ID: 2gRPE
ブラウザに備え付けのリロードを行うことで対話状況はリセットされますが、対話システム側から指示がない限りリロードを行わないで下さい。
解答は全画面に行ってください、レイアウトが崩れる場合があります。
ロボットの質問に対して行ったことのあるレストランや食べたことのある品名を答える際に、複数の回答を含む発言をしないでください。
発言は一つにまとめて、「発言」ボタンを押してください。
「発言」ボタンを押すためには入力フォームに文字を入力して下さい、空欄の場合は押すことが出来ません。
ロボットの質問に答える時に「○○けど○○ですよね」等の、複数の意味を持つ回答をしないでください。
スペースや特殊記号などを入力しないで下さい。

START

ロボット: 突然失礼します!

ロボット: あなたの食べ物の好みについて知りたいと思っていますので、短い間ですがおつきあい頂けたらと思います

ロボット: 外食だと、どこに行ったことがありますか?

丸亀製麺に行ったことがあります :あなた

ロボット: なるほど

ロボット: 丸亀製麺では何を食べたんですか?

釜玉うどんを食べました :あなた

ロボット: 釜玉うどんって美味しいと思いますか?

(a) image

Let's talk to a chatbot!

You will have a short conversation with the robot on this site.
For experimental purposes, the robot is basically unable to answer questions.
When the conversation is over, please fill out this ID on the Google Form. (However, you do not need to fill in the form for the second time or later)
ID: 2gRPE
Reloading the browser will reset the dialog status, but please do not reload unless instructed by the dialog system.
Please answer in full screen. The layout may be corrupted.
When answering the robot's question with the name of a restaurant you have been to or an menu you have eaten, do not make utterances containing more than one answer.
Please combine all utterances into one and press the "Speech" button.
Please enter text in the input form to push the "Speech" button. (If the form is left blank, the button cannot be pushed).
Do not enter spaces, special symbols, etc.

(b) English translate

Figure 2. Chat Interface

The dialogue interface provides notes at the top of the screen and chat-style dialogue at the bottom of the screen. The dialogue system speaks to the user by entering start. The name of the dialogue system is unified with "Robot". The dialogue system ends when the following conditions are met

- Opinion dialogue is performed a total of 4 times.
- The topic presentation is performed 12 times in total

This dialogue system generates a delay in speech. This is done for the following two reasons: first, to avoid giving the user the feeling that the dialog system is responding mechanically; second, to evoke the user's attention to what the dialog system says. Specifically, this is to prevent the user from responding randomly without reading sentences in detail. It is known that in person-to-person dialogue, turns are exchanged alternately and a certain pause occurs between them [30]. It is also known that differences in robot reaction times can greatly affect the prediction of the robot's response that the user expects [31]. Therefore, while always delaying for a fixed length of time, the dialogue system was delayed significantly for statements that the user particularly wanted attention to be paid to. We divided the delay of this dialogue system into two parts: the delay between the user's utterance and the system's response (turn start delay) and the delay between the system's utterances (intra-turn delay time). The following rules were used for the delay at the start of the turn

- The user speaks 3 seconds after the dialogue system completes its generation of utterances

With the above rule, the delay time varied depending on the time taken to process the generation of utterances. Since the processing of [Selection Algorithm Based on Information Content] takes time in this dialogue system, the delay time increases only for opinion utterances. This may lead the user to predict that the system's opinion utterances are more important than the other utterances.

In addition, the following rules are used for in-turn delay

- Separate by a punctuation mark (end of sentence)
- Speak 3 seconds after the previous utterance

This delay takes into account the time it takes for a person to type a text. If the utterance is repeated without delay, this is a behavior that is difficult for a person to achieve, thus taking into account the possibility of giving the impression of machine-like behavior.

5. Result

The Japanese version of the TIPI [27] was used for the users' personality traits, and five factors [32-33] were assessed: extraversion, cooperativeness, diligence, neuroticism, and openness. We also analyzed those who retained moderate and high scores on each personality trait factor. Since the possible values of the Japanese version of the TIPI ranged from 0 to 16, we divided the TIPI into three groups: 0 to 5.3 as the low group, 5.3 to 10.6 as the middle group, and 10.6 to 16 as the high group. Excluding the low group, the data for the medium and high groups are defined as open-minded, extroverted, cooperative, industrious, and neurotic in this paper. The reason for excluding the low group is to ensure a sufficient number of data and to analyze each case in which the characteristics of each factor are strong. A one-factor analysis of variance was performed using these data. The results are shown in Table 6.

