AN ASSESSMENT METHOD OF INTERVIEWS FOR MODELING ENGINEERING DESIGN PROCESS

Rempei Nishida[1], Yutaka Nomaguchi[1], Kikuo Fujita[1], Le Anh Hoang[2], Takahiro Omori[2], Yuji Sasaki[2], Yuji Mori[2], Yasutada Nakagawa[2], Michinobu Inoue[2], and Daisuke Hosokawa[2]
[1] Department of Mechanical Engineering, Osaka University
[2] Corporate Manufacturing Engineering Center, Toshiba Corporation

ABSTRACT
The design and development process that produces innovative products and services has been considered an important organizational capability. Modeling of the design and development process is generally conducted through interviews with experienced engineers. However, the success of such modeling depends heavily on the ability of the interviewer to converse with experienced engineers, and it is not clear what kind of conversation is effective for modeling. This has become a bottleneck in knowledge sharing and reuse within organizations. To solve this problem, this study aims to establish a method for clarifying and assessing the conversation structure of interviews for modeling the engineering design processes. The proposed method consists of two parts: a method for analyzing the conversation structure using an integrated three-layered process model, which the group of Osaka University has developed to represent the system design process, and a method for evaluating the relationship between questions and answers in the conversation based on text similarity measurement based on a large language model. As a case study, we take up an example of modeling a solder joint design and evaluation process carried out in a manufacturing company. We apply the proposed method to the analysis and assessment of the modeling process conducted by an expert interviewer and a novice interviewer for an engineer. We discuss the validity of the proposed method under the results of the case study.

Keywords: Design process, Modeling, Engineering knowledge, Knowledge management, Text similarity, Conversation assessment

1 INTRODUCTION

With the diversification of customer needs and the globalization of product development, there is a growing need for rapid design development based on more information. About 20% of a designer's time is spent searching and absorbing information [1], and the importance of capturing and managing engineering knowledge is increasing. In this context, there is an effort to model processes [1] and standardize them to share and reuse the experimental knowledge of engineers. These are efforts to accumulate the knowledge and know-how of experts in the form of procedure manuals and best practices collection, etc., and to extract and utilize the necessary information. Modeling such a design and development process is generally done through interviews with skilled engineers. However, effective knowledge-capturing methods through interviews have not been established. This study proposes a method to clarify and evaluate the conversation structure in interview-style modeling of the engineering design process toward effective knowledge capture.

The rest of this paper consists of four sections. Section 2 addresses research issues and describes our approach to clarifying the interview structure and its evaluation method. Section 3 explains the proposed method. Section 4 describes the application of solder joint design to modeling as a case study. Section 5 examines and discusses the proposed method, and finally, Section 6 summarizes this study.

2 RESEARCH ISSUES AND OUR APPROACHES

2.1 Prior Works on Interview Methods

Interviews are positioned as a representative method of data collection in qualitative research [2,14], and play an essential role in ethnography, which describes people and cultures by observing the actual activities of groups in fields such as sociology and psychology. The data collected in interviews are transcribed and then structured through the KJ method [3] and data coding to make sense. Such conversational analysis is based on the analyst's expertise and may contain subjective elements. There are still open issues in evaluating the interview quality quantitatively.

In the context of design, data collection for human-centered design and interviews with designers who have used generative design tools to analyze the impact of the tools on the design process and designers [4] are proposed. However, since engineering activities are under time constraints, it is necessary to develop and analyze efficient interview methods and effectively utilize the know-how of expert interviewers. Therefore, we objectively analyze interviews to capture engineering knowledge from engineers efficiently.

2.2 Conversation Analysis in Interview

Capturing engineering knowledge through interviews with engineers is conducted by the interviewer during the design and development process, or through interviews and workshops after a series of design and development has been completed. Although the former allows the engineers to make decisions in the design and development process in real-time, it may burden the engineers during the work and interfere with the work. In the latter case, formalized knowledge that is highly reusable can be acquired by taking the time to interview engineers after the design is completed. On the other hand, in the limited time of the interview, engineers often cannot recall their judgments and may not be able to record sufficient knowledge.

This study focuses on the latter approach and analyzes the problem that the success or failure of the interview in the latter method depends highly on the interviewer's skill. We compare two cases: an expert interviewer and a novice interviewer. Until now, various studies have been conducted on analyzing dialogue or communication as a dialogue system. As the origin of this, there is ELIZA [5], a system that repeats and asks questions of the person's utterances based on predefined rules, and SHRDLU [6], which aims to achieve tasks using frames. In recent years, significant research has focused on dialogue systems utilizing deep learning, such as updating dialogue states based on dialogue history [7] and incorporating external knowledge to avoid generic responses and ensure consistency [8]. These dialogue systems depend on the execution of specific tasks or are targeted at everyday conversation, such as small talk, and no studies have analyzed the modeling of the engineering design process so far. This study proposes a method for analyzing the conversation structure using an integrated three-layered process model [9,10] and evaluating communications using text similarity based on a large language model.

