Development Paper

INFORMATION AND TECHNOLOGY IN EDUCATION AND LEARNING, Vol. 2, No. 1 2022
https://doi.org/10.12937/itel.2.1.Dev.p001

Exploratory Analysis of Usage Statistics of Dialogue Systems by Visualization

Yosuke Seki*

(Received 16 June 2021 and accepted in revised form 1 March 2022)

Abstract It is important to analyze the usage statistics of dialogue systems in order to identify improvements that can increase user satisfaction. However, further beneficial information may be hidden among the information (such as logs and registered usage statistics in the database) that is not output as a result of the general analysis methods and can be discovered through exploratory analysis. To this end, this study proposes a method for the exploratory analysis of the usage statistics of dialogue systems using visualization to detect beneficial information, either intentionally or unintentionally. Specifically, usage statistics are visualized as network graphs showing the queries and transitions among them, and functions to extract particular visualized objects and identify characteristic queries are implemented to support the analysis. The results of this evaluation show that the proposed method can detect beneficial information through an exploratory analysis of visualized usage statistics, thereby demonstrating its effectiveness in the analysis of dialogue systems.

Keywords exploratory analysis support, visualization, dialogue systems

1. Introduction

Various industries have adopted dialogue systems to provide users with information quickly and easily. For example, airlines have introduced dialogue systems to guide tourists to sightseeing resorts, and online shopping sites have introduced these systems to support item selection. Because dialogue systems enable users to access the required information at any time and location, they are considered to be one of the most effective methods to support the collection of information. As a new method for supporting students' campus life and public relations activities in universities, dialogue systems have captured the attention of educational institutions[1], [2]. At Tokushima University, a task-oriented dialogue system called TOKUPON-talk (hereinafter referred to as "T-talk")[3] was introduced for public relations with high school students and other stakeholders. In this system, various types of information (such as university features and lifestyles in Tokushima) are provided as textual information.

The analysis of usage statistics (such as detecting user types and the number of queries) is important for system operation because the analysis results can, for example, contribute to identifying any improvements required in the dialogue system to increase user satisfaction. Using the general analysis methods for dialogue systems, system administrators acquire the results (such as the number of users and their query content) generated by the standard analysis functions. Information that is not provided by these standard functions may be beneficial and can be obtained using exploratory methods. However, detecting such beneficial information is challenging because logs and other information sources include a considerable amount of text information, necessitating a significant amount of time and effort to analyze targeted information, such as relationships, trends, characteristic queries, and other statistics. Furthermore, methods that enable administrators to perform such analyses for dialogue systems have not yet been reported.

This study proposes a method to visualize the usage statistics of dialogue systems, enabling exploratory analysis.[1] Specifically, network graphs are used to visualize usage statistics to show queries and the transitions between them, and functions that can extract

* Research Center for Higher Education, Tokushima University, Japan

[1] This study developed the method[4] and conducted a new evaluation.

©Japan Society for Educational Technology & Japanese Society for Information and Systems in Education 2022. This is an open access article published under a Creative Commons Attribution 4.0 International (CC BY-NC-ND 4.0) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

particular visualized objects and identify characteristic queries to support the analysis are implemented. The visualization of usage statistics will enable the discovery of characteristic objects, which can start an analysis intuitively and allow for various trends and user behaviors to be detected intentionally or unintentionally. Based on the analysis results from various usage statistics using the proposed method, implementing recommendation functions considering query trends and improving the accuracy rate by retraining dialogue models are, for example, expected. Thus, it is possible to collect information concerning students' campus life (for university students), their selection of universities after graduation (for high school students), and other related information through improved dialogue systems.

The proposed method targets general task-oriented dialogue systems and is designed using a general dialogue system for the public relations of a university as a representative case. The main contributions of this study are as follows:

- An exploratory analysis of the usage statistics of dialogue systems was conducted using a novel visualization method.
- A visualization method for usage statistics, including the definitions of characteristic queries for analysis, was designed.
- The proposed method enables the detection of beneficial information that cannot be easily detected using general analysis methods.

2. Related Work

Studies on dialogue systems[5] in the fields of natural language processing and artificial intelligence have made significant progress. Many of these studies have focused on user support, including recommending information[6], predicting user satisfaction[7], [8], supporting decision-making[9], education support[10], [11], and providing health advisory systems[12]. These studies have contributed to solving user issues and increasing user satisfaction through dialogue systems. However, they did not focus on methods that enable the analysis of the introduced dialogue systems.

Alternatively, some studies have analyzed the effectiveness of dialogue systems following their introduction. Among them, a few studies have analyzed the conversations between users and dialogue systems in order to determine where failures in progressing the conversation lie[13], in addition to the usage patterns of first-time users, to serve as a reference for the design of future dialogue systems[14]. Other studies have analyzed the expectations and interactions of users with dialogue systems[15] as well as their advertising effectiveness, such as helpfulness and usefulness[16]. However, because they are not designed for daily system operations, they are unsuitable for capturing necessary information. Furthermore, these studies used only logs, interviews, and experiments designed for specific purposes during their analysis; however, the analysis still requires a significant amount of effort. Although it is desirable to analyze usage statistics easily and regularly to identify improvements, methods to support these analyses for satisfactory system operation have not yet been fully realized.

Since visualization can facilitate an intuitive understanding of the overall situation, it is considered an effective method for detecting beneficial information in logs and other sources. Researchers have used network graphs to visualize different types of information. For example, some studies have visualized academic papers, presentation slides, and knowledge for information sharing in laboratories[17], [18], while others have visualized the relationships among academic papers to elucidate the relationships between articles and search results[19], [20]. Researchers have also developed analysis tools using legal precedent data to address the complexity of many nodes and links[21]; however, although the effectiveness of visualization in intuitively understanding information was discussed in these studies, methods specifically suited to dialogue systems were not proposed.