Table 6. Subjective evaluation of each dialogue strategy.

Personality traits	Questionnaire item	condition
--------------------	--------------------	-----------

		max	count	Sum	ANOVA p-value	sub effect tests Ryan-method p-value(pair)
Openness(n=33)	Talk again	2.88	3.00	2.48	0.0251*	0.00952 (count-sum)
	Estimation naturalness	2.97	3.06	2.79		
	Understand	2.85	3.00	2.82		
	Try to understand	3.64	3.82	3.33	0.0356*	0.0109 (count - sum)
	Satisfaction	2.91	2.85	2.39	0.0122*	0.00652 (max - sum) 0.0157 (count - sum)
Extroversion(m=26)	Talk again	2.81	2.92	2.58		
	Estimation naturalness	2.88	3.08	2.96		
	Understand	2.77	2.92	2.85		
	Try to understand	3.58	3.73	3.35		
	Satisfaction	2.73	2.85	2.38	0.0631+	
Agreeableness(n=21)	Talk again					
	Estimation naturalness					
	Understand	2.87	2.89	2.58		
	Try to understand					
	Satisfaction					
Conscientiousness(n=32)	Estimation naturalness	3.08	3.03	2.89		
	Understand	2.89	2.97	2.87		
	Try to understand	3.63	3.76	3.42		
	Satisfaction	2.95	2.79	2.55		
	Talk again	2.78	2.94	2.44	0.0496 *	
Neuroticism(n=22)	Estimation naturalness	3.00	3.06	2.84		
	Understand	2.75	2.97	2.81		
	Try to understand	3.63	3.78	3.34	0.0859 +	
	Satisfaction	2.88	2.75	2.44	0.0756 +	
	Talk again	3.07	3.00	2.66		
	Estimation naturalness	3.21	3.14	2.93		
	Understand	3.03	3.03	2.93		
	Try to understand	3.66	3.79	3.52		
	Satisfaction	3.21	2.93	2.72		

The "ANOVA p-value" does not show p-values below 0.1. Also, "sub effect tests Ryan-method p-value(pair)" does not show p-values below 0.05. + : <0.1, * : <0.05, Talk again: "Would you like to talk to them again?", estimation natural: "Was the estimation natural?", try to understand: "Did they try to understand you?", satisfaction: "Satisfied with the dialogue?".

The results showed that max and count were superior to sum in the "satisfaction with the dialogue" indicator and in the "openness" indicator. In addition, "Would you like to talk again?" and "Was the other party trying to understand you?" also showed higher scores for count than sum in the openness condition.

Table 7 shows the amount of user information at the end of the dialogue for each method. The amount of information of the direct opinion is the amount of information obtained from the user model obtained when the dialog system asked "Do you think X is good/expensive/cheap? The amount of information of the directly asked opinion is the amount of information of the user model obtained when the dialog system asked "Do you think X is good / expensive / cheap? The amount of information of the estimated opinion

is the amount of information of the opinion estimated by the system without confirming it directly with the user.

Table 7. Parameters after dialogue.

	max	count	sum
Sum of information content of directly heard opinions (average)	11.26	9.619	9.741
	(1.413)	(1.380)	(1.337)
Sum of information content of opinions estimated (average)	39.37	51.67	50.33
	(2.141)	(2.253)	(2.218)
Average of the number of opinions directly heard	5.211	4.816	4.895
Average of the number of opinions estimated	19.61	21.87	24.29
Number of types of opinions held by the user model (38 total)	286	279	320
Average number of duplicates of one type of opinion	2.606	2.979	2.884

To investigate the impact of estimation accuracy on the impression of the dialogue, the correlation between estimation accuracy (percentage of correct answers) and impression evaluation was also examined. The results are described from Table 8 to 11.

Table 8. Correlation coefficients between the percentage of correct answers for directly asked items and the questionnaire index.

