



Article

A Dialogue System That Models User Opinions 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, 1-3, Machikaneyama, Toyonaka 560-8531, Japan

² Advanced Telecommunications Research Institute International (ATR), Kyoto 619-0237, Japan

³ Rikagaku Kenkyūjyo (RIKEN), 2-1 Hirosawa, Wako 351-0198, Japan

* 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 the creation of an elaborate design by the designer is a challenge. 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 in which the system models the user. In the field of opinion dialogue, there has been little study on the topic of transition methods for modeling users while maintaining their motivation to engage in dialogue. To model them, we adopted information content. Our contribution includes the design of a rule-based dialogue system that does not require an 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 the questions gives 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 to engage in dialogue with the system.

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



Citation: Ohira, Y.; Uchida, T.; Minato, T.; Ishiguro, H. A Dialogue System That Models User Opinions Based on Information Content. *Multimodal Technol. Interact.* **2022**, *6*, 91. <https://doi.org/10.3390/mti6100091>

Academic Editor: Georgios N. Yannakakis

Received: 20 August 2022
Accepted: 30 September 2022
Published: 13 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, systems that are able to engage in 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: machine learning approaches [4] and rule-based approaches [5]. The former generally requires large corpus data. When using existing corpora, however, it is difficult to control the utterances of the dialogue system. There are context management difficulties, which make 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 large-scale content. Compared to this, rule-based methods have advantages in terms of controlling utterances and context management, and they 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 that can produce them as intended by the designer.

Understanding (modeling) the user is also important for dialogue systems. Uchida et al. [6] 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 recommendations through dialogue [7]. Therefore, it is desirable for dialogue systems to model users to increase their dialogue motivation as well as to improve the estimation accuracy. Additionally, there is a benefit of reducing the cost of implementing a dialogue system by clarifying how to realize a dialogue system that understands users based on the collected data.

In this study, we focus on non-task-oriented dialogue among dialogue systems and, among them, opinion dialogue. Opinion dialogue is dialogue in which subjective opinions are exchanged, and it is said that exchanging subjective opinions is a type of self-disclosure [8]. Therefore, it reflects the user's personal values, etc. There are several possible reasons 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. [9] describe dialogue as a way to collectively observe hidden values, intentions, and cultural differences. In other words, it can be interpreted as users being motivated to engage in dialogue when they become aware of values that they were unaware of through dialogue. The desire to have one's opinion known (self-expression) is one of the generally recognized motivations to engage in dialogue, and the function of self-expression is listed as one of the functions that activate interaction. Hiraki et al. [10] 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 they can 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 in which 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 the questions of users. In contrast, there have been few approaches aimed at the latter. That is, there are 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 has been pursued.

To address this issue, this paper builds an opinion expression model that can elicit user 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.

When a dialogue system is engaging in dialogue, speech presenting a general opinion or that questions the user's opinion in the form of a specific opinion is considered effective, such as in the 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 the topic. It is undesirable to ask such users ambiguous questions with the expectation that they will spontaneously give specific opinions. Furthermore, if they are repeatedly asked questions about topics that they are not interested in, they may become less willing to interact with the system.

In human-human dialogue, the dialogue changes gradually depending on the relationship. For example, when there is no information about the user, the topics that the user holds subjective opinions on are unknown. Therefore, it is possible to determine the opinions that they hold by asking questions that are easy for many people to answer or by learning about their experiences first. If the user modeling is advanced, however, presenting topics that the user is likely to have opinions on will be efficient for collecting their subjective opinions. In this study, we examine the topic selection rules for what kinds of questions should be selected to enable 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 experiments with this dialogue system and impression evaluation, we examine what kinds of estimation methods cause a user to

feel that a dialogue system understands them better. 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

In this paper, we focus on three key aspects of dialogue system design: making the dialogue system smaller, modeling the user, and maintaining the willingness to interact. Dialogue systems that model users represent a different approach compared to question-answering agents such as ELIZA [5] or machine-learning dialogue systems such as LUNA [11]. User modeling can be broadly divided into "identity-based user modeling" and "knowledge-based user modeling" [12]. Identity-based user modeling uses static information that is connected to individuals, such as gender and age. Knowledge-based user modeling, on the other hand, uses modeling that matches structured data, predefined rules, and existing user information. This paper focuses on "knowledge-based user modeling", as we aim to realize a dialogue system that models users while reducing the cost of system construction by utilizing knowledge models with simple structures. A commonly used user modeling method that utilizes knowledge models is collaborative filtering [13]. It can estimate user opinions from known user data based on the degree of similarity and the co-occurrence of opinions among users. However, this method has been reported to be insufficiently effective when the amount of user data is small [14]. A situation in which there is not enough user data in a dialogue suggests that the user has not spent much time using that dialogue system. Therefore, methods to maintain motivation to interact in situations in which there are not enough data strongly contribute to preventing users from leaving the system. Methods such as object grouping [15] and knowledge exploitation [16] have been proposed to solve this problem. However, these do not take into account interactive methods that require additional information input from the user to maximize estimation. On the other hand, there are several studies that consider human-machine interactions in order to learn user preferences. [17–19]. However, there are two issues with these methods: first, they do not provide sufficient estimation when the amount of user data is small, and second, they require a large data set. Uchida et al. [6] and Macros et al. [20] address these two issues. Uchida et al. use a thought model on preference data to estimate the perspective from which preferences are made. Macros et al. [20] solved these problems by using the label ranking method [21], a type of ranking method. However, these estimations can only handle likes and dislikes and cannot handle multiple adjectives. Adopting an evaluation axis other than "like-dislike" requires taking the imbalance of data and the differences in the averages for each evaluation axis into account. Additionally, the number of options is the product of the number of concepts and the number of evaluation axes, so the number of options is much larger. However, in order to realize a dialogue that handles opinion models, an expression model that allows multiple evaluation axes and that is able to handle more general chat dialogue situations is necessary.

On the other hand, it is also necessary to consider dialogue design architectures that reduce the cost of designing dialogue systems. Approaches based on big data are suitable for the open domain, but the process of collecting data is costly. However, there are also approaches that reduce the cost of dialogue system construction by using transition learning and other methods. These approaches involve relearning the dialogue system to produce the ideal response. Wang et al. [22] solved these problems by proposing the GPT-Adapter-CopyNet. On the other hand, in the case of flow-based rule-based dialogue, the dialogue designer must pay a time cost proportional to the amount of content. Furthermore, the quality of the dialogue system depends on the dialogue designer. Therefore, a practical dialogue system that can generate dialogue content from a knowledge database is desired; Manuhara et al. [23] and Muthugala et al. [24] proposed a dialogue service robot that manages dialogue by means of finite-state interaction modules. In [23], besides responding to questions, the human nature of the interaction is reinforced by the random selection of

extended dialogues, such as the ability to request further explanation or to make appropriate comments. However, this dialogue system mainly consists of one question and one answer, making it difficult to realize a long-term dialogue that spans multiple turns. For the purposes of user modeling, the architecture must be able to realize multiple turns of interaction on the topic to be modeled and to execute optimal planning.

In addition, when the number of data that a knowledge model can represent is large, a knowledge model with proportionally increasing relational data, such as a table, is not desirable. This is because the data become sparse, making it difficult to utilize the relationship data between concepts. A knowledge model suitable for referring to local relational data is a semantic network. For example, when searching for topics related to the current topic, it is sufficient to refer to adjacent nodes if they are represented in a semantic network. Additionally, when defining the distance between concepts, the length of the path between two nodes can be utilized. There is a study in [25] that uses semantic networks for dialogue management. In [25], a method that naturally leads to the assigned long-term goal while also achieving the short-term goal is proposed. It can be used to achieve the short-term goal of maintaining consistency while achieving the long-term goal of modeling users.

This paper proposes a method for modeling user opinions interactively, assuming multiple third-party models and allowing a high degree of freedom in the expression of opinions but with little data associated with the user. It has been shown that the dialogue system's willingness to understand the user's dialogue contributes to the motivation to engage in dialogue on a given topic [6]. However, the relationship between estimation methods and motivation has only been studied to a limited extent. Therefore, in this paper, we propose a system for estimating user opinions and investigate its effect on the subjective impressions of users.

3. Proposed Method

3.1. Opinion Model

Knowledge of a 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 every noun and adjective is described on a different semantic network. The advantage of this is that it is easier to express the polysemy of a particular noun. For example, consider a case where we describe the relationship among three nouns: "pizza", "pizza served at restaurant A", and "restaurant A". The word "pizza" is a general noun that is not limited to being served at a restaurant, while "pizza served at restaurant A" is a noun that is limited to pizza that can be served at a specific restaurant. Here, both "pizza" and "pizza served at restaurant A" are food items and can be included in the expression "Did you eat X?" In this way, processing differs depending on the type of relationship with other nodes and is the same regardless of whether the relationship exists or not. 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 our previous study [6], it was 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 by simply 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), we want to know which method is superior: referring to all of the data or only to 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 to determine whether one topic is useful for estimating opinions on the other. However, the opinions are limited to the same topic. In our previous study [6], it was not determined whether this method is effective in the case of mixed topics,

such as “pizza” and “cake”, where it is not certain that the tendencies of opinions will be the same. Hence, this study adopts semantic networks to model user opinions.