In analysis methods targeting dialogue systems, tools to evaluate dialogue models by simulated dialogues or automatic analysis have been developed[22], [23]. The analysis results, such as missed transitions in the models, were presented using these tools. However, they targeted pre-introduced evaluation, and usage statistics after introduction were not analyzed; thus, a real usage statistics analysis method for improving dialogue systems to further support the collection of information is required.

3. System Design

3.1 Research Direction

This study enables the exploration and discovery of beneficial information through the visualization of logs and other information sources. Since users submit queries one at a time in a dialogue system, usage statistics can be expressed in network graphs. The queries submitted by users are represented as nodes, and the transitions between queries are represented as edges in the network graphs. The network graphs contribute to the understanding of the usage statistics intuitively, such as "who submitted what queries in what order." This visual awareness can lead to several discoveries. For example, characteristic queries (such as the most frequently submitted queries) can be discovered by focusing on the dense parts of the network graphs. In addition, because network graphs can easily display other queries and transitions around the target queries, they enable unexpected discoveries, such as query tendencies, relevant queries, and other information related to the target queries. Thus, it is possible to analyze usage statistics in detail by exploring the graphs. Clarifying the potential of visualization for dialogue systems is important in order to establish it as a standard analysis method. This study proposes and defines the requirements of an analysis method based on the visualization of dialogue systems.

3.2 Requirements

To visually express usage statistics, it is necessary to organize the information obtained from dialogue systems and design the visualization. In this study, usage statistics were defined as the information gathered from user profiles and dialogue histories. Table 1 lists the usage statistics obtained from the general dialogue systems used for public relations at universities. Network graphs are designed using this information. Notably, the information in Table 1 may vary depending on the dialogue system used. This can be achieved by incorporating flexible customization functions in the proposed method.

To identify possible improvements in dialogue systems, it is important to understand whether unsuitable responses are provided. Thus, a method for detecting such information was also considered.

The longer a dialogue system is in operation, the larger the network graphs. In these complex network graphs, it is difficult to gather information that satisfies specific conditions (e.g., queries submitted by users, users who submitted queries more than three times, and users in the third grade of high school) and characteristic queries (e.g., queries submitted by multiple users and those resulting in several unsuitable responses). Thus, it is necessary to implement support functions for analysis. Based on these considerations, the following requirements for the proposed method are defined:

(A) Visualization of usage statistics as network graphs
(B) Filtering of visualized objects
(C) Definition and identification of characteristic queries

Item A was designed based on the usage statistics listed in Table 1 to produce the network graphs. Items B and C provide a detailed analysis of network graphs. Specifically, item B is a function that extracts particular visualized objects in order to analyze them under specific conditions. For item C, characteristic queries are defined and then identified in the network graphs. By displaying the characteristic queries and surrounding information, it is possible to support the analysis of these queries and those related to them. The combination of items B and C enabled further analysis. A method for analyzing usage statistics was designed and proposed based on these requirements.

Table 1. List of usage statistics obtained from logs and other sources of general dialogue systems for universities' public relations

Classification	Details
User profile	User ID Gender Grade (status) Hometown
Dialogue histories	Content of queries Order of queries Query fields Query frequencies Time between queries Responses Usage frequency of dialogue systems Date User evaluation against each response

3.3 Design of Visualization

(A-1). Visualization of queries and their transitions

In the network graphs, the nodes are labeled with unique node IDs. Pairs of system queries and system responses[2] are registered in generic dialogue systems. For the network graphs, the user queries were converted into the corresponding system queries so that user queries with the same content type could be combined into a single node. Information (such as system queries and user queries) is provided to the nodes to show the content of the user queries and other information.

Query fields (such as entrance exams and university features) were classified according to the dialogue system used. For example, some types of query fields are defined in the dialogue systems used by universities.[3] Visualizing the query fields makes it easier to understand the query frequency in each query field. Thus, the nodes were colored according to query field for easy identification.

An edge can include more than one transition, and the edge thickness varies depending on the number of transitions between nodes, where a thicker edge represents a greater number of transitions. The edges were colored differently for easy identification of the queries submitted by each grade (Table 1).

Fig. 1 shows an example of a network graph. A user (whose queries are indicated by green edges) submitted queries in the following order: No. 2, No. 51, No. 9, and No. 100. The query fields of the first three queries are different from those of the last query, as can be seen from the different node colors.

The proposed method enables the display of system queries, user queries, and other information by clicking on the nodes, as shown in node no. 100. For example, if a user submits a query such as, "tell me the location of Tokushima University," "where is Tokushima University located" and "tell me the location of Tokushima University," will be displayed as the system and user queries respectively. Node 71 indicates that the users submitted a query only once. The purple edge between nodes 51 and 41 is thicker than the other edges because many users submitted these queries.

(A-2). Visualization of unsuitable responses

When a dialogue system is first introduced, unsuitable responses may occur more frequently. Therefore, it is important to quickly identify user queries that result in unsuitable responses in order to improve the system. These unsuitable responses occur when the system cannot respond to queries (hereinafter called "no responses") or respond incorrectly (hereinafter called "incorrect responses") owing to undefined queries and false retrievals, among other reasons. In this study, unsuitable responses were classified as one or the other, and methods were designed to identify user queries that resulted in these unsuitable responses.