3 PROPOSED METHOD OF INTERVIEW ASSESSMENT

Figure 1 shows an overview of the proposed method. It consists of two parts. The first is a method to clarify the argument structure based on the process model and design operation template (the bottom right of Figure 1). The second is the evaluation of an interview by calculating dialogue distances based on the distributed representation.

3.1 Model of Engineering Design Process

There are various representation methods for the engineering design process, such as IDEF0 [11], Design Structure Matrix (DSM) [12], and so on. Although their details vary, they commonly see the engineering design process as a system. It is a formal, structured, and interdisciplinary methodology that integrates technical and human elements to address complex engineering problems. It is organized hierarchically and focuses on whole-system, whole-life dimensions, ensuring the required quality, cost, and duration.

We use a simple systemic model following the premises. The engineering design process hierarchically consists of several processes. Subprocesses are interrelated through the inputs and outputs of deliverables. The process has attributes such as standard required duration, standard cost, and standard resources required to complete the process.

3.2 Clarifying the Conversation Structure of Interviews

The integrated three-layered process model has a mechanism to convert the log of design work acquired as design tool operations, videos, and audio into a representation of the discussion structure model through a template based on design operations that integrate several basic operations and consists of three levels of processes: action, model operation, and discussion [9,10]. In addition, to perform a detailed analysis of each remark, we define remark type (RT) for the interviewer and engineering expert's remarks. The conversation structure of the interview is clarified by classifying all remarks in the modeling process into RT and examining the frequency of their occurrences, time transition, and sequence.

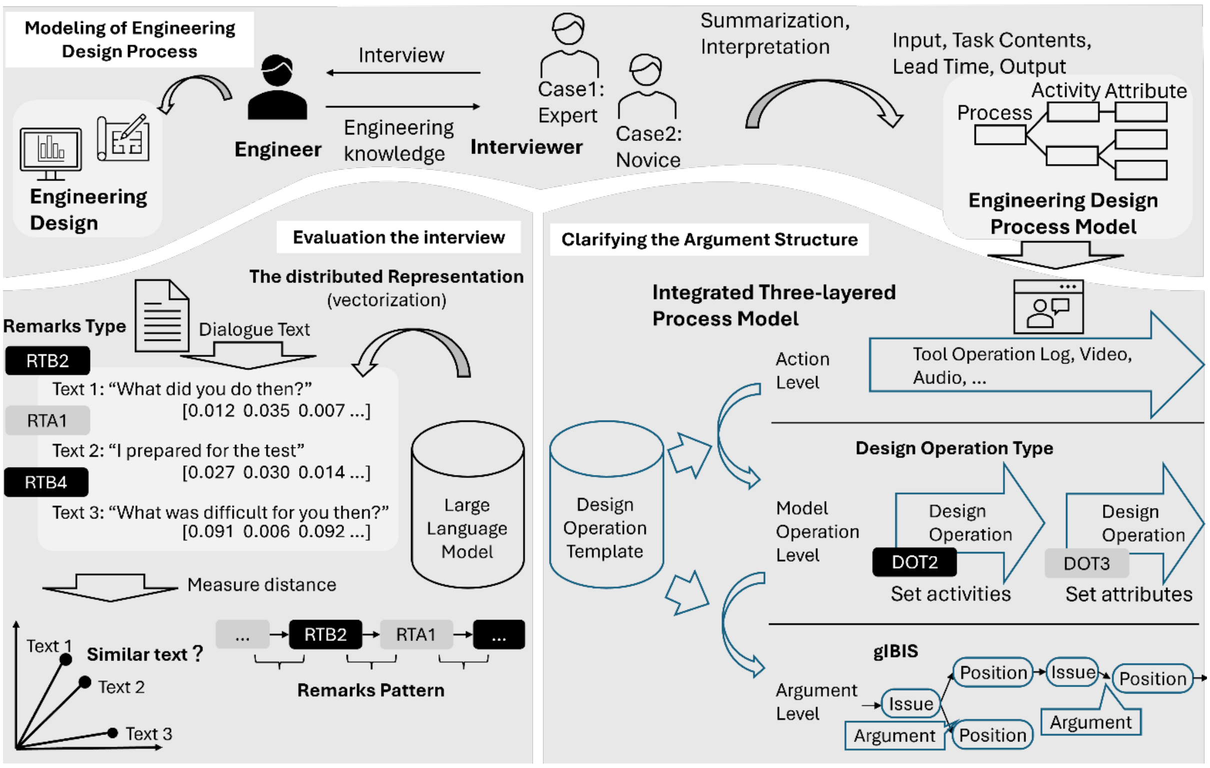

Figure 1. Overview of the proposed method

3.2.1 Description of Argument Structure

To describe an argument structure in the modeling process, Graphical Issue-Based Information Systems; gIBIS [13] are utilized. gIBIS is a knowledge representation method that records people's logical thought processes in a hierarchical structure according to time series, which corresponds to the discussion process level of the integrated three-layered process model. gIBIS consists of three nodes called, "Issue", "Position", and "Argument". In gIBIS, multiple positions to solve an issue are lined up and represented as branches. The process for reaching the final solution proposal and the background of the modeling results are recorded in detail. In other words, the number of branches can explicitly express the point where the discussion has deepened and then the process model representation is modified. To obtain such a description of knowledge in gIBIS, it is important to create issue-and-position nodes formally. The design operation template [9,10] works to formally create their relationships.