3.2. Dialogue System

The purpose of this dialogue system is to model users in order to maintain their motivation to use the model. Modeling users refers to estimating the opinions of specific users. However, 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 [26]. The information that the dialogue system is looking for to model the user does not necessarily match what questions that can be asked in the current situation. Additionally, the semantic network is needed in the process of modeling the user. While related research has been conducted in [27,28], little research has been developed for the purpose of being used in dialogue. Since there are 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, 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 that can increase the accuracy of the dialogue system by utilizing data with a simple structure and that requires as little manual maintenance as possible.

Opinions are expressed as a combination of nouns and adjectives. Therefore, the opinion dialogue system is divided into two major phases: a noun identification phase and an adjective identification phase. The first phase is the function of correctly recognizing the user’s intention (Intention level). It identifies the user’s intention in response to the dialogue system’s experience questions and considers the answers to these questions as nouns that constitute opinions. The second phase is a function that asks questions that facilitate user modeling based on the user’s opinions collected through the dialogue and the system’s knowledge (opinion level). The intention level is a dialogue process in which the user’s answers to a single system experience question are obtained and can be regarded as task-oriented dialogue led by the dialogue system. On the other hand, the opinion level is similar to that of non-task-oriented dialogue in that the user chooses one of the possible statements based on the current context. Young et al. [29] have seamlessly connected task-oriented and non-task-oriented dialogues without making a distinction between the two. In this paper, while non-task-oriented dialogue and task-oriented dialogue are not independent of each other, the dialogue model separates the determination of “nouns” and “adjectives” that constitute opinions into their respective dialogue levels and repeats “noun convergence” and “adjective divergence” to achieve both the breadth and depth of the chosen topics.

The intent level has two additional functions: the first is to determine if the user has an answer relevant to the topic or the question and to determine if it is appropriate to ask it (interest or knowledge); the second function is to correctly recognize what the user has said and intended in response to the user’s answer and to recover if the user’s response was not recognized. These constructs adopt the dialogue modeling approach of [30] et al. This approach has strong capabilities necessary for dialogue, such as error recovery, and supports the realization of natural dialogue. Table 1 shows the types of error recovery, and Table 2 shows the types of utterances. The definitions of “Open”, “Closed”, and “Talk” in Table 2 are shown in Table 3. The model of [30] consists of four Acts and four CAs, with the same correspondences shown in Table 4 in this paper as in [30].

Table 1. 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.
Not interested	Open topic	"I'm not interested"	"Are you interested in a meal?" and if yes, end the dialogue.

Table 2. Type of utterance.

Label	Description
Open genre	Ask a question about GENRE in a "WHAT".
Closed genre	Ask a question about genre in a "YesNo".
Open menu	Ask a question about menu in a "WHAT".
Closed menu	Ask a question about menu in a "YesNo".
Open restaurant	Ask a question about restaurant in a "WHERE".
Closed restaurant	Ask a question about restaurant in a "YesNo".
Open topic	While changing the topic, ask the experience question "What have you eaten before?"
Closed 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

Table 3. Types of utterance and perlocutionary acts.

Label	Description
Open x	Extract a word from the word type corresponding to x from the user's utterance and add it to the context.
Closed x	If the user answers yes to a question from the dialogue system that contains the word corresponding to x, x is included in the context; if the user answers no, it is stored in the no-exp variable.
Open topic	Empty context. Only keep the elements that appear in the previous dialogue system utterance and leave the rest empty.
Closed opinion	If the user answers "yes" to the system question, the option stored in context is added to the user model; if the user answers "no", it is added to the no user model.
Talk opinion	Add nouns that appear in the system's statements to the context.

Table 4. Communicative acts and opinion dialogue system responses.

Acts	Definition in Opinion Dialogue System
Illocutionary act	One or more are set for each locutionary act. The definition depends on the locutionary act.
Locutionary act	Table 2 presents the correspondence with the illocutionary act. One locutionary act is defined for each word type and for each combination of word types.
Perlocutionary act	More details in Table 3.
Communicative act	Intention level: The purpose is to slot-fill a CONTEXT. Opinion level: Close-opinions are generated for opinions that can be expressed by nouns in the context. The opinion selection rules are explained in the “Opinion Dialogue” section. Interested or not: See Table 1: “Not interested”. No intention: Indicated in Table 1: “Forget”.

The above is the minimum structure required to construct an opinion dialogue system using hierarchically structured noun data. For each word, there is a word type, and for each word type, multiple words are defined. For example, the word type “genre” includes the word “pizza”. Context is expressed as a combination of word type and the word. In this paper, we implemented the dialogue system shown in Figures 1 and 2.

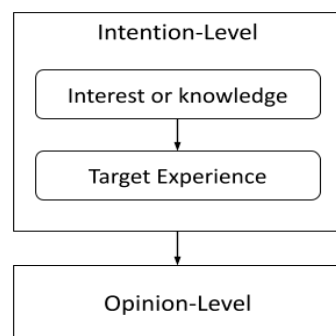


Figure 1. Process flow.

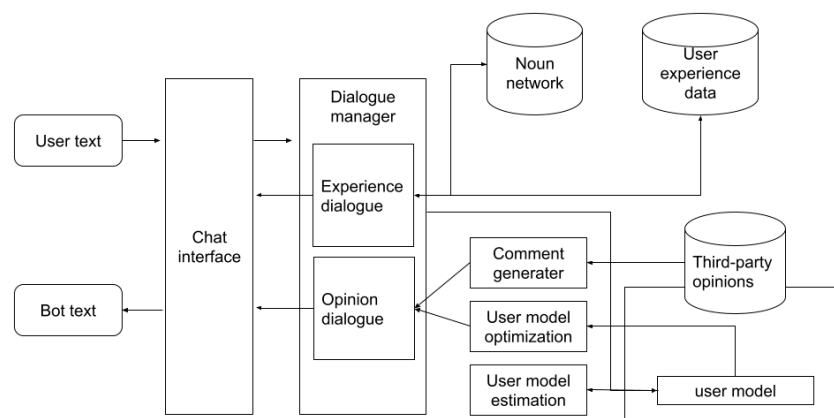


Figure 2. Dialogue system architecture.

Intention-level dialogue corresponds to “Experience dialogue”. The second category 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 the “Experience dialogue” and “Opinion dialogue” phases. Figure 3 shows the flow from “Experience dialogue” to “Opinion dialogue”. This represents one cycle.

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 this process is repeated twice in this system. The reason for repeating this process twice is that if three or more questions are asked, then the user will feel that the dialogue system is asking questions in a monotonous and one-sided manner, and their motivation for engaging in dialogue with the system will decrease. We thought that the user would be able to determine 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 two questions. If the user’s intention was unclear or if their intention did not exist in the system’s knowledge, an error recovery operation was performed.

In this study, opinions are composed 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 the “Experience dialogue” and “Opinion dialogue” phases. In order to identify nouns in the “Experience dialogue” phase, as shown in Figure 1, it is necessary to specify the candidates by considering the nouns 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 are a combination of polarity and nouns that express whether the user has experience.

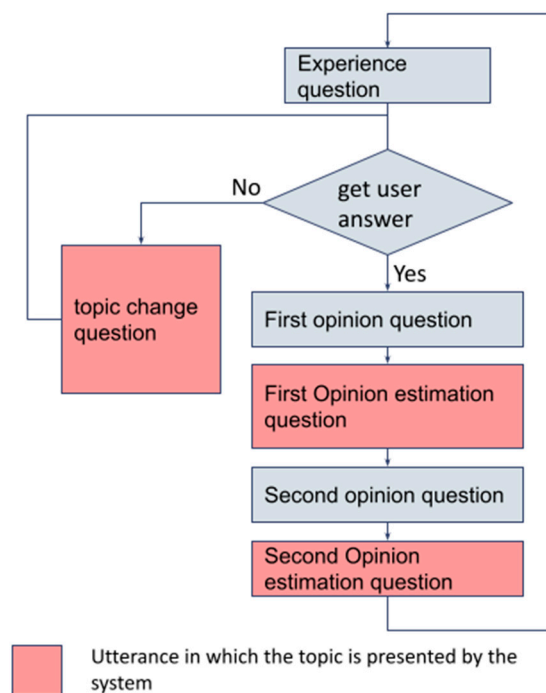


Figure 3. Dialogue flow.

The “Comment generator” generates general opinions from “third-party opinions”, representing opinion data. This is used in the “Opinion dialogue” phase to reduce the proportion of questions in the dialogue act. This is because a dialogue consisting of only questions directed from the dialogue system to the user will decrease in the user’s motivation to engage in dialogue. The “User model optimization” phase estimates opinions that should be added to the current user model and generates questions. The “User model estimation” phase estimates and extends the user model from the current user model and is not explicitly stated in the dialogue. It is not directly involved in utterance generation, but it does indirectly influence 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 that are deeply related to people's daily lives, and it is highly likely that many people have opinions about these topics. For example, with opinions such as "pizza is good". First, opinion data on food were collected through a crowd-sourced questionnaire survey. The questionnaire items are listed in Table 5:

Table 5. Opinion data collection items.

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

Eight adjectives were adopted to evaluate the impressions of specific menus. Restaurants were selected from restaurants that have a nationwide presence in Japan and that are considered well-known throughout the country. Some other restaurants offering similar food and beverages were omitted. This was to avoid bias in the opinion data collected. The 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 them, were removed, and semi-solid items such as shakes were classified as sweets. Opinion data were collected for an average of 31 items per menu and 367 items per person for 493 food and beverage items offered by the selected restaurants.