To identify user queries resulting in no responses, all undefined queries were visualized in network graphs. This allowed all undefined queries and their contents to be perceived visually. In addition, because queries related to specific information tend to be submitted continuously, responses with insufficient content can be

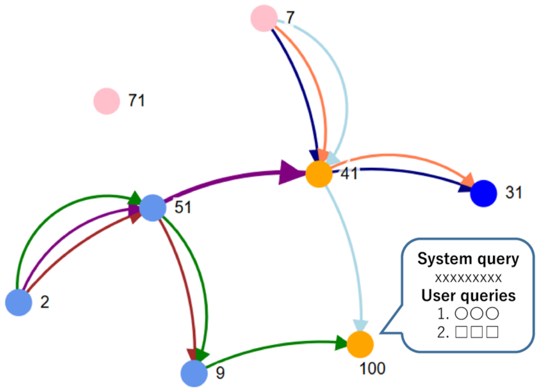


Figure 1. Example of a network graph

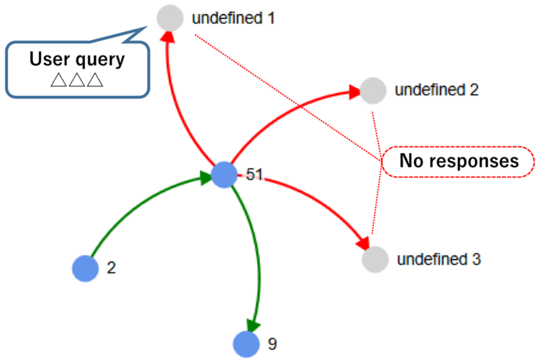


Figure 2. Example of undefined queries

[2] System queries and system responses are types of information registered in systems for dialogue. User queries are submitted by users.

[3] AI-Campus, http://www.saga-u.ac.jp/admissions/
T-talk, https://taiwa.honbu.tokushima-u.ac.jp/taiwa/ui/toku

estimated by focusing on the order of undefined queries.

Fig. 2 shows an example of an undefined query in a network graph. The grey nodes, which indicate undefined queries, are not unique and contain only one user query. Undefined user queries can be displayed by clicking on the nodes. In Fig. 2, there are three undefined queries (undefined 1–3) originating from node 51. The network graph further shows that users submitted queries in the order of no. 2, no. 51, undefined 1–3, and no. 9, which implies that the dialogue systems did not respond three times. It can be assumed that users may have submitted queries related to the content of no. 51 or no. 9; therefore, undefined 1–3 arose from insufficient response content regarding the queries represented by nodes 51 and 9.

Unlike undefined queries, incorrect responses are more difficult to identify. Since dialogue systems typically provide some type of response, it is difficult to determine the correctness of a response. Many dialogue systems contain user evaluation functions for each response. For example, buttons can be used to evaluate each response as good or bad, which are available in this study. Although the evaluation scale differs depending on the dialogue system used, this study included two values (i.e., “good” and “bad”). A bad evaluation implies that an incorrect response is generated, and the edges are colored in red in Fig. 3. For example, the edge between nodes 9 and 100 appears red, reflecting the poor evaluation; therefore, the response to the user queries at node 100 is considered to be incorrect. All user queries that resulted in incorrect responses are displayed in red.

When users submit queries that are related or

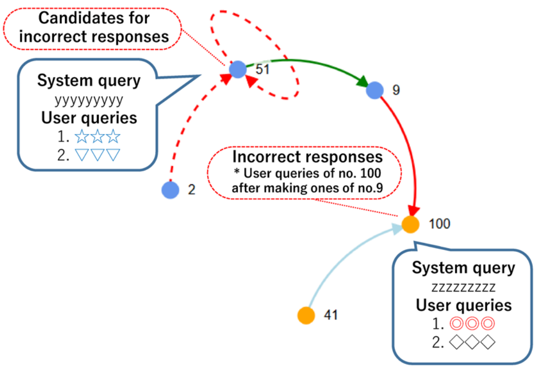


Figure 3. Example of incorrect responses

similar to previous queries, the same responses are sometimes generated as incorrect responses. For example, if users query “the schedule of the briefings for the university” after querying “the number of participants in the briefings for the university” in T-talk, the same responses are provided owing to false retrieval. Providing the same responses implies that only one response is correct or that all responses are incorrect. To identify user queries that result in these incorrect responses, the edges of the same nodes for each user are considered as candidates for incorrect responses. In the network graphs, such edges are indicated by red dotted lines. In Fig. 3, the self-edge of node 51 and the edge between nodes 2 and 51 are candidates for incorrect responses. User queries that result in such candidates are depicted in blue.

(B). Filtering of visualized objects

Filtering functions to extract particular visualized objects were designed and applied to the network graphs. The filtering conditions are listed in Table 2. Items (1)–(4) were acquired from user profiles and items (5)–(13) were acquired from dialogue histories, as presented in Table 1. For example, specific conditions, such as “queries in the field of lifestyle submitted by users who are from the third grade of high school in Tokushima” and “queries submitted by users who are women and have used dialogue systems more than three times in 2020” can be implemented, and only nodes and edges that satisfy these specific conditions are dis-

Table 2. List of filtering condition

Object	Example Value
(1) User ID	u_12
(2) Gender	Woman
(3) Grade	Third grade of high school
(4) Hometown	Tokushima
(5) Node ID	51
(6) Query fields	Exam
(7) Query frequencies	More than 10 times
(8) Time between queries	More than 30 s
(9) Undefined queries	Display
(10) Usage frequency of dialogue systems	More than five
(11) Term	1/4/2019–31/3/2020
(12) Query generating incorrect responses	Display
(13) Query from candidates generating incorrect responses	Display

played. For item (5), all edges and nodes for users who submitted queries to the selected nodes are displayed. A more focused visual discovery can be realized by removing unnecessary information from complex network graphs.

(C). Definition and identification of characteristic queries

Characteristic queries are defined and identified using mathematical techniques. It is important to define query sets that comprise the characteristic queries. Those necessary to operate dialogue systems were determined through interviews with T-talk administrators and were classified into eight different types, as shown in Fig. 4. Any number of selected nodes can be designated as a characteristic query using a red circle. Since there is a high probability that new types of queries may be discovered through operation of the dialogue system, these new types are added as needed.