	Agreeable-ness	Extraversion	Conscientiousness	Neuroticism	Openness
Did the interlocutor (robot) try to understand you?	.0404	.128	.0249	-.0821	.0404
	(.691)	(.264)	(.819)	(.493)	(.691)
Did the robot understand you?	.0172	.0652	-.0314	-.0269	.0172
	(.866)	(.571)	(.773)	(.823)	(.866)
Would you like to talk again?	.0531	.0966	.00909	.0321	.0531
	(.602)	(.400)	(.933)	(.798)	(.602)
Satisfied with the dialogue?	-.0360	.0126	.0974	.0345	.0360
	(.723)	(.913)	(.370)	(.774)	(.723)
Whether the presumption is natural	.0350	.0881	.0201	-.00574	.0350
	(.731)	(.443)	(.853)	(.962)	(.731)

Table 9. Correlation coefficients between the percentage of correct answers for the estimated items and the questionnaire index.

	Agreeable-ness	Extraversion	Conscientiousness	Neuroticism	Openness
Did the interlocutor (robot) try to understand you?	.132	.204+	.168	-.0118	.132
	(.193)	(.0730)	(.120)	(.921)	(.193)
Did the robot understand you?	.204+	.360**	.304**	.143	.258**
	(.0730)	(.00121)	(.00425)	(.229)	(.00992)
Would you like to talk again?	.168	.319**	.265*	.120	.252*
	(.120)	(.00442)	(.0130)	(.315)	(.0118)
Satisfied with the dialogue?	-.0118	.377**	.305**	.179	.280**
	(.921)	(.000675)	(.00404)	(.132)	(.00497)
Whether the presumption is natural	.132	.315**	.250*	.124	.244*
	(.193)	(.00491)	(.0193)	(.298)	(.0150)

Table 10. Correlation coefficients between the information content for directly asked items and the questionnaire index.

	Agreeable- ness	Extraver- sion	Conscien- tiousness	Neuroti- cism	Openness
Did the interlocutor (robot) try to understand you?	.0261 (.797)	.127 (.269)	.0117 (.914)	-.117 (.330)	.0261 (.797)
Did the robot understand you?	.127 (.269)	-.0399 (.729)	-.0937 (.388)	-.125 (.300)	-.0639 (.530)
Would you like to talk again?	.0117 (.914)	.00657 (.954)	-.0534 (.623)	-.0725 (.545)	-.0267 (.793)
Satisfied with the dialogue?	-.117 (.330)	-.0421 (.714)	-.135 (.213)	-.0959 (.423)	-.0880 (.387)
Whether the presumption is natural	.0261 (.797)	-.0170 (.883)	-.0698 (.521)	-.103 (.388)	-.0577 (.571)

Table 11. Correlation coefficients between the information content for the estimated items and the questionnaire index.

	Agreeable- ness	Extraver- sion	Conscien- tiousness	Neuroti- cism	Openness
Did the interlocutor (robot) try to understand you?	.123 (.225)	.181 (.113)	.141 (.191)	.0381 (.751)	.123 (.225)
Did the robot understand you?	.183+ (.0697)	.291** (.00969)	.205+ (.0569)	.101 (.400)	.183+ (.0697)
Would you like to talk again?	.146 (.149)	.180 (.115)	.154 (.155)	.0791 (.509)	.146 (.149)
Satisfied with the dialogue?	.174+ (.0842)	.222+ (.0507)	.199+ (.0649)	.152 (.202)	.174+ (.0842)
Whether the presumption is natural	.113 (.263)	.180 (.116)	.117 (.280)	.0929 (.438)	.113 (.263)

0.05. + : <0.1, * : <0.05.

The "opinions directly questioned" refer to the opinions that the dialogue system asked the user during the course of one dialog. The "estimated opinions" are the opinions of the user model at the end of a dialogue, excluding the opinions that were directly questioned. The percentage of correct answers is the percentage of the opinions that were included in the opinion survey data conducted after the end of the dialogue. The questionnaire index is the value of the subjective survey of the dialogue conducted after the dialogue. The correlation between the estimated percentage of correct answers and the questionnaire index showed weak correlations for cooperativeness, extraversion, industriousness, and openness. Although it could not be confirmed for neuroticism tendency, the number of mothers tended to be smaller for neuroticism tendency than for the other indicators. Furthermore, the correlation between the estimated information content on opinions and the questionnaire index showed weak correlations in extroversion and diligence with the index "Did the other party understand you? The correlations were confirmed only for the estimated opinions, and none of the correlations were confirmed for the opinions directly asked.