From the opinion data, a set of specific restaurants (restaurant) and their menus (menu) were 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 was adopted for each menu item.
3. If there are multiple menu components, the main one is used.
4. A menu genre that can be inferred from the restaurant where the menu item is served is assigned.
5. If there are abbreviations or synonyms, they are treated as the same.

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 added to menus as such, 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 stated on menus when offered at specialty stores. For example, some menus offered at hamburger restaurants use the phrasing "X burger", while others are referred to as hamburgers and omit the "burger" part. These were adjusted manually. These subcategories of nouns and adjectives such as "menu", "restaurant", "genre", and "adjective", which are the components of opinions, are hereinafter referred to as categories. The "noun network" in Figure 1 describes a semantic network of the relationships between noun categories, the relationships between restaurant and menu, and the relationships between genre and menu. The genre–menu relationship is referred to as "IsA", and the restaurant–menu relationship is called "SERVE". IsA represents the relationship "is a kind of" and is one of the most commonly used relationships in semantic networks [31]. SERVE is the relationship between a restaurant and the food and drink served there. This paper has the following hierarchical structure: restaurant, genre, and menu. Therefore, these relationships are used to infer opinions about restaurants and genres from opinions about the menu and

to transition topics from higher-level concepts to lower-level concepts. The relationship between *IsA* and *SERVE* is treated basically as the same, but the difference is only utilized during sentence generation. “User model” and “third-party opinions” are described as pairs of menus and adjectives.

3.4. Dialogue to Specify Intention

In this dialogue system, the “Intention level” identifies “what have you eaten at restaurants before?”. At the opinion level that follows, the dialogue system asks opinion questions such as “Do you think it tastes good?” and “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 it makes an utterance to return to a dialogue that achieves the system’s goal. This is called error recovery. Error recovery in the “Determining strategies based on interest and 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”. Additionally, the user may not remember the question. In this case, answering the question with “I don’t remember” may clearly indicate that the answer to the question cannot be provided at that time. For example, in response to a question such as “Have you ever been to (a restaurant)?” the respondent may explain their 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, then the user may be forced to recall memories that he or she does not remember, or the user’s motivation to engage in 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.

An open topic is an utterance that resets the current context and that presents a new topic. System response (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 cheeseburgers has not been mentioned immediately beforehand. The purpose of (1) is to encourage the user to recall the experience. The details of this are explained in the section “Intention Estimation by Elaboration”. The “Not interested” rules indicate 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 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 answered the system’s questions about). 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 two methods described below 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 menus and options. Since each menu has its own name, users suggesting the name of the menu directly would be ideal. However, it is not always possible for the user to provide a specific 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 non-specific answers such as “I have eaten a variety of foods”. This is assumed to be because the intention of the question is too vague, too many things come to mind to narrow down to one, or the memory being 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 “Common Sense Candidate Reasoning”. 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 closed questions is to encourage users to recall their experiences. Therefore, it is desirable that the candidate answers the questions presented by the dialogue system through closed questions referring to concepts that are close to or that 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 item. The user who receives this utterance considers the relevance of utterance (3) based on the immediately preceding context given in (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, then 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 of the user thinking that the experience presented by the system is what the user has experienced. 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. A specific example follows:

- SYSTEM: “What have you eaten (a restaurant)?”
- USER: “Various.”
- SYSTEM: “Have you ever eaten pizza, for example?”

Pizza is a genre, a concept that encompasses multiple menus, that is included in the “noun network” in 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 item. Therefore, in this dialogue system, the flowchart in Figure 4 was implemented to make the “noun network” correspond to the intention-specific dialogue.

The extract category determines the categories for which the context is empty in the order of restaurant, genre, and menu. The categories of open question and closed question are determined depending on this category. If no answer is obtained after one open question, then a closed question is asked up to two times. Since a closed 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 an open question, there is a

wide range of user responses, and this can lead to errors, such as ambiguous responses or the 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 an open question is placed before a closed question is that a closed question is the one-way repetition of questions from the dialogue system, which is undesirable from the viewpoint of motivation to engage in dialogue. The reason for repeating a closed question up to two times is also the same: repeating the same dialogue action multiple times may remind the user of the mechanical behavior of the dialogue system.

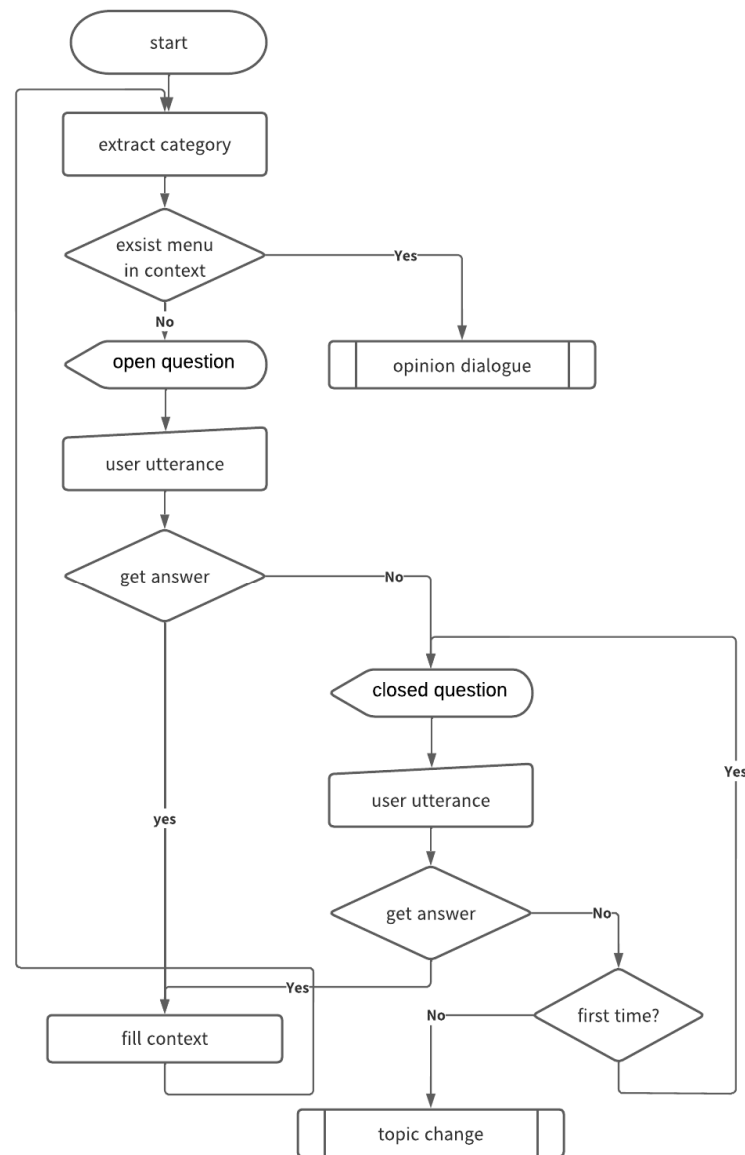


Figure 4. Dialogue system flowchart.

The concept of context was described above and refers to a set of variables that describe the state of the dialogue according to the format in Table 6:

Table 6. Context used in dialogue to specify intention.

Label	Description	Rule
Restaurant	Name of restaurant	
Genre	Name of menu genre	Allow unknown when menu is known
Menu item	Name of food	
Opinion	Opinion on menu item	

In Table 6, the levels of abstraction are higher at the top and lower at the bottom. Question generation, such as which genre to select from a particular restaurant and which menu item to select from a particular genre–restaurant combination, is explained in the next chapter. It also identifies the restaurant if the genre is filled first and it identifies the restaurant even if the menu is determined. However, in situations in which the menu is determined, we added a rule stating that the restaurant could not be identified. The reason for this is that the relationship between a restaurant and menu is SERVE, while the relationship between genre and a menu is IsA. When using the experience of the user, the menu–restaurant relationship is not specified (the dialogue system cannot determine it without confirming that the presented genre or menu is offered by the restaurant with the user), whereas in the case of IsA, the menu–restaurant relationship is not specified. In the case of IsA, this is because of the social structure of the menu itself and is not dependent on the user’s experience. For example:

- SYSTEM: “What have you eaten at (a restaurant)?”
- USER: “I have eaten Margherita pizza.”
- SYSTEM: “Is Margherita pizza a pizza?”

Asking this question to the user is an act that emphasizes the lack of common sense of the dialogue system and may decrease the user’s motivation to engage in dialogue, and is also useless in terms of efficiently identifying the user’s intention. For this reason, we have added a rule that if a sub-concept of “IsA” exists, then the blank space in the upper-level concept is acceptable. The dialogue labels to fill in the above context are listed in Table 2.