The definitions of the characteristic queries depicted in Fig. 4 are as follows. Type (I) nodes correspond to queries submitted by multiple users. Nodes that follow this pattern indicate that the query is popular among users. Type (II) nodes are the origins of many other nodes and act as hubs in network graphs. Type (III) nodes exhibit high transition rates between nodes, implying that multiple users tend to submit certain queries in the same order. In this case, there is a possibility that the response content of each query is similar and may need to be combined. Type (IV) nodes represent the final queries submitted by multiple users. As such, they can be regarded as low priority and less important than their preceding queries; however, users are still interested in solving them. A type (V) node represents the first query submitted by multiple users, and thus these queries are of high priority. Type (VI) nodes occur either before or after gray nodes (i.e., undefined queries) and indicate that the queries' responses lack sufficient content; therefore, enriching the response content is necessary. Type (VII) nodes denote queries that were evaluated as "bad" by multiple users and can thus be classified as incorrect responses. Nodes classified as type (VIII) represent the case where identical responses were provided to the same users for different queries. Type (VII) and (VIII) nodes may include user queries that resulted in incorrect responses.

Table 3 presents the mathematical representations of the characteristic queries. The parameters are as follows: q_i is the query for node number i; u_i is the number of users who submitted query q_i; out_i is the number of users who submitted other queries after submitting query q_i; u_{ij} is the number of users who submitted queries q_i and q_j in that order; S is the set of all nodes; sa_i is the number of undefined queries that occur after submission of query q_i; sb_i is the number of undefined queries that occur before submission of query q_i; b_i is

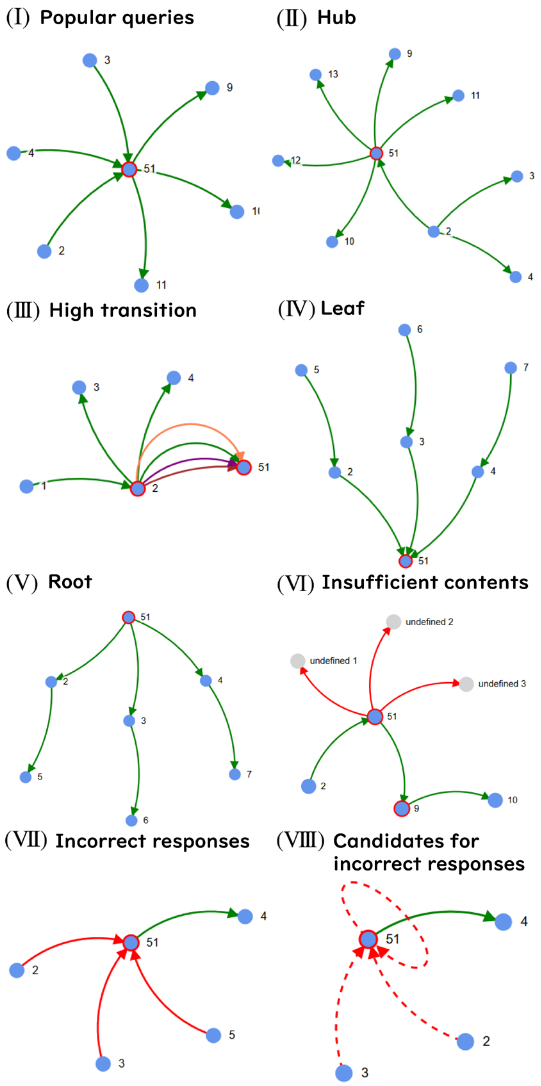


Figure 4. Examples of characteristic queries

Table 3. Expressions of characteristic queries for calculation

Types	Expression
(I)	u_i
(II)	out_i
(III)	$\frac{u_{ij}}{out_i}$
(IV)	$\sum_{j \in S} u_{ij}$
(V)	$\sum_{j \in S} u_{ji}$
(VI)	sa_i, sb_j
(VII)	$\sum_{i \in S} b_i$
(VIII)	$\sum_{i \in S} c_i$

the number of users who evaluated the responses of query q_i as bad; and c_i is the number of users who submitted query q_i, signaling it as a candidate for incorrect responses.

In (I), (III), and (IV), the scores were calculated by excluding the red edges (incorrect responses), where the red dotted edges of the same nodes (candidates for incorrect responses) were counted as one edge. Any type of characteristic node can be selected for analysis, and nodes of any higher score (i.e., (I)–(III), (VI)–(VIII)) or lower score (i.e., (IV), (V)) are circled in red, as shown in Fig. 4.

In cases where the number of queries is small, (III), (IV), and (V) will have standard values because the mean values of all the nodes are used as an initial value in the calculation. The standard values applied by each node type are as follows: (III), out_i; (IV) $\Sigma_{j\in S}\ u_{ji}$; and (V) $\Sigma_{j\in S}\ u_{ij}$. For example, if out_i in (III) is small, high transitions will result from calculations when out_i is greater than the standard value.

The edge distance from the characteristic queries was set to a value of three in order to define the visualization range that displays only the information surrounding the characteristic queries. This default value can be changed to suit different visualization requirements.

3.4 Development and System Scalability

Network graphs and functions supporting the analysis were developed using the software listed in Table 4. The virtualization software ESXi was used for development efficiency. The nodes and edges were displayed as a force-directed graph using D3.js[24]. Functions for filtering and identifying characteristic queries were developed using PHP and HTML. General dialogue systems contain databases to record information, such as user profiles and dialogue histories, and network graphs are displayed by referring to this database. Node ID; system queries; number of queries; queries with unsuitable responses; number of users from/to other nodes, user ID; and user queries, including those with incorrect responses and candidates for these were displayed in pop-up windows when the nodes were clicked. The number of users, nodes, gray nodes, and queries were displayed in the top-left corner of the window displaying the network graphs. An option menu that enabled setting changes (e.g., color of the nodes, number of characteristic nodes displayed, and length of edges) was also added.