Examples of dialogues are presented below.

ROBOT: Excuse me for coming on such short notice! I would like to know about your food preferences, and I hope you will stay with me for a short time. If you eat out, where have you been?
 USER: I have been to [name of family restaurant].
 ROBOT: I see. USER: Yes, I have been to [name of family restaurant]. Have you ever been to [name of Chinese restaurant], for example?
 USER: I have been there.
 ROBOT: I see. What did you eat at [name of Chinese restaurant]?
 USER: I had gyoza.
 ROBOT: I see. What kind of dumplings did you have?
 USER: I had garlic zero ginger dumplings.
 ROBOT: Do you think garlic-zero ginger dumplings are good?
 USER: I think they are good.

ROBOT: Hmm. Then, do you think [name of Chinese restaurant] ramen is good?
USER: I think it is good.
ROBOT: I guess. I have the impression that those people like healthy food. Do you think ramen is inexpensive?
USER: I think it is cheap.
ROBOT: Hmm. Then, do you think [name of sushi]'s asari-dashi ramen is inexpensive?
USER: I think it is cheap.
ROBOT: I guess. I have a vague impression that people like that would say that asari-dashi ramen is cheap. What else have you eaten at [name of Chinese restaurant]?
USER: I have had fried rice.
ROBOT: Do you think fried rice is good?
USER: I think it is good.
ROBOT: I see. Do you think fried rice is cheap?
USER: I think it is cheap.
ROBOT: I see. It's time for us to go, so I'm going to end now. Thank you very much.

※ The parts that can identify actual stores are indicated in generic notation with [].

4. Discussion

In this study, we developed a dialogue system to model users' opinions, and differences in users' motivation for dialogue were examined depending on the methods of the opinions to be modeled. The results showed that sum was inferior in terms of motivation for dialogue under the high openness condition, while count scored higher than sum in terms of understanding attitude and dialogue satisfaction under the same openness condition. In addition, the information content and the correctness rate of the estimated opinions were positively and weakly correlated with some subjective ratings in some conditions.

The conditions in this dialogue are the differences due to the methods of the opinions to be modeled. Chat dialogues are generally not only about exchange information, but also aiming to continuing the dialogue and the dialogue itself. Therefore, the attitude of understanding the dialogue system is not necessarily important for the user. On the other hand, even in chat dialogues, it is thought that some people's motivation for dialogue is greatly influenced by what the other party has said rather than by the social rules of the dialogue. In Table 6, only openness showed significant differences among the three conditions of "whether or not you were satisfied with the dialogue", "whether or not you would like to talk again" and "whether or not the other party was trying to understand you". This suggests that for those with high openness, differences in the nature of the presumed opinion of the user may affect their motivation for dialogue, etc. In [28], which investigated the relationship between personality traits and dialogue, it was suggested that openness and main effects are factors, and the results here are similar.

Table 7 shows that sum and count are almost equal in terms of the total information content, while max is inferior. In terms of the number of opinions, sum is greater than count, and count is greater than max. This suggests that max is not a superior method from the viewpoint of advancing user modeling and that sum is a superior method. On the other hand, Figure x shows that sum was inferior to the other two methods in subjective evaluation in terms of "whether you were satisfied with the dialogue", "whether you would like to talk to the person again" and "whether the person tried to understand you".

There is no significant difference in the information content of the opinion directly mentioned between sum and count, and there is a significant difference in the information content of the opinion inferred. In other words, there is a suggestion that users were able to sense the tendency of opinions that can be inferred from their opinions through the dialogue. However, this does not deny the possibility that there are other hidden factors. The results indicate that there may be a relationship between the difference in the estimated information content and subjective evaluation. To investigate this result, the correlation coefficients between the information content and subjective evaluation are shown in Table 11 [Correlation coefficients between the estimated information content and the questionnaire index]. The results show a significant trend toward a weak correlation