For each of the categories listed in Table 2, open and closed are implemented, as is open topic. However, open opinion was not implemented; two question types, open and closed, were provided for each column. This is the minimum structure that was needed to implement the following hypotheses:

1. Rules for prioritizing the subject of the topic presentation. Since the purpose of dialogue is to determine 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 according to 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. Rules for choosing between open and closed questions. In some cases, users may not be able to recall their own experiences clearly in response to the system questions. In a chat dialogue, a highly abstract question such as “What did you do yesterday?” may not allow the user to recall concrete details immediately. Therefore, if the user does not provide a clear answer, it is necessary to provide an example from the dialogue system.
3. Rules for prohibiting duplication of synonymous questions. If the user does not give a meaningful answer to a particular question, 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 for the system to repeatedly ask “What did you eat?” Repeated mechanical responses will decrease the user’s motivation to engage in dialogue. The question “What did you eat at that time at X?” is a solution to some extent, as it changes the question to be different from “What did you eat at that time?” However, this method cannot solve the problems that occur when the question’s intention is accurately conveyed to the user and no answer is obtained. In addition, asking the user “Did you eat X?” followed by “Then did you eat X?” and “Then did you eat X?” repeatedly and the user saying the same thing over and over again but with different words indicates that the probability of the system being able to infer the user’s intention is low. However, if there are only two or three candidates for estimation, then it is acceptable to say, “Then, did you eat X?” in the sense of “If it is not X, then it is only X”. 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, then that 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 closed 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 closed questions, the question type cannot be repeated due to condition (3). Therefore, it is necessary to change the topic by using the open topic label. Talk opinion and closed opinion are described in the “Opinion Dialogue” section.

3.5. Dialogue to Specify Intention

3.5.1. Intention Estimation Handling Ambiguity

Dialogue systems need to identify user responses from user statements. In the chapter “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 the “Intention Estimation by Elaboration” chapter, specific intentions were identified from ambiguous intentions one by one. However, if the user directly states a word that indicates the item (menu) to be inferred, then the menu can be identified without the above means. Intention estimation by slot filling, as in the “Intention Estimation by Elaboration” chapter, has the advantage of the dialogue proceeding under the initiative of the dialogue system, but on the one hand, the number of dialogue turns taken 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 according to the 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 to engage in 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 to engage in dialogue. Therefore, it is desirable to use the two intention estimation methods together.

Even when the system has narrowed down 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 outlined in "Intention Estimation by Elaboration", we would make statements such as "Do you think Margherita pizza is good?" However, if the user feels that the conversation is being interpreted in a way that is different from their intention, then it becomes difficult to continue the dialogue. The solution is to change the response depending on the level of confidence. This can be achieved with unknown words, as described below, or with confirmatory dialogue cues 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, then 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, decreasing the user's motivation to engage in dialogue. However, we believe that the dialogue system itself will be more likely to be accepted by the user if it indicates that it is unsure about 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?" 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 this case, the user may not be able to correctly word the formal menu name word for word. 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 answers from other candidates that the dialogue system keeps in its internal memory. Additionally, 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 6 are included is determined. KNP [32] was used for morphological analysis. In this example, it was judged that the user's utterance "hamburger store" contains the words "hamburger" and "shop". Therefore, since all of 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 making up this designation included in the user's utterance is hereafter referred to as the percentage of agreement. Figure 5 shows an example of a notation model.

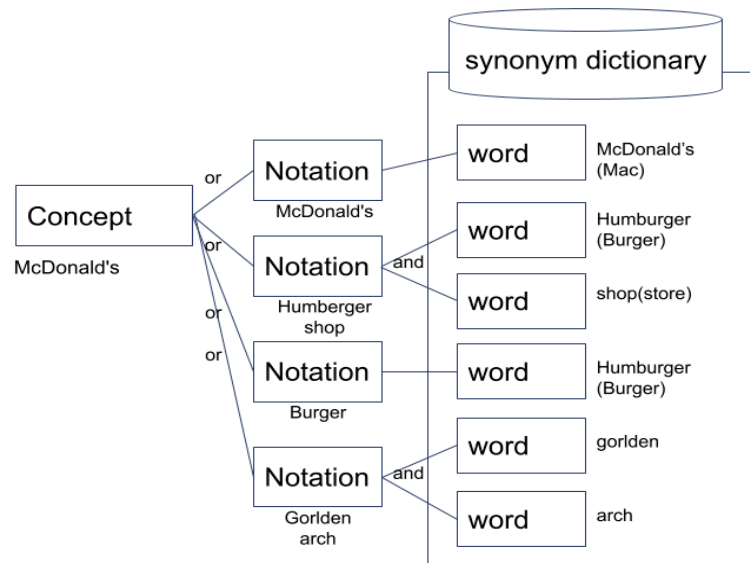


Figure 5. Notation blur of hamburger restaurant.

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 to represent 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 calculated the maximum percentage of matches for each invocation. The closer this value is to 1, the more likely it is that the user accurately referred to the name held by the system, and the lower the value, the more likely it is that 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 correspond 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 to the category itself in order to identify the object to which it refers. For this reason, we have nouns such as “place” that refer to a 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 a notation-warping dictionary when there were multiple expressions for each instance. When a restaurant uses the expression “ordinary” for a standard menu item included in a specific genre, the combination of ordinary + menu item name, etc., is registered.

- (3) Dictionary of word distortion.

While the instance notation variation dictionary 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 if the context of a dialogue is maintained are

registered. Specifically, word variations in katakana, hiragana, kanji, and Japanese-English and abbreviations such as “hamburger” and “burger” are registered.

3.5.2. Common Sense Candidate Reasoning

When asking a closed question, a specific noun should be presented from the context. Specifically, after the user states that he has been to restaurant A, the system should ask, “Have you ever eaten at X?” and should generate a question when the user asks a 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, all menus that have a relationship with the 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 a menu:

- 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 are represented in Table 5, which includes columns for people, adjectives, and nouns. Taking the adjectives and nouns as inputs, we can obtain a list of Boolean values for the number of third parties stored by the system from the overall data. At this stage, the context of the dialogue does not include adjectives. In other words, in order to execute the algorithm above, it is necessary to fix the adjectives to one. Here, the adjective is fixed to “good”. A frequency of 1 represents the percentage of true values in the list of Boolean values for the number of third parties. The first method refers to method 1, and the second method refers to method 2. Method 2 differs from method 1 in terms of the third parties referenced: while method 1 refers to all people, method 2 only refers to the data of a subset of people who are judged to be close to the user. The details of method 2 are described in the “Opinion Dialogue” section.

During the inference of intent by elaboration, the inference of common-sense candidates corresponds to closed restaurant, closed menu, and closed genre. These are always performed after the open candidates, as shown in Figure 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 the section “Selection Algorithm Based on Information Content”.

3.6. Opinion Dialogue

After identification of intended concepts, opinion dialogue is executed. The purpose of the identification of intended concepts is to recall the user’s experience and to 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. Inferring other opinions from the user model.

In case 1, an example question might be “Do you think the cheeseburger (sold at restaurant A) is greasy?”, and the user provides a direct answer of Yes or No. While this is highly accurate as a user model, it requires N turns of conversation to obtain N user models. Value 2 is the presumption that other opinion data that are not explicitly affirmed by the

user are also considered as a user model. This is an efficient modeling method because it allows 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 modeling with fewer dialogues.

Additionally, if random question generation is used, the user may wonder “Why did the dialogue system ask these questions?” 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 is referred to here as “question intention”. For example, if the dialogue system repeatedly asks questions that everyone agrees on, then 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 things that are “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 to engage in dialogue is likely to change depending on whether or not the user can sense “question intent” from the dialogue system’s questions and the type of question intent the user has.

During opinion dialogue, users are modeled in two ways: in method 1, the type of questions generated is important for modeling and motivation to engage in 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 during opinion dialogue. In the “Estimated Opinion Model” section, 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 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 obtained from 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.

Studies classifying dialogue breakdowns [26] refer to breakdowns caused by unrelated topics. Unrelated topics refer to utterances that deviate from the previous topic, and they are one of the causes of dialogue breakdown. This dialogue system asks questions to the user multiple times. In such cases, if the question that immediately follows the previous one is not related, it may be considered to be an irrelevant topic. To prevent this, the system evaluates the closeness of the opinions obtained by the immediately preceding question to the opinion to be 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 there is not a distance measure that can classify them with a high level of granularity. However, it is possible to consider the association between adjectives for opinions about food, for example, “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 from different users. The 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” [33], and its application to word similarity was suggested to make certain contributions in [34] as well as in other publications. Examples of semantic networks include wordnet [28] and word2vec [35]. However, for the reasons explained in the “Data Collection” section, it was necessary to use data for real restaurants as well as to use the names of the menu items that they sell. The data were partially created manually based on wordnet. The above sets the restriction that only opinions whose distance in the semantic network is within a certain proximity of the immediately preceding opinion and the next opinion to be presented are presented. This prevents the appearance of unrelated topics [26]. This threshold is hereafter referred to as the “similarity topic threshold”.

2. Evaluation of the proximity of opinion tendencies among people.

During opinion dialogue, it is necessary to implement an estimation method to determine other opinions from user models. We utilized some third-party data as a method for this, in contrast to the method of Uchida et al. [6], in which opinions are estimated by a statistical evaluation that takes all of the 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 conventional methods in that it does not refer to “the data of others who are not judged to be close”. This is a device that forces the method to only use local data in order to make it applicable at all times, even if the opinion expression space or the overall data are 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 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 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 who has been judged to have preferences similar to those of the user is sought out, then the opinion data of that user should be considered as the user’s estimated opinion. However, the third party’s opinion data contain 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 point, 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; using b, we can estimate third parties with similar opinion tendencies to those obtained from the user’s statements; and using 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 described in Table 7.