Table 4. Software used for development

Types of software	Name
Virtualization software	VMware ESXi 6.7
Operating system	CentOS 7.6
Visualization tool	D3.js
Programming language	PHP, JavaScript
Markup language	HTML

The change in the usage statistics (Table 1) based on the dialogue system used necessitates a change in the visualization objects, filtering conditions (Table 2), and characteristic queries (Fig. 4). This is possible owing to the customization flexibility of the proposed method. In network graphs, it is possible to redesign visualizations with slight customization (e.g., changing the colors of the nodes and edges). In addition, the proposed system can easily enable, disable, and add new functions because the functions for filtering and identifying characteristic queries are modularized.

4. Evaluation

The proposed method was evaluated using real data from the T-talk. The purpose of the evaluation was to demonstrate the types of analytical results that could be obtained. Specifically, three administrators of T-talk analyzed real data using the proposed method. The administrators included one faculty member and two

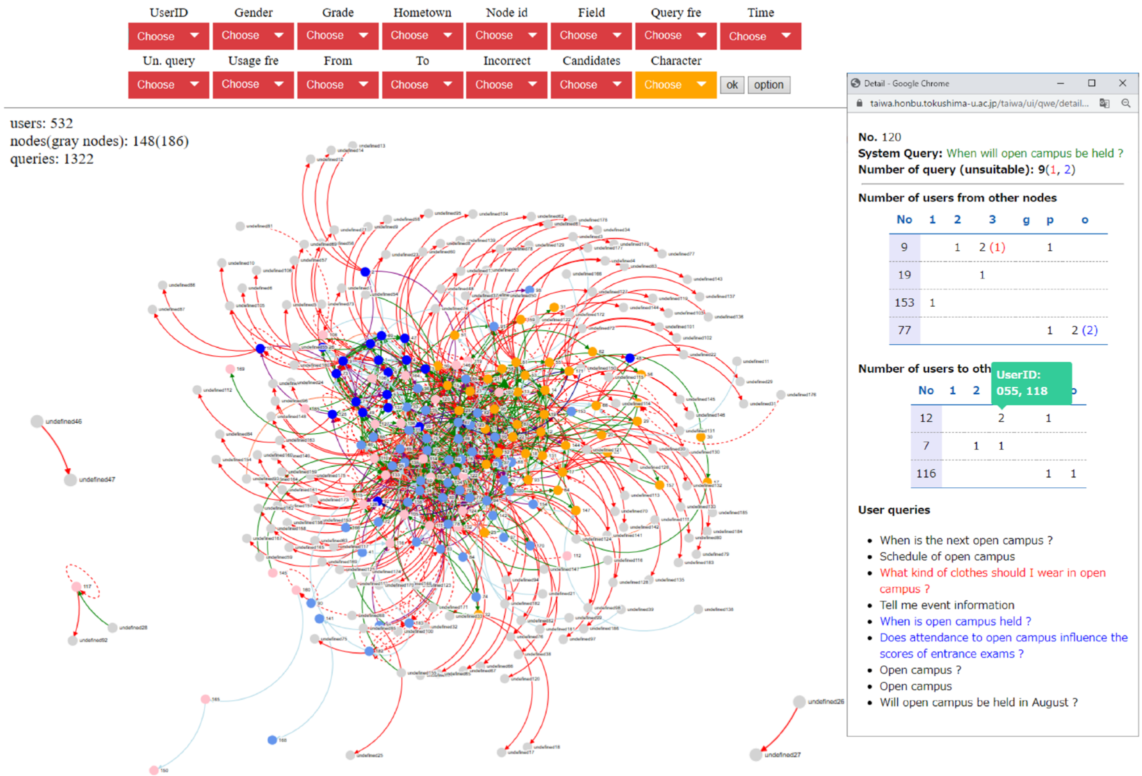


Figure 5. Network graph using real data

office workers, and the analysis (e.g., discovered content, discussion, and improvement plans) was performed for 1 h. Furthermore, the proposed method was qualitatively compared with general analysis methods.

4.1 Dataset

Tokushima University introduced T-talk in April 2019, and user profiles and dialogue histories were stored in a database. Real data collected between April 01, 2019 and March 31, 2020, were evaluated. Fig. 5[4] displays network graphs using real data. The numbers of users, nodes (gray nodes), and queries were 532, 148 (186), and 1,322, respectively. The top of the image displays functions for filtering and detecting characteristic queries, and the right window provides information on the selected nodes. Table 5 lists the numbers of users and queries for each grade. The colors of the nodes represent the following items: pink, exams; cornflower blue, university features; yellow, lifestyle; and blue, other. The colors of the edges are defined as follows: brown, first grade of high school; purple, second grade of high school; green, third grade of high school; navy, recent graduates; orange, parents; and light blue: other.

Table 5. Number of users and queries for each grade

Grades	Users	Queries
Third grade	237	625
Second grade	49	154
First grade	30	51
Graduates	9	24
Parents	81	142
Other	126	326
Total	532	1,322

[4] Fig. 5 was processed to suit this study by enlarging parts of the image and translating Japanese text to English.

4.2 Analysis Results

Supplementary explanations (e.g., background) were added to the administrators' analysis results to improve intelligibility, where necessary. To unify the representations, those of the analysis results were modified to achieve a uniform representation while retaining the content. The analysis results and comments are presented below.