between motivation for dialogue and the estimated information content in the extrovert group. This indicates that those with extroverted personality traits have a better impression of the dialogue system's ability to estimate more user models. On the other hand, there was no significant difference in [the correlation coefficient between the information content directly asked and the questionnaire index]. This indicates that the information content that can be estimated as a result is more important than what is directly presented as the information content that affects the motivation for dialogue. This indicates that users are able to experience the estimates of the dialogue system through the dialogue, and that their motivation for dialogue is influenced by the information content that the estimates contain. The considerations that led to these results are as follows. First, "opinions that directly question" are constrained by the structure of the dialogue. Specifically, when asking a question to a user, the user may feel uncomfortable if the question is suddenly asked with a high information content (unusual). Furthermore, the dialogue environment in this experiment is a dialogue that takes place in a situation similar to a first meeting and ends in a few minutes. In such a dialogue environment, users may negatively perceive opinions with a high information content. Therefore, there are two possible factors that affect the motivation for dialogue in terms of the information content: "whether the information content is appropriate for the dialogue environment" and "the degree to which the dialogue system understands the user" for "opinions to be asked directly". Therefore, it is not a relationship where the greater the information content, the higher the motivation for dialogue. On the other hand, if we interpret that "the information content of the estimated opinion" has nothing to do with the dialogue environment, but only the degree of understanding, it could explain why a positive correlation between the information content and motivation for dialogue was confirmed only for the estimated opinion. Clarifying this factor is a future issue.

The relationship between the percentage of correct answers and motivation for dialogue is discussed. Table 9 [Correlation Coefficients between Correct Answer Rates of Estimated Items and Questionnaire Indicators] confirms a weak positive correlation between correct answer rates and four question items. Only the tendency toward neuroticism was not found to be significantly different, but this may be due to the small population size of the neurotic tendency. Therefore, it is suggested that regardless of personality traits, a high rate of correct answers for presumptive opinions may improve the impression of dialogue. However, there was no association with the rate of correct answers in the posture of understanding. This can be interpreted as the attitude of trying to understand the user is not an evaluation of the possibility of estimation by the dialogue system, but rather of the behavior of trying to estimate. Therefore, it is possible that there was no correlation with the percentage of correct answers, which is one of the indicators of the degree of estimation. On the other hand, no correlation with subjective evaluation was found in the [percentage of correct answers to direct questions]. We believe that this may be due to the tacit understanding between the user and the dialogue system that the dialogue system is trying to understand the user interactively. In other words, the user understood that the dialogue system did not fully understand the user in the early stages of the dialog, and the user was able to tolerate the dialogue system's low-precision questions. On the other hand, the rate of correct answers to the estimated opinions was correlated with the motivation for dialogue. As mentioned above, users can experience the results and trends of the estimation of the dialogue system, which are not explicitly stated, through the dialogue. In other words, we thought that the users were able to feel that the user model estimated by the dialogue system from the previous dialogues was not consistent with their own self-perception. In Table 9 [Correlation Coefficients between Correct Answer Rate of Estimated Items and Questionnaire Indicators], there may be no relationship between the correct answer rate of the estimated opinions and personality characteristics, except for the neurotic tendency. A low correct response rate means that if one of the objectives of the user's dialogue is mutual understanding, the user feels that they could not fully achieve the dialogue objective. This feeling may have a negative effect on such indicators as motivation for dialogue and dialogue satisfaction, regardless of the

personality trait. It is difficult to determine whether the neurotic tendency's different results from other personality traits may be due to the small number of data, or whether there are factors that appear only in the neurotic tendency.