Table 7. 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 selected to verify the similarity of nodes is a menu, if this threshold is 2, then all other menus belonging to a 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?” refers to 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 a 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 by 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. [6] 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 dialogue system “understands the user better”, and the level of dialogue 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.
- Ensuring that the opinions that are obtained are “rare opinions”.
- Both of the above.

The following are possible: Knowledge of many opinions contributes to “user understanding”. However, even if the dialogue system has knowledge of many opinions that are not rare, if there is no knowledge that differentiates the user from the majority, then 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 the “dialogue strategy” below.

If the method of estimation is acceptable to the user, then 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”, then 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 using 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:

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

where $P(E)$ is the probability that event E occurs, where E is the event that the user holds a particular opinion on. This probability is the probability of a third party also holding that opinion. For example, if the dialogue system remembers 50 user models and

20 users hold the opinion that “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:

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

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

where est is a function that estimates the user model when a question q is selected. In (3), $f(\{ I(o) | o \in O \})$ is calculated for the set of candidate opinions O that can be obtained through dialogue, and the 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”. The function $f(X)$ used in this study is shown below:

$$count := |X| \quad (4)$$

$$max := \max(X) \quad (5)$$

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

where 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 the “Selection Algorithm Based on Information Content” section, we presented three types of indicators to evaluate the final user model: count, max, and sum. The count indicator 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 that are agreed on by many users can be obtained through dialogue means that the model to be estimated tends to have more opinions that are agreed on by many users. The maximum indicator includes the amount of information in the equation, but the number of pieces estimated does not affect the results. In other words, the dialogue system only focuses on the rarity of the opinions and does not consider how many can be estimated. This type of high information content and the low number of relevant opinions is a common tendency among uncommon user types. The sum indicator 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 [36], so the system is also analyzed using these data.

The main experiment was conducted after a preliminary survey. In the preliminary survey, third-party opinion data were collected from 40 people online. The subjects were between 20 and 59 years old and resided in Japan. The questions consisted of three items: basic information (gender, age, occupation, prefecture of residence, and hobbies), user personality characteristics (Japanese version of the TIPI-J (Japanese version of the Ten-Item Personality Inventory) [37]), and opinion data (Table 5). This dialogue system was constructed based on the data obtained from the preliminary data collection session. The main experiment was then conducted. Subjects were selected following the same method used for preliminary data collection. In total, 75 subjects responded to this experiment.

Dummy questions were asked during both the pre-survey and during 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 collecting basic information, personality characteristics [37], and opinion data.

The conditions are Formulas (4)–(6), and the order is randomized for each subject. When determining the impressions about the conversation section, a five-point scale from "does not apply" to "applies well" was set for the following items:

- Did they try to understand you?
- Did they understand you?
- Would you like to talk to them again?
- Satisfied with the dialogue?
- Was the presumption natural?

Interface

The dialogue interface provides notes at the top of the screen and a chat-style dialogue at the bottom of the screen. Figure 6 shows the interface.

The dialogue system speaks to the user after the user presses "Start". The name of the dialogue system is coded as "Robot". The dialogue system ends when the following conditions are met:

- Opinion dialogue is performed a total of four times.
- Topic presentation is performed 12 times in total.

This dialogue system generates a delay in speech. This is for the following two reasons: to avoid giving the user the feeling that the dialogue system is responding mechanically and to draw the user's attention to what the dialogue system is saying. Specifically, this is to prevent the user from responding randomly without reading the sentences in detail. It is known that in person-to-person dialogue, participants alternate turns, and a certain pause occurs between them [38]. It is also known that differences in robot reaction times can greatly affect the prediction made in the robot's response [39]. Therefore, the dialogue system was significantly delayed for statements that we wanted the participant to pay attention to in particular. 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 to delay the start of the turn:

- The user speaks 3 s after the dialogue system generates its utterances.

With the above rule, the delay time varied depending on the time taken to process the generated utterances. Since the opinion estimation process takes time in this dialogue system, the delay time only increases for opinion utterances. This may lead the user to predict that the system's opinion utterances are more important than the other utterances.

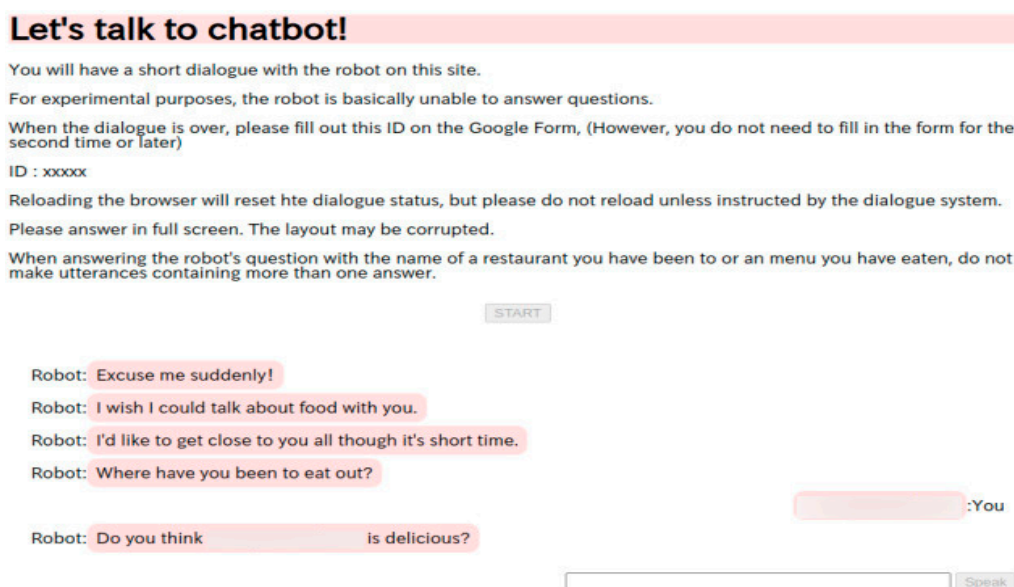


Figure 6. Chat interface. Note: Mosaic is applied to the part where the restaurant name is identified. Japanese was used for the experiment, and this image has been translated into English.

In addition, the following rules were applied for in-turn delay:

- Separation with a punctuation mark (end of a sentence).
- Speak 3 s after the previous utterance.

This delay takes into account the time it takes for a person to type a response. When an utterance is repeated without delay, this represents a behavior that is difficult for a person to achieve, thus giving the impression of machine-like behavior.

5. Result

In total, 74 subjects were recruited, and analyses were conducted on 38 subjects who met all of the following requirements:

- No dialogue breakdowns occurred throughout the entire dialogue.
- The termination conditions for all dialogues were met.
- All dummy questions placed in the opinion survey questionnaire were answered correctly.

The TIPI-J was used to determine the users' personality traits, and five factors [40,41] 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 TIPI-J ranged from 0 to 16, we divided the TIPI-J into three groups: 0 to 5.3 as the low group, 5.3 to 10.6 as the medium 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 openness, 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 8.

Table 8. Subjective evaluation of each dialogue strategy.

Personality Traits	Questionnaire Item	Condition			ANOVA <i>p</i> -Value	Sub Effect Tests Ryan-Method <i>p</i> -Value (Pair)
		Max	Count	Sum		
Openness (<i>n</i> = 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		
	Satisfaction	2.91	2.85	2.39		
Extroversion (<i>m</i> = 26)	Talk again	2.81	2.92	2.58	0.0631 +	
	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		
Agreeableness (<i>n</i> = 21)	Talk again	2.87	2.89	2.58		
	Estimation naturalness	3.08	3.03	2.89		
	Understand	2.89	2.97	2.87		
	Try to understand	3.63	3.76	3.42		
Conscientiousness (<i>n</i> = 32)	Talk again	2.78	2.94	2.44	0.0496 *	
	Estimation naturalness	3.00	3.06	2.84		
	Understand	2.75	2.97	2.81		
	Try to understand	3.63	3.78	3.34		
Neuroticism (<i>n</i> = 22)	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		

The “ANOVA *p*-value” does not show *p*-values below 0.1. Additionally, “sub-effect test using Ryan method *p*-values (pair)” does not show *p*-values below 0.05. + $p < 0.1$, * $p < 0.05$. Try to understand: “Did they try to understand you?”, Understanding: “Did they understand you?”, Talk again: “Would you like to talk to them again?”, Satisfaction: “Satisfied with the dialogue?”, Estimation naturalness: “Was the estimation natural?”.

The results showed that max and count were superior to sum in the “satisfied with the dialogue” indicator and in the “Openness” indicator. In addition, “Would you like to talk to them, again?” and “Did they try to understand you?” also showed higher scores for count than sum in the openness condition.

Table 9 shows the amount of user information at the end of the dialogue for each method. The amount of information collected for direct opinions is the amount of information obtained from the user model obtained when the dialogue system asked “Do you think X is good/expensive/cheap?” The amount of information obtained from the estimated opinion is the amount of information derived from the opinion estimated by the system without confirming it directly with the user.

Table 9. Parameters after dialogue.

	Max	Count	Sum
Sum of information content of directly heard opinions (average)	11.26 (1.413)	9.619 (1.380)	9.741 (1.337)
Sum of information content of opinions estimated (average)	39.37 (2.141)	51.67 (2.253)	50.33 (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 the estimation accuracy on the impression of the dialogue, the correlation between the estimation accuracy (percentage of correct answers) and the impression evaluation was examined. The results are described in Tables 10–13.