With the exception of the red edges, many green edges are displayed centrally in Fig. 5, which indicates that the third-grade students submitted various queries. Most nodes resulted in clusters in each field, indicating that users tended to submit queries in each field.

A cluster of cornflower-blue nodes is located centrally, indicating a tendency for many users to submit queries regarding university features. The distribution of nodes (queries) was as follows: pink nodes, 36 (335); cornflower-blue nodes, 55 (503); yellow nodes, 43 (325); and blue nodes, 20 (159). This data reveals that the number of nodes and queries related to university features is the largest.

Fig. 5 shows a disconnected network graph that includes small components. Most of these small components included only gray nodes. Since T-talk could not respond to the users' queries in these cases, the trust of these users was undermined, and these users may have stopped submitting queries. Providing additional information (e.g., query ranking) may be effective in encouraging such users to continue submitting queries.

A query about open campus[5] was selected as a popular query between July 01, 2019, and August 31, 2019. This is because the open campus of Tokushima University is held every August. However, queries about the open campus were indicated as having insufficient content. Therefore, responses may have lacked the sort of information expected by users. Undefined queries included "What kind of clothes should I wear?" "When will application acceptance start?" "Does attendance influence the scores of entrance exams?" among others. Since information about open campus attracted significant attention, enriching the response content is necessary.

[5] "Open campus" is a promotional admissions event that showcases the inside of universities and orientates participants to encourage and strengthen their interest in the universities.

A filter could be applied to selectively reveal the types of responses where the user allowed more than 180 s to pass between queries. It was found that users tended to exceed this time after obtaining one of the following three responses: information about exam schedules and venues, dormitories, and exam content. Some content in T-talk is difficult to formulate as a response owing to the large amount of information required. To solve this problem, links to websites or PDF files (e.g., examination guidelines and university brochures) are included in the system responses. The three aforementioned responses included links or PDF files, and users may have taken the time to read them before submitting another query. Furthermore, some users submitted similar queries repeatedly after obtaining responses containing linked files. Users may have avoided reading these files owing to difficulty accessing the target information and consequently may have tried to access the information by making similar queries. Thus, it was suggested that access to information should be supported by highlighting the table of contents and displaying the target information page.

Queries on the schedule of seminars, including open campuses, were selected as popular queries. Since participants can obtain various types of information in these seminars, many users may request schedules to join them. Users who had used T-talk two or more times also submitted queries regarding seminars. There is a possibility that these users developed an interest in Tokushima University through T-talk and may have submitted queries to obtain further information regarding the seminars.

Queries on the features of Tokushima University were selected as hubs. Users tended to continue making additional queries (e.g., study-abroad programs, types of clubs, and laboratories) after making these queries. Since this type of information increases users' interest in Tokushima University, encouraging users to submit these other queries by recommending them may be possible. Queries about dormitories and part-time jobs were selected under high-transition conditions. At Tokushima University, students experiencing economic strain would prioritize living in dormitories, and it is thus important for them to find part-time jobs to cover expenses. Additionally, various nodes around the selected nodes were related to lifestyle (e.g., apartment rent and percentage of living alone). To prevent users from overlooking information and to relieve anxiety,

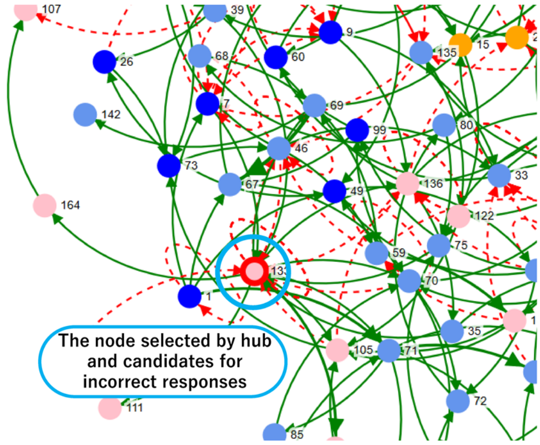


Figure 6. Part of the network graph, excluding undefined queries, filtered by the third grade of high school

responses that include communal information about lifestyle may be effective.

A query indicated by the leaf pattern, for example, was “How should I make a hotel reservation during exams?” This query is important for users who live far from the university. However, because its priority is low compared with other queries, this query may be the last one submitted.

Queries that users tended to submit first were analyzed as the root. A query that many users submitted (e.g., seminar schedules) was expected to be selected as an example. However, the greeting “hello” was selected, as various queries were submitted afterward. This could be due to the users encountering dialogue systems for the first time, who may have submitted such queries to see the reaction. To familiarize users with dialogue systems, incorporating a greeting and introduction by T-talk at the beginning of a conversation may be effective.

Fig. 6 shows part of the network graph, excluding undefined queries, filtered by third-grade high school students. A query about the exam for selected candidates (hereinafter called “special exam”) was a hub and also a candidate for incorrect responses. Queries (e.g., number of applicants, evaluation view, and acceptance standard) were submitted after obtaining special exam responses. Responses to such queries should be combined to decrease the users’ workload. Alternatively, there were incorrect responses (e.g., a cutoff point of the special exam and the number of special exams taken), and the same responses were provided owing to undefined queries and false retrieval. Thus, registering responses and synonyms is required to improve retrieval accuracy.

The filtering of some fields resulted in triangle-shaped nodes and edges, which indicated that users submitted the queries consecutively in any order. In the university feature fields, nodes that formed triangles were queries about laboratories, career paths after graduation, and acquirable qualifications. Their responses lead to a future increase of users, as they attracted high user interest. These queries showed a linkage relationship, and users may benefit from additional information provided as related information.