Here, we discuss the differences in estimation methods (max, count, sum). Refer to Figure 6. It can be seen that sum is inferior to the other methods in terms of subjective evaluation. That is, it suggests that the dialogue system's attempt to maximize the information content acquired is not favorably accepted by the user. We believe that this is due to the fact that the intent of the question is unclear. What we want to know through a few turns of dialogue is "Do you think cheeseburgers are good?" but rather an opinion that emerged as a result of the combination of multiple menu items. The user who receives the question considers the intent of the dialogue system's question, but the lack of an intuitive answer to it is likely to have a negative effect. In chat dialogues, dialogue can be divided into units called discourse [34]. Within a discourse, specific topics are explored in depth, and there are scattered repetitions of the dialogue process, with topic changes taking place between the discourse and the discourse. When mapped against this structure, it feels more natural to have a simple and specific question intention within the discourse. It is possible that the question intent referring to the sum of the rarity of multiple opinions was perceived as complicated by the user. sum's method is also contrary to the structure of the dialogue, and may have given the user the impression that only topic transitions were system-driven and no in-depth exploration was done. On the other hand, there is not much difference in the subjective evaluation of max and count. In Uchida et al.[9], it is reported that the count method is optimal. max, which is a method to maximize the information content of a specific opinion, has the same level of evaluation as count in subjective evaluation, and is considered to be one of the most reliable methods as an estimation algorithm in topic presentation. On the other hand, max has the lowest average number of overlapping opinions in Table 7, and sum and count have the same level of overlapping opinions. This means that sum and count methods have less spread of estimation compared to max. Although not introduced in this study, this indicator is important for a dialogue system that updates its knowledge model through dialogue with the user and improves with each successive dialogue. On the other hand, in Table 7 [Number of types of opinions held by user models], max only slightly increases in relation to count. This is due to the low average number of opinions estimated for max. Although the difference between max and count is small for the data used in this experiment, max is expected to be a better method when dealing with sparse data with a large number of opinion types and small data. On the other hand, when dealing with dense data with a small number of opinion types and a large amount of data, count is considered to be superior from the standpoint that the number of estimates per dialogue is larger. On the other hand, the sum method maximized the [number of opinion types held by the user model]. Although this method is inferior in terms of subjective evaluation by the interviewee, it is superior in terms of updating and improving the system with fewer interactions. This is expected to be useful for interfaces with less interactive elements, such as questionnaires.

This experiment is a chat-style dialogue, and there is no interface for the user to infer the personality or characterization of the dialogue system. For the user, the dialogue system is not able to infer the personality of the person. Therefore, it is possible that the user did not feel that they wanted the system to understand him/her. Furthermore, in a first face-to-face dialogue, the user should start by asking innocuous questions that are easy for the user to answer, and questions that require an understanding of the context before and after, or sensitive questions, should be asked only after the relationship between the two parties has been established. This dialogue structure may have contributed to the increase in count's subjective evaluation in this experiment. In other words, other estimation methods may be effective in a long-term dialogue or in a dialogue environment where information on the user is provided in advance, rather than in a first face-to-face dialogue. If the motivation for dialogue is based on the desire to communicate without limiting the recipient, as in the case of the desire for self-disclosure [35], this dialogue experiment can achieve this. If the desire is based on the social act of dialogue itself, the "count" method,

which repeatedly asks easy-to-answer questions, can be used. If the desire is based on the construction of a long-term relationship, such as the "sum" and "max" methods, can be used.

In this experiment, 50 people's opinion data were implemented in the dialogue system prior to the dialogue experiment. This number of people is referred to as the number of third-party data. If there is no bias in the data collection method, the estimation accuracy is expected to increase with the number of persons. Under the conditions of this experiment, a higher number of people is preferable because the higher accuracy of the estimated opinions has a positive effect on the motivation for dialogue. On the other hand, to estimate the tendency of other people's opinions, there is a degree of knowledge about the tendency of third-party opinions that is generally stored as prior knowledge. That is, even when the number of third-party data is not large, it is possible to engage in natural dialogue as people do. If the number of third-party data is small, the user may behave like a person who has little understanding of others and little experience in dialogue, and if the number is large, the user may feel as if the person has repeatedly communicated with others. In such a category of human-like behavior, it is thought that differences in impressions may occur depending on personality and experience. In this dialogue, there was no utterance that told the user how much third-party data the dialogue system holds. Therefore, it is possible that the user implicitly expected the dialogue system to have sufficient estimation capability, and it is thought that the estimation capability correlated with the motivation for dialogue. In other words, there is a possibility that the user will accept even a small amount of third-party data by making utterances that show behavior that is not confident even in the dialogue system's words and actions, and by telling the user that the number of third-party data is small. Impression evaluation in dialogues where dialogue system knowledge is low, the user has more confidence in the utterance, and the imbalance of confidence in the utterance may have a different tendency than in this experiment. The fact that a method for maintaining the user's motivation for dialogues with a small amount of third-party data and low estimation accuracy can be clarified means that the cost of collecting information to improve accuracy when building a dialogue system can be reduced. In other words, it indicates that the cost of creating a dialogue system will be lowered. This is an issue to be addressed in the future.