Table 10. Correlation coefficients between the percentage of correct answers for directly asked items and the questionnaire index.

	Agreeableness	Extraversion	Conscientiousness	Neuroticism	Openness
Did they try to understand you?	0.0404 (0.691)	0.128 (0.264)	0.0249 (0.819)	−0.0821 (0.493)	0.0404 (0.691)
Did they understand you?	0.0172 (0.866)	0.0652 (0.571)	−0.0314 (0.773)	−0.0269 (0.823)	0.0172 (0.866)
Would you like to talk to them again?	0.0531 (0.602)	0.0966 (0.400)	0.00909 (0.933)	0.0321 (0.798)	0.0531 (0.602)
Satisfied with the dialogue?	−0.0360 (0.723)	0.0126 (0.913)	0.0974 (0.370)	0.0345 (0.774)	0.0360 (0.723)
Was the presumption natural?	0.0350 (0.731)	0.0881 (0.443)	0.0201 (0.853)	−0.00574 (0.962)	0.0350 (0.731)

Table 11. Correlation coefficients between the percentage of correct answers for the estimated items and the questionnaire index.

	Agreeableness	Extraversion	Conscientiousness	Neuroticism	Openness
Did they try to understand you?	0.132 (0.193)	0.204 + (0.0730)	0.168 (0.120)	−0.0118 (0.921)	0.132 (0.193)
Did they understand you?	0.204 + (0.0730)	0.360 ** (0.00121)	0.304 ** (0.00425)	0.143 (0.229)	0.258 ** (0.00992)
Would you like to talk to them again?	0.168 (0.120)	0.319 ** (0.00442)	0.265 * (0.0130)	0.120 (0.315)	0.252 * (0.0118)
Satisfied with the dialogue?	−0.0118 (0.921)	0.377 ** (0.000675)	0.305 ** (0.00404)	0.179 (0.132)	0.280 ** (0.00497)
Was the presumption natural?	0.132 (0.193)	0.315 ** (0.00491)	0.250 * (0.0193)	0.124 (0.298)	0.244 * (0.0150)

+: $p < 0.1$, *: $p < 0.05$, ** $p < 0.01$.

Table 12. Correlation coefficients between the information content for directly asked items and the questionnaire index.

	Agreeableness	Extraversion	Conscientiousness	Neuroticism	Openness
Did they try to understand you?	0.0261 (0.797)	0.127 (0.269)	0.0117 (0.914)	−0.117 (0.330)	0.0261 (0.797)
Did they understand you?	0.127 (0.269)	−0.0399 (0.729)	−0.0937 (0.388)	−0.125 (0.300)	−0.0639 (0.530)
Would you like to talk to them again?	0.0117 (0.914)	0.00657 (0.954)	−0.0534 (0.623)	−0.0725 (0.545)	−0.0267 (0.793)
Satisfied with the dialogue?	−0.117 (0.330)	−0.0421 (0.714)	−0.135 (0.213)	−0.0959 (0.423)	−0.0880 (0.387)
Was the presumption natural?	0.0261 (0.797)	−0.0170 (0.883)	−0.0698 (0.521)	−0.103 (0.388)	−0.0577 (0.571)

Table 13. Correlation coefficients between the information content for the estimated items and the questionnaire index.

	Agreeableness	Extraversion	Conscientiousness	Neuroticism	Openness
Did they try to understand you?	0.123 (0.225)	0.181 (0.113)	0.141 (0.191)	0.0381 (0.751)	0.123 (0.225)
Did they understand you?	0.183 + (0.0697)	0.291 ** (0.00969)	0.205 + (0.0569)	0.101 (0.400)	0.183 + (0.0697)
Would you like to talk to them again?	0.146 (0.149)	0.180 (0.115)	0.154 (0.155)	0.0791 (0.509)	0.146 (0.149)
Satisfied with the dialogue?	0.174 + (0.0842)	0.222 + (0.0507)	0.199 + (0.0649)	0.152 (0.202)	0.174 + (0.0842)
Was the presumption natural?	0.113 (0.263)	0.180 (0.116)	0.117 (0.280)	0.0929 (0.438)	0.113 (0.263)

+: $p < 0.1$, ** $p < 0.01$.

“Opinions directly questioned” refers to the opinions that the dialogue system asked the user about in the dialogue session. The “estimated opinions” are the opinions obtained from the user model at the end of a dialogue and exclude the opinions that were asked about directly. 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 regarding 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, the mothers who participated in the survey tended to have a lower neuroticism tendency compared to 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 they understand you?” The correlations were only confirmed for the estimated opinions, and none of the correlations were confirmed for the opinions that were asked about directly (see Box 1).

Box 1. An example dialogue is presented in the box below. Note: The parts that can identify actual stores are indicated in generic notation with [].

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]'s 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 restaurant]'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 am going to end now. Thank you very much.

6. Discussion

6.1. User Modeling Strategies

In this study, we developed a dialogue system to model user opinions, and differences in user motivations to engage in dialogue were examined depending on how the opinions were modeled. The results showed that sum was inferior in terms of motivation to engage in 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 in how the opinions are modeled. Chat dialogues are generally not only about exchanging information, and there is also the goal of continuing the dialogue itself. Therefore, attitudes towards understanding the dialogue system are not necessarily important to the user. On the other hand, even in chat dialogues, it is thought that some people's motivation to engage in dialogue is greatly influenced by what the other party has said rather than by the social rules of the dialogue. In Table 8, only openness showed significant differences among the three conditions of "satisfied with the dialogue", "Would you like to talk to them again", and "Did they try 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 to engage in dialogue. In [24], which investigated the relationship between personality traits and dialogue, it was suggested that openness and main effects are factors, and the results obtained here are similar.

Table 9 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, the sum is greater than the count, and the count is greater than the max. This suggests that max is not a superior method from the point of view of advancing user modeling and that sum is a superior method. On the other hand, Table 8 shows that sum was inferior to the other two methods

during the subjective evaluation of “satisfied with the dialogue”, “Would you like to talk to them again”, and “Did they try to understand you”.

No significant differences were observed in the information content of the opinions that were explicitly stated between sum and count. However, there was a significant difference in the information content of the opinions inferred. In other words, it can be suggested that the users were able to sense the tendency of the opinions being inferred from their own opinions through the dialogue. However, this does not deny the possibility of there being other hidden factors. The results indicate that there may be a relationship between differences 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 13 (Correlation coefficients between the estimated information content and the questionnaire index). The results show a significant trend toward a weak correlation between motivation to engage in 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 user models. On the other hand, no significant differences were observed in the correlation coefficient between the information content asked about directly 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 affecting motivation to engage in dialogue. This indicates that users are able to experience the estimates of the dialogue system through the dialogue itself and that their motivation to engage in dialogue is influenced by the information content that the estimates contain. The considerations that led to these results are as follows: First, “opinions that are asked about directly” are constrained by the structure of the dialogue. Specifically, when asking a question to a user, the user may feel uncomfortable if a question with high information content is asked (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 that contain a high level of information. Therefore, there are two possible factors that affect the motivation to engage in 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 that are asked about directly”. Therefore, it is no relationship where the greater the information content, the higher the motivation to engage in dialogue. On the other hand, if we interpret that “the information content of the estimated opinion” has nothing to do with the dialogue environment and is only related to the degree of understanding, then it could explain why a positive correlation between the information content and motivation to engage in dialogue was only confirmed for the estimated opinions. Clarifying this factor is an issue to be discussed in future research.

The relationship between the percentage of correct answers and motivation to engage in dialogue is discussed. Table 11 (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 a tendency toward neuroticism was not found to be significantly different, but this may be due to the small population size of participants with neurotic tendencies. Therefore, it is suggested that regardless of personality traits, a high rate of correct answers for presumptive opinions may improve impressions of the dialogue. However, there was no association with the rate of correct answers in the posture of understanding. This can be interpreted as trying to understand the user is not something that is able to be evaluated when the dialogue system is making estimations. Therefore, it is possible that there is 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. This is not a clever dialogue system that accurately estimates the user's opinion model, but rather a dialogue system that acts as an agent for the user and the dialogue system, mutually deepening their understanding of each other. Therefore, it should not be used as a method for modeling actual users but as a partner system that maintains a willingness to engage in dialogue through two-way understanding. On the other hand, the rate of correct answers to the estimated opinions was correlated with the motivation to engage in dialogue. As mentioned above, users can experience the results and trends of the estimations made by 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 previous dialogues was not consistent with their own self-perceptions. In Table 11 (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, with the exception of neurotic tendencies. A low correct response rate means that if one of the objectives of the user's dialogue is mutual understanding, then the user feels that they could not fully achieve this dialogue objective. This feeling may have a negative effect on such indicators as motivation to engage in dialogue and dialogue satisfaction, regardless of whether someone has this personality trait. It is difficult to determine whether the neurotic tendency's different results compared to other personality traits may be due to the small number of data or whether there are factors that only appear in the neurotic tendency.

This experiment was carried out via a chat-style dialogue, and there was no interface through which the user could infer the personality or characterization of the dialogue system. For the user, the dialogue system was 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 the first face-to-face dialogue, the system 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 only be asked after a relationship has been established between the two parties. 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 to engage in dialogue is based on the desire to communicate without limiting the recipient, such as in contexts of self-disclosure [42], then this dialogue experiment can achieve this. If the desire is based on the social act of dialogue itself, then 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, the "sum" and "max" methods can be used.