There were some comments from users who tested the proposed method. First, visual awareness was useful for exploratory analysis; however, textual information was more convenient for understanding the characteristic queries. Although the characteristic queries were highlighted in the network graphs, it was necessary to search for them, which may be a burden. Thus, if exploratory analysis is not required, it is necessary to display only the results as text information on a separate screen. Second, an understanding of query trends for each user desiring specific faculties is preferable. Filterable objects depend on collectible information from dialogue systems. Questionnaires can be implemented to obtain information before using dialogue systems, which would make it possible to expand the filterable objects. Third, it is desirable to freely combine characteristic queries. For example, by selecting the nodes of popular queries and roots, the most important queries were determined. Since analysis purposes depend on the users, functions that flexibly combine various conditions are required.

Through an exploratory analysis of the visualized usage statistics, much beneficial information was discovered, including some unexpected results. Notably, there were complex discoveries (e.g., shapes such as triangles). Since this information can be discovered easily by analysis support, registering new types of queries as characteristic queries and designing other support functions for finding them are required.

4.3 Comparison with General Analysis Methods

The proposed method was qualitatively compared with general analysis methods. General analysis methods provide analysis results and are easily applied. However, the content of the results is limited by predefined conditions, and methods to support further analysis of logs and other information sources freely are not provided. It is possible to directly analyze logs and other information sources using SQL command lines and spreadsheet software.

The administrators analyzed the real data using SQL commands or spreadsheet software under the same conditions as in the evaluation of the proposed method. One administrator used SQL commands, whereas the others used spreadsheet software. Some information (such as the most frequently submitted queries in each grade and queries resulting in incorrect responses) was easily discovered. However, it is a labor-intensive task to analyze the information in relation to the target information (such as query content before and after submitting queries of dormitories) owing to the need to combine multiple conditions and analyze data from various perspectives. In addition, it was difficult to analyze information that could not be discovered by statistical analysis (such as information gained by observing triangle-shaped nodes and edges, as described in Section 4.2). This was because the analysis targets were determined intentionally, and unintentional discoveries were not expected.

In the proposed method (Section 4.2), by focusing on visualized objects (e.g., all usage statistics and peripheral information on characteristic queries, clusters, disconnected networks with only gray nodes, and edge shapes), the overall query trends, presence of early leavers owing to undefined queries, factors causing users to submit similar queries repeatedly, responses that should be unified and enriched, candidates for incorrect responses, and queries with a linkage relationship were discovered. It is noteworthy that intuitively discovering characteristic objects and starting the analysis from this point contributed to a more detailed analysis. Intuitive awareness through visualization facilitated the exploratory analysis, and the support functions contributed to various discoveries. Thus, the proposed method was effective in discovering both expected and unexpected beneficial information.

However, the comparison carried out in this study did not quantitatively evaluate the proposed method. By comparing the retrieval time, amount of analysis results, types of information discovered, and other results, the quantitative effectiveness can be clarified. Other comparison points include determining the amount of information that can be collected for implementing recommendation functions and retraining dialogue models.

5. Conclusions

This study proposes a visualization method to enable exploratory analysis of the usage statistics of dialogue systems in order to discover beneficial information. Specifically, usage statistics were visualized as network graphs displaying queries and their transitions, and functions to extract particular visualized objects and identifying characteristic queries were implemented to support the analysis. The results of the evaluation of real data of the dialogue system at Tokushima University showed that the proposed method could determine various trends and user behaviors, such as overall query trends, the presence of early leavers owing to undefined queries, factors causing users to submit similar queries repeatedly, responses that should be unified and enriched, candidates for incorrect responses, and queries with a linkage relationship. Visualization using network graphs enables easy and intuitive analysis, allowing administrators to identify beneficial information. Thus, dialogue systems can further support the collection of information for users by implementing recommendation functions, retraining dialogue models, and other improvements based on the analysis results.

The proposed method was introduced at Tokushima University to support the analysis of T-talk system operation. Before its introduction, it was difficult to analyze usage statistics in detail, and the analyzable data range was limited because the information needed for the analysis (e.g., user profiles and dialogue histories) was registered in a database as text data. The analyzable range was extended by the proposed method, and various pieces of information, such as the analysis results presented in Section 4.2, could be easily obtained. Therefore, the proposed method contributes to the operation of dialogue systems.

The proposed method targets general task-oriented dialogue systems, and the usage statistics of other, similar dialogue systems can be analyzed, resulting in discoveries different to those found in the current evaluation. Tokushima University introduced another dialogue system for university students in 2020. By analyzing the usage statistics of this updated dialogue system using the proposed method, smooth system operation can be expected.

Further research is required to extend the proposed method to support further analysis. Expanding the filtering targets, registering new types of characteristic queries, and designing other support functions can be considered for implementation. Additionally, some users may not know what queries to submit, while others may submit their queries without accessing all reference information; therefore, methods to support users in accessing this information are required. One solution is to provide recommendations for popular hub queries. In addition, ranking algorithms (e.g., PageRank and HITS) for analyzing networks can be applied to network graphs for recommendations. Further studies are needed to design analytical methods for these recommendations.

Acknowledgments

This work was supported by a JSPS KAKENHI Grant Number JP19K14317.

References

[1] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, “Chatbot for university related FAQs,” in *Int. Conf. Adv. Comput. Commun. Inform., ICACCI 2017*, vol. 2017-January, pp. 1525–1530, 2017.

[2] Y. Sumikawa, M. Fujiyoshi, H. Hatakeyama, and M. Nagai, “An FAQ dataset for e-learning system used on a Japanese university,” *Data Brief*, vol. 25, p. 104001, 2019.

[3] Y. Seki and Y. Ueno, “A recommendation-type dialogue system responding to potential requests in consideration of personal attributes,” (in Japanese), *Jpn. J. Educ. Technol.*, vol. 45, no. 1, pp. 103–112, 2021.