The topic of food impressions was adopted in this study. We discuss the generalization of the experimental results to other topics. First, the relationship between the estimation method and the amount of the data is discussed above. Food is something that is related to many people every day, so some overlap in preferences and opinions between subjects can be expected. In terms of adjectives, "good" is considered to have a large subjective factor, while "sweet and spicy" can be explained by the amount of sugar and stimulants contained in the food, and thus have relatively few subjective factors. In other words, the data tends to have a wide distribution of opinions that are common to all users and opinions that are specific to each user. On the other hand, if the data has a large subjective factor and it is difficult to find common opinions in general, we can expect that the estimated number of max and sum will be better than in this experiment. In the case of data where there are almost no subjective factors and common opinions are somewhat self-evident, the difference between count and max is expected to be more pronounced.

This dialogue system can be applied to other dialogues by creating a noun list, an adjective list, third-party opinion data, noun distortion data, and noun-noun relationship data. Third-party opinion data, which has the largest number of data, can be collected through questionnaires. Therefore, the dialogue system can be constructed without much effort on the part of the dialogue designer. Since it is possible to accumulate impression data on concepts (e.g., pizza, Tokyo, etc.) through the chatbot, it is expected to be applied to quantitative evaluation of impressions of advertisements and catch phrases. Since the opinion model of an individual interlocutor is estimated, it can be applied to product recommendation. By utilizing the opinion models of individuals, it is also expected to be applied to person-to-person matching and clustering based on similarity of interests and preferences. This classification of people's opinion tendencies is considered useful for

analyzing personality characteristics and classification tendencies limited to specific topics, rather than universal personality characteristics such as the Big Five.

Author Contributions: Conceptualization, Y.O., T.U. and T.M.; methodology, Y.O., T.U. and T.M.; software, Y.O., T.U. and T.M.; validation, Y.O., T.U.; formal analysis, Y.O., T.U.; investigation, Y.O., T.U.; resources, Y.O., T.U., T.M. and H.I.; data curation, Y.O.; writing—original draft preparation, Y.O., T.U.; writing—review and editing, Y.O., T.U.; visualization, Y.O.; supervision, T.U., T.M. and H.I.; project administration, T.U., T.M. and H.I.; funding acquisition, T.U., T.M. and H.I.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JST ERATO Ishiguro Symbiotic Human Robot Interaction Project JPMJER1401, JSPS KAKENHI, grant number 19H05692, JP20K11915, 19H05693 and 22K17949.