The dialogue system collects third-party data in advance. 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, then the estimation accuracy is expected to increase with the number of people. Under the conditions of this experiment, a higher number of people is preferable because there is a positive effect on the motivation to engage in dialogue when the estimations have a high level of accuracy. On the other hand, when trying 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 amount of third-party data is not large, it is possible to engage in natural dialogue in the same way as people do. If the number of third-party data is small, the system 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 system has repeatedly communicated with others. In terms of this type 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 held. 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 is correlated with the motivation to engage in dialogue. In other words, there is a possibility that the user will accept even a small amount of third-party data if the system makes utterances that show behavior that is not confident in its words and actions and by telling the user that the amount of third-party data is small. According to impression evaluations in dialogues in which the dialogue system's 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 to engage in dialogues with a system with a small amount of third-party data and low estimation accuracy 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. As such, we will discuss the general experimental results as they pertain to other topics. First, the relationship between the estimation method and the amount of data is discussed above. Food is something that people think about 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 tend 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 have 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.

Here, we discuss the differences in estimation methods (max, count, and sum). If we refer to Figure 6, it can be seen that the 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 intention 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 emerged as a result of the combination of multiple menu items. When a user receives a question, they consider the intention of the dialogue system, but the lack of an intuitive answer to the question may have a negative effect. When chatting, the dialogue is composed of multiple topics. The dialogue is limited to a specific topic for several turns, and when it becomes difficult to continue the conversation, the topic is shifted. Thus, simple and specific questions seem more natural within a topic. It is possible that when the intention of a question refers to the sum of the rarity of multiple opinions, it was perceived as complicated by the user. The sum method is also contrary to the structure of the dialogue and may have given the user the impression that topic transitions were system-driven and that there was no in-depth exploration. On the other hand, not many differences were observed in the subjective evaluation of max and count. In Uchida et al. [6], it is reported that the count method is optimal. The max method, which is a method that maximizes the information content of a specific opinion, has the same level of evaluation as the count method does during subjective evaluation and is considered to be one of the most reliable algorithm estimation methods in topic presentation studies. On the other hand, the max method has the lowest average number of overlapping opinions in Table 9, and the sum and count methods have the same level of overlapping opinions. This means that the sum and count methods have less spread-out estimations compared to the max method. 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 that improves with each

successive dialogue. On the other hand, in Table 9 (Number of types of opinions held by user models), the max method only increases slightly in relation to the count method. This is due to the low average number of opinions estimated for max. Although the difference between the max and count methods 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, the count method 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 fewer interactive elements, such as questionnaires. Li et al. [43] also showed that context is useful for sentiment classification in terms of sentiment analysis. It is possible that the user's opinion in this experiment was influenced by the experience or opinion utterances presented by the user immediately before. Therefore, if a different flow of dialogue is used than that used in this experiment, the results may differ from those in this experiment. The results of this experiment cannot necessarily be claimed with generality.

6.2. Ambiguity in User Speech

This chapter discusses how dialogue systems recognize ambiguous and inaccurate user statements. The ability of a dialogue system to smoothly recognize inaccurate information from a person contributes to reducing the load caused by human communication. This is especially important in non-task-oriented dialogues in which it is difficult to provide clear motivation. Spontaneously spoken instructions and responses in spoken communication contain uncertain information, lexical symbols, and concepts [44]. It is unlikely that the user can remember all of the noun strings that the dialogue system has. Therefore, ambiguous expressions from users are likely to occur frequently during text communication in dialogue. This is common in that it is a burden to remember expressions that match perfectly. In other words, expressions that reduce the amount of information to a level that is subjectively meaningful to the user tend to be preferred over expressions with a large amount of information. At the intention level of this specific dialogue system, the menu is specified. However, for the question "What have you eaten recently?", the response "I had a Margherita pizza" must be assumed as well as the response "I had a pizza". "Pizza" is a concept that encompasses "Margherita", and a knowledge model that can be traced from an ambiguous concept to a more concrete concept is desirable. For this reason, this dialogue system represents noun data in a tree structure. A knowledge model that classifies the ambiguity and concreteness of concepts by hierarchy is considered to be friendly to human communication.

Cordes et al. [44] and Muthugara et al. [45] also argue that one of the ambiguities expressed by people is the expression of "degree". This paper deals with opinions, and many of the adjectives used to describe subjective expressions contain degree information, such as "expensive", "sweet" and "delicious". Consider the case of "tuna salad". This dialogue system only keeps "tuna salad is healthy" as data. The interpretation of whether this is a relative evaluation or an absolute evaluation is divided. Other possible responses include "Tuna salad is healthy compared to seafood salad", "I usually eat a lot of fatty foods, so tuna salad is healthy compared to them", "I simply have the impression that tuna salad is healthy", and "Tuna salad is especially healthy among salads". Information on degree is also missing, and the overall criteria for whether or not there is a difference may differ. A way to resolve the ambiguity caused by human degrees is the robot experience model (REM) developed by Muthugala and Jayasekara et al. [24]. The REM is a hierarchical structure that organizes a robot's knowledge about the environment, actions, and context, and uncertainty depends on the environment and experience. We have shown that whether a user says a food is "healthy" or not depends on the user's environment and previous

experience. To translate this into the topic of this experiment, which is the evaluation of food and drink, we can consider the distribution of foods that the user sees within the user's area of activity and how frequently the user eats certain foods. This ambiguity is difficult to resolve because this experience and environment cannot be unified by the user. Conversely, if this experience and environment can be estimated, the ambiguity of the user's own opinion can be resolved.

Mavridis et al. [46] propose ten items and sequences that a dialogue robot should have. Among them, "Affective interaction" and "Motor correlates and Non-Verbal Communication" are not implemented, and the lower-ordered "Purposeful speech and planning" is the argument of this paper. The experiment does not use images that give the user an impression of the dialogue system, nor does it use emotional expressions, and the dialogue system does not make verbal utterances. The multimodal component is a temporal delay affecting the timing of the dialogue system's utterances, which expresses the sense that the user is communicating at the same time as the dialogue system and also gives a sense of importance to the utterance that is reflected in the thinking time of the dialogue system. In the design of this dialogue system, we eliminated noise from the multimodal expression when viewed from the perspective of validating the algorithm for estimating opinions about the user. However, it is suggested that the lack of "emotional expressions" and "multimodal interfaces", which are more primitive than the "opinion estimation algorithm for users", may affect the interaction of the "opinion estimation algorithm for users" or may contribute to the recognition of user statements. The validation of an "opinion estimation algorithm for users" that satisfies these multimodal interfaces and that takes their interactions into account is a challenge for future research.

7. Conclusions

For people with personality traits that are highly open to dialogues in which a dialogue system asks questions to the user, the dialogue system is more willing to adopt the number of user models rather than the sum of the amount of information as a measure for optimizing user modeling. When the user models obtained after the dialogue were evaluated from the viewpoint of the total amount of information, the method that prioritized the largest amount of information in the user models was inferior, and when evaluated by the number of opinions, the method considering the total amount of information was the most superior. It was also found that the amount of information and the percentage of correct answers for indirectly presented questions had a more positive effect on subjective evaluation than it did on questions that were presented directly. This suggests that users can perceive differences in the criteria for understanding the user of a dialogue system through dialogue rather than through superficial question exchanges.

This dialogue system can be applied to other dialogues by creating a noun list and an adjective list and by collecting third-party opinion data, noun distortion data, and noun–noun relationship data. Third-party opinion data, which are large-scale data, can be collected through questionnaires. Therefore, a dialogue system can be constructed without the dialogue designer having to put in much effort. Since it is possible to accumulate impression data on concepts (e.g., pizza, Tokyo, etc.) through the chatbot, this method can be applied to the quantitative evaluation of impressions of advertisements and catch phrases. Since opinion models of individual interlocutors can be estimated, they can be applied to product recommendations. Opinion models of individuals can also be applied to person-to-person matching and clustering based on similar interests and preferences. This classification of people's opinion tendencies is useful for analyzing personality characteristics and classification tendencies limited to specific topics rather than for 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. and T.U.; formal analysis, Y.O. and T.U.; investigation, Y.O. and T.U.; resources, Y.O., T.U., T.M., and H.I.; data curation, Y.O.; writing—original draft preparation, Y.O. and T.U.; writing—review and editing, Y.O. and 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 and JSPS KAKENHI under grant numbers 19H05692, JP20K11915, 19H05693, and 22K17949.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of ATR (protocol code 21-605, 23 March 2021 approval).

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

Data Availability Statement: Data are included in this paper.