[4] Y. Seki, “Visualization for analyzing usage status from dialogue systems,” in *Proc. 2020 IEEE 44th Ann. Comput., Software, and Applications Conf. (COMPSAC) Workshop 2020 Soc. Netw. Comput. and Big Data*, pp. 1701–1706, 2020.

[5] H. Chen, X. Liu, D. Yin, and J. Tang, “A survey on dialogue systems: Recent advances and new frontiers,” *SIGKDD Explor.*, vol. 19, no. 2, pp. 25–35, 2017.

[6] K. Yoshino, S. Mori, and T. Kawahara, “Spoken dialogue system based on information extraction and presentation using similarity of predicate argument structures,” (in Japanese), *IPSJ J.*, vol. 52, no. 12, pp. 3386–3397, 2011.

[7] M. Mizukami, *et al.*, “Example based dialogue system, response selection, user adaptation, satisfaction prediction,”(in Japanese), *Trans. Jpn. Soc. Artif. Intell.*, vol. 31, no. 1, pp. 1–12, 2016.

[8] Z. Yang, G.-A. Levow, and H. Meng, “Predicting user satisfaction in spoken dialog system evaluation with collaborative filtering,” *IEEE J. Sel. Top. Signal Process*, vol. 6, no. 8, pp. 971–981, 2012.

[9] G. K. Hoon, L. J. Yong, and G. K. Yang, “Interfacing chatbot with data retrieval and analytics queries for decision making,” in *RITA 2018. Lecture Notes in Mech. Eng.*, A. P. P. Abdul Majeed, J. Mat-Jizat, M. Hassan, Z. Taha, H. Choi, and J. Kim, Eds., Singapore, Springer, 2020, pp. 385–394.

[10] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, “Chatbot: An education support system for student,” *Int. Symp. Cyberspace Safety and Security*, pp. 291–302, 2018.

[11] S. N. Gilani, *et al.*, “Multimodal dialogue management for multiparty interaction with infants,” in *Proc. 20th ACM Int. Conf. Multimodal Intaract.*, pp. 5–13, 2018.

[12] B. Liu and S. S. Sundar, “Should machines express sympathy and empathy?: Experiments with a health advice chatbot,” *Cyberpsychol. Behav. Soc. Netw.*, vol. 21, no. 10, pp. 625–636, 2018.

[13] C.-H. Li, K. Chen, and Y.-J. Chang, “When there is no progress with a task-oriented chatbot: A conversation analysis,” in *Proc. 21st Int. Conf. Hum. Comput. Interact. with Mobile Devices and Services*, 2019.

[14] M. Jain, P. Kumar, R. Kota, and S. N. Patel, “Evaluating and informing the design of chatbots,” in *Proc. 2018 Designing Interact. Syst. Conf.*, pp. 895–906, 2018.

[15] M.-C. Jenkins, R. Churchill, S. Cox, and D. Smith, “Analysis of user interaction with service oriented chatbot systems,” *Lect. Notes Comput. Sci.*, vol. 4552 LNCS, no. PART 3, pp. 76–83, 2007.

[16] E. Van den Broeck, B. Zarouali, and K. Poels, “Chatbot advertising effectiveness: When does the message get through?” *Comput. Hum. Behav.*, vol. 98, pp. 150–157, 2019.

[17] Y. Miyadera, S. Nakamura, S. Yokoyama, and T. Yaku, “Personal management and sharing of research infor-

mation using research information transition graphs," (in Japanese), *IEICE Trans. Inf. Syst.*, vol. 91, no. 3, pp. 639–653, 2008.

[18] K. Umeda, T. Yasuda, and S. Yokoi, "Proposal of a research information sharing system utilizing knowledge-memos," (in Japanese), *IPSJ J.*, vol. 42, no. 11, pp. 2562–2571, 2001.

[19] Y. Miyadera, A. Taji, K. Oyobe, S. Yokoyama, and H. Konya, "Algorithms for visualization of paper-relations based on the graph drawing problem," (in Japanese), *Trans. Inst. Electron. Inf. Commun. Eng.*, vol. D-I 87 no. 3, pp. 398–415, 2004.

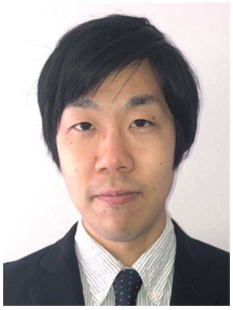
[20] C. L. Giles, K. D. Bollacker, and S. Lawrence, "CiteSeer: An automatic citation indexing system," in *Proc. ACM Int. Conf. Digital Libraries*, pp. 89–98, 1998.

[21] B. Shneiderman and A. Aris, "Network visualization by semantic substrates," *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 5, pp. 733–740, 2006.

[22] Q. Zhu, *et al.*, "ConvLab-2: An open-source toolkit for building, evaluating, and diagnosing dialogue systems," in *Proc. 58th Ann. Meeting Assoc. Computat. Linguistics*, pp. 142–149, 2020.

[23] H. Dybkjær and L. Dybkjær, "DialogDesigner: Tools support for dialogue model design and evaluation," *Lang. Resour. Eval.*, vol. 40, no. 1, pp. 87–107, 2020.

[24] M. Bostock, V. Ogievetsky, and J. Heer, "D3: Data-driven documents," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2301–2309, 2011.



Yosuke Seki received his Ph.D. from Tokushima University in 2021. He is an associate professor at the Research Center for Higher Education at Tokushima University. He graduated from the Department of Science and Technology at Konan University. After several years working as a systems engineer, he completed a Ph.D. program without a dissertation in Systems Innovation Engineering at Tokushima University. His research interests include dialogue systems and the entrance examinations of universities. He is a member of JSET, JSAI, IPSJ, and IEICE.