Data Availability Statement: Data are included in this paper.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Dominey, P. F.; Paléologue, V.; Pandey, A. K.; Ventre-Dominey, J. Improving quality of life with a narrative companion. In 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) 2017, pp. 127-134.
2. Sabelli, A. M.; Kanda, T.; Hagita, N. A conversational robot in an elderly care center: an ethnographic study. In 2011 6th ACM/IEEE international conference on human-robot interaction (HRI) 2011, pp. 37-44
3. Uchida, T., Ishiguro, H., & Dominey, P. F. Improving Quality of Life with a Narrative Robot Companion: II—Creating Group Cohesion via Shared Narrative Experience. In 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) 2020, pp. 906-913.
4. Bessho, F., Harada, T., & Kuniyoshi, Y. 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 2012, pp. 227-231.
5. Savenkov, D., Weitzner, S., & Agichtein, E. Crowdsourcing for (almost) real-time question answering. In Proceedings of the Workshop on Human-Computer Question Answering 2016, pp. 8-14.
6. Shibata, M., Nishiguchi, T., & Tomiura, Y. Dialog system for open-ended conversation using web documents. *Informatica*, 33(3) 2009
7. Weizenbaum, J. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1) 1966, pp. 36-45.
8. Wallace, R. S. The anatomy of ALICE. In *Parsing the turing test*. 2009, pp. 181-210. Springer, Dordrecht.
9. Uchida, T., Minato, T., Nakamura, Y., Yoshikawa, Y., & Ishiguro, H. 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. *International Journal of Social Robotics*, 13(6) 2021, pp. 1499-1516.
10. Konishi, T., Sano, H., Ohta, K., Ikeda, D., Katagiri, M. Item Recommendation with User Contexts Obtained through Chat Bot. *Multimedia, Distributed, Cooperative, and Mobile Symposium, 2017*, 487-493. (in Japanese)
11. Altman, I., & Taylor, D. A. *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston 1973.
12. Bohm, D., Factor, D., & Garrett, P. *Dialogue: A proposal* 1991.
13. Rogers, C. R. *Carl Rogers on encounter groups*. 1970.
14. Hiraki, N. *assertion training, 1st ed.; Kaneko shobo: Tokyo Bunkyo Ohtsuka 3-3-7, Japan, 1993*
15. Dinarelli, M., Stepanov, E. A., Varges, S., & Riccardi, G. The luna spoken dialogue system: Beyond utterance classification. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing 2010, pp. 5366-5369.
16. Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. Using collaborative filtering to weave an information tapestry. *Communications of the ACM* 1992, 35(12), pp. 61-70.
17. Shardanand, U., & Maes, P. Social information filtering: Algorithms for automating “word of mouth”. In Proceedings of the SIGCHI conference on Human factors in computing systems 1995, pp. 210-217.
18. Ungar, L. H., & Foster, D. P. Clustering methods for collaborative filtering. In *AAAI workshop on recommendation systems 1998*, Vol. 1, pp. 114-129.
19. Burke, R. Integrating knowledge-based and collaborative-filtering recommender systems. In Proceedings of the Workshop on AI and Electronic Commerce 1999, pp. 69-72.
20. Higashinaka, R., Funakoshi, K., Araki, M., Tsukahara, Y., Kobayashi, Y., Mizukami, M. Text Chat Dialogue Corpus Construction and Analysis of Dialogue Breakdown. *Journal of Natural Language Processing* 2016. 23(1), pp. 59-86.
21. Speer, R., Chin, J., & Havasi, C. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-first AAAI conference on artificial intelligence 2017*.
22. Miller, G. A. WordNet: a lexical database for English. *Communications of the ACM* 1995, 38(11), pp. 39-41.

23. Ohira, Y., Uchida, T., Minato, T., Ishiguro, H. A Study of Error Recovery Methods for User Information Acquisition in Daily Dialogue. The 35th Annual Conference of the Japanese Society for Artificial Intelligence 2021, pp. 4E1OS11a04-4E1OS11a04. (in Japanese)
24. Lee, J. H., Kim, M. H., & Lee, Y. J. Information retrieval based on conceptual distance in IS-A hierarchies. *Journal of documentation* 1993.
25. Mihalcea, R., Corley, C., & Strapparava, C. Corpus-based and knowledge-based measures of text semantic similarity. In *Aaai* 2006. Vol. 6. pp. 775-780.
26. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
27. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
28. Ogawa, Y., Kikuchi, H. Effect of Agent's Self-Disclosures on its Personality. *Human-Agent Interaction Symposium* 2011.
29. Oshio, A., Shingo, A. B. E., & Cutrone, P. Development, reliability, and validity of the Japanese version of Ten Item Personality Inventory (TIPI-J). *Japanese Journal of Personality*. 2012, 21(1).
30. Clark, H. H. *Using language*. Cambridge university press. 1996
31. Norman, D. A. *The psychology of everyday things*. Basic books. 1988
32. Steyn, R., & Ndofirepi, T. M. Structural validity and measurement invariance of the short version of the Big Five Inventory (BFI-10) in selected countries. *Cogent Psychology* 2022, 9(1), 2095035.
33. McCrae, R. R., Yamagata, S., Jang, K. L., Riemann, R., Ando, J., Ono, Y., Angleitner, Y., Spinath, F. M. Substance and artifact in the higher-order factors of the Big Five. *Journal of personality and social psychology* 2008, 95(2), 442.
34. Tsutsui, S. *Structural analysis in Japanese casual conversation*, 1st ed.; Kurosio Publishers: Tokyo Chiyoda 2-4-3-8F, Japan, 2012
35. Jourard, S. M. *The transparent self*. Van Nostrand Reinhold Company.1971
36. Kawahara, D., & Kurohashi, S. A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference* 2006, pp. 176-183.