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

- Dominey, P.F.; Paléologue, V.; Pandey, A.K.; Ventre-Dominey, J. Improving quality of life with a narrative companion. In Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Lisbon, Portugal, 28 August–1 September 2017; pp. 127–134.
- Sabelli, A.M.; Kanda, T.; Hagita, N. A conversational robot in an elderly care center: An ethnographic study. In Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI) 2011, Lausanne, Switzerland, 6–9 March 2011; pp. 37–44.
- 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 Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 906–913.
- Thoppilan, R.; De Freitas, D.; Hall, J.; Shazeer, N.; Kulshreshtha, A.; Cheng, H.T.; Le, Q. Lambda: Language models for dialog applications. *arXiv* **2022**, arXiv:2201.08239.
- Weizenbaum, J. ELIZA—A computer program for the study of natural language communication between man and machine. *Commun. ACM* **1966**, *9*, 36–45. [[CrossRef](#)]
- 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. *Int. J. Soc. Robot.* **2021**, *13*, 1499–1516. [[CrossRef](#)]
- Konishi, T.; Sano, H.; Ohta, K.; Ikeda, D.; Katagiri, M. Item Recommendation with User Contexts Obtained through Chat Bot. In Proceedings of the Multimedia, Distributed, Cooperative, and Mobile Symposium, Sapporo, Hokkaido, 28–30 June 2017; pp. 487–493. (In Japanese).
- Higashinaka, R.; Dohsaka, K.; Isozaki, H. Effects of empathy and self-disclosure in dialogue systems. In Proceedings of the Association for Natural Language Processing, 15th Annual Meeting, Tokyo, Japan, 2–7 August 2009; pp. 446–449.
- Bohm, D.; Factor, D.; Garrett, P. Dialogue: A Proposal 1991. Available online: http://www.david-bohm.net/dialogue/dialogue_proposal.html (accessed on 1 September 2022).
- Hiraki, N. *Assertion Training*, 1st ed.; Kaneko Shobo: Tokyo, Japan, 1993.
- Dinarelli, M.; Stepanov, E.A.; Varges, S.; Riccardi, G. The luna spoken dialogue system: Beyond utterance classification. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14–19 March 2010; pp. 5366–5369.
- Ma, Y.; Nguyen, K.L.; Xing, F.; Cambria, E. A survey on empathetic dialogue systems. *Inf. Fusion* **2020**, *64*, 50–70. [[CrossRef](#)]
- Goldberg, D.; Nichols, D.; Oki, B.M.; Terry, D. Using collaborative filtering to weave an information tapestry. *Commun. ACM* **1992**, *35*, 61–70. [[CrossRef](#)]
- 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, Denver, CO, USA, 7–11 May 1995; pp. 210–217.
- Ungar, L.H.; Foster, D.P. Clustering methods for collaborative filtering. In Proceedings of the AAAI Workshop on Recommendation Systems, Madison, WI, USA, 26–31 July 1998; Volume 1, pp. 114–129.
- Burke, R. Integrating knowledge-based and collaborative-filtering recommender systems. In Proceedings of the Workshop on AI and Electronic Commerce, Orlando, FL, USA, 18–19 July 1999; pp. 69–72.
- Kobayashi, S.; Hagiwara, M. Non-task-oriented dialogue system considering user’s preference and human relations. *Trans. Jpn. Soc. Artif. Intell.* **2016**, *31*, 1. [[CrossRef](#)]
- Sakai, K.; Nakamura, Y.; Yoshikawa, Y.; Kano, S.; Ishiguro, H. Dialogal robot that estimates user’s preferences by using subjective similarity. In Proceedings of the IROS 2018 Workshop Fr-WS7 Autonomous Dialogue Technologies in Symbiotic Human-Robot Interaction 2018, Madrid, Spain, 1–5 October 2018.

19. Sumi, K.; Sumi, Y.; Mase, K.; Nakasuka, S.; Hori, K. Information presentation by inferring user's interests based on individual conceptual spaces. *Syst. Comput. Jpn.* **2008**, *31*, 41–55. [[CrossRef](#)]
20. Maroto-Gómez, M.; Castro-González, Á.; Castillo, J.C.; Malfaz, M.; Salichs, M.Á. An adaptive decision-making system supported on user preference predictions for human-robot interactive communication. *User Model User-Adapt. Interact.* **2022**, *9*, 1–45. [[CrossRef](#)] [[PubMed](#)]
21. Cléménçon, S.; Depecker, M.; Vayatis, N. Ranking forests. *J. Mach. Learn. Res.* **2013**, *14*, 39–73.
22. Wang, W.; Zhang, Z.; Guo, J.; Dai, Y.; Chen, B.; Luo, W. Task-Oriented Dialogue System as Natural Language Generation. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 2698–2703.
23. Manuhara, G.W.M.; Muthugala, M.A.V.J.; Jayasekara, A.G.B.P. Design and Development of an Interactive Service Robot as a Conversational Companion for Elderly People. In Proceedings of the 2018 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 30 May–1 June 2018; pp. 378–383.
24. Muthugala, M.V.J.; Jayasekara, A.B.P. MIRob: An intelligent service robot that learns from interactive discussions while handling uncertain information in user instructions. In Proceedings of the 2016 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 5–6 April 2016; pp. 397–402.
25. Ni, J.; Pandelea, V.; Young, T.; Zhou, H.; Cambria, E. Hitkg: Towards goal-oriented conversations via multi-hierarchy learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Crete, Greece, 17–20 June 2022; Volume 36, pp. 11112–11120.
26. Higashinaka, R.; Funakoshi, K.; Araki, M.; Tsukahara, Y.; Kobayashi, Y.; Mizukami, M. Text Chat Dialogue Corpus Construction and Analysis of Dialogue Breakdown. *J. Nat. Lang. Processing* **2016**, *23*, 59–86. [[CrossRef](#)]
27. Speer, R.; Chin, J.; Havasi, C. Conceptnet 5.5: An open multilingual graph of general knowledge. In Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
28. Miller, G.A. WordNet: A lexical database for English. *Commun. ACM* **1995**, *38*, 39–41. [[CrossRef](#)]
29. Young, T.; Xing, F.; Pandelea, V.; Ni, J.; Cambria, E. Fusing task-oriented and open-domain dialogues in conversational agents. In Proceedings of the AAAI Conference on Artificial Intelligence, Crete, Greece, 17–20 June 2022; Volume 36, pp. 11622–11629.
30. Fernández-Rodicio, E.; Castro-González, Á.; Alonso-Martín, F.; Maroto-Gómez, M.; Salichs, M.Á. Modelling multimodal dialogues for social robots using communicative acts. *Sensors* **2020**, *20*, 3440. [[CrossRef](#)] [[PubMed](#)]
31. Havasi, C.; Speer, R.; Alonso, J. ConceptNet 3: A flexible, multilingual semantic network for common sense knowledge. In Proceedings of the Recent Advances in Natural Language Processing, Borovets, Bulgaria, 27–29 September 2007; pp. 27–29.
32. 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; Association for Computational Linguistics*: Stroudsburg, PA, USA, 2006; pp. 176–183.
33. Lee, J.H.; Kim, M.H.; Lee, Y.J. Information retrieval based on conceptual distance in IS-A hierarchies. *J. Doc.* **1993**, *49*, 188–207. [[CrossRef](#)]
34. Mihalcea, R.; Corley, C.; Strapparava, C. *Corpus-Based and Knowledge-Based Measures of Text Semantic Similarity*; American Association for Artificial Intelligence: Palo Alto, CA, USA, 2006; Volume 6, pp. 775–780.
35. Goldberg, Y.; Levy, O. word2vec Explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv* **2014**, arXiv:1402.3722.
36. Miyamoto, T.; Iwashita, M.; Endo, M.; Nagai, N.; Katagami, D. Influence of Utterance Strategies to Get Closer Psychologically on Evaluation of Dialogue in a Nontask-oriented Dialogue System. *Trans. Jpn. Soc. Artif. Intell.* **2021**, *36*, AG21-I. (In Japanese) [[CrossRef](#)]
37. Oshio, A.; Shingo, A.B.E.; Cutrone, P. Development, reliability, and validity of the Japanese version of Ten Item Personality Inventory (TIPI-J). *Jpn. J. Personal.* **2012**, *21*, 40–52. [[CrossRef](#)]
38. Clark, H.H. *Using Language*; Cambridge University Press: Cambridge, UK, 1996.
39. Norman, D.A. *The Psychology of Everyday Things*; Basic Books: New York, NY, USA, 1988.
40. Costa, P.T.; McCrae, R.R. *Neo Personality Inventory-Revised (NEO PI-R)*; Psychological Assessment Resources: Odessa, FL, USA, 1992.
41. 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. *J. Personal. Soc. Psychol.* **2008**, *95*, 442. [[CrossRef](#)] [[PubMed](#)]
42. Ogawa, Y.; Kikuchi, H. Effect of Agent's Self-Disclosures on its Personality. In Proceedings of the Human-Agent Interaction Symposium, Kyoto, Japan, 3–5 December 2011.
43. Li, W.; Shao, W.; Ji, S.; Cambria, E. BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing* **2022**, *467*, 73–82. [[CrossRef](#)]
44. Cordes, L. Who speaks?—Ambiguity and Vagueness in the Design of Cicero's Dialogue Speakers. In *Strategies of Ambiguity in Ancient Literature*; De Gruyter: Berlin, Germany, 2021; Volume 114, p. 297.
45. Muthugala, M.A.V.J.; Jayasekara, A.G.B.P. A Review of Service Robots Coping with Uncertain Information in Natural Language Instructions. *IEEE* **2018**, *6*, 12913–12928. [[CrossRef](#)]
46. Mavridis, N. A review of verbal and non-verbal human-robot interactive communication. *Robot. Auton. Syst.* **2015**, *63*, 22–35. [[CrossRef](#)]