3.2.2 Formalization of Argument Structure with Design Operation Template

Depending on the content and intent of the interviewer's operations in modeling, we define a total of five design operation types (DOT). We define the formalized relationship between position and issue node as a template. The definitions for each operation type are as follows:

- DOT1: Set a target process

This type is an operation to determine a target process for modeling. The interviewer decides which process knowledge to elicit from the engineer and standardize based on the company's policies and issues, as well as academic and technical significance.

- DOT2: Set activities

This operation type is an operation that refers to a target process or an activity and adds activities that compose the target process. To describe the general flow of the target process, the interviewer divides the process into some activities based on his or her experience and the engineer's answers acquired through them.

- DOT3: Set attributes

This is an operation to refer to an activity and add attributes that compose the activity. Attributes are common elements, regardless of activities.

- DOT4: Divide an attribute

This type is an operation to refer to an attribute and add its details.

- DOT5: Set a value of an attribute

This operation refers to an attribute and adds a specific value.

3.2.3 Remarks Type for Interviewer

We have extended and redefined the categories of interviewer questions defined by Brinkmann and Kvale [14]. Questions to acquire knowledge through conversation (from RTB1 to RTB5) and questions about the progress of conversations (from RTB6 to RTB9) are defined. The definitions for each RT for the interviewer are shown in Table 1.

Table 1. Definition of Remarks type for interviewer

label	Remarks type	Details
RTB1	Follow-up remarks	Follow-up remarks and reactions to the engineer’s answers, such as repetition of responses and phasing.
RTB2	Specificity questions	Clear questions that encourage the engineer to answer, e.g., “What did you do then?”
RTB3	Interpretive questions	Questions to confirm the validity of what the engineer said and the interviewer’s interpretation, e.g., “Is this what you mean?”
RTB4	Exploratory questions	Questions to ask when the interviewer wants to learn more about what the engineer said, e.g., “Can you tell me a little more?”
RTB5	Direct questions	Questions that are implied by responses and conversation flow.
RTB6	Introductory questions	Questions to present a topic and start a conversation.
RTB7	Survey questions	Questions to see if the engineer has left anything unsaid.
RTB8	Framed remarks	Questions to move on to the next topic when the interviewer feels that he or she has received sufficient answers from the engineer.
RTB9	Advice	Advice such as when the engineer is struggling to answer, or the interviewer has an idea to respond.

3.2.4 Remarks Type for Engineering Expert

In the interview, there are questions with a large amount of room for the engineer to decide the content of the answer and questions with a predetermined answer format and content. We define the former as open-ended answers; RTA1, and the latter as closed answers; RTA2. The definitions of RT for the engineering expert are shown in Table 2.

Table 2. Definition of Remarks type for engineering expert

label	Remarks type	Details
RTA1	Open-ended answers	An engineer’s response in their own words.
RTA2	Closed answers	Yes/No answers or those repeating the question’s wording
RTA3	Specificity remarks	Remarks about supplementing or concretizing his/her responses.
RTA4	Interpretive remarks	Remarks confirming the validity of their answer to the interviewer.
RTA5	Follow-up remarks	Reactions to the questions, such as repetition of them and phasing.
RTA6	Framed remarks	Remarks to move on to the next topic when the engineer feels they have sufficiently answered about the topic.
RTA7	Advice	Advice from the engineer to the interviewer, such as when the interviewer is struggling in the modeling process.
RTA8	Questions for questions	Remarks when the respondent does not understand the question or the intent of the question.
RTA9	Remarks to buy time	Remarks or reactions that the engineer makes time to respond.

3.3 Measuring Text Similarity Based on a Large Language Model

The key issues in capturing knowledge through interviews are determining whether the questions and answers are communicating and identifying effective questions. We use an embedding model based on a large language model to calculate the distributed representation of utterance text and evaluate the conversation in the interview by the distance between the texts. The distributed representation is based on the distributional hypothesis that the meaning of a word is formed by the surrounding words, i.e. context [15], and is based on the learning of many documents and converting the text into a vector representation. Therefore, the representation makes the text in a form that is easy to process by computer, and the more similar the meanings of the texts, the closer the distance between the two. Recent embedding models can handle longer and more complex meanings by converting text into higher-dimensional vectors. To calculate the distributed representation, we used OpenAI's Embedding model called, "text embedding-3-large" and converted each utterance text to a normalized vector of 3,072 dimensions. In interviews, a single question and the answer are principally repeated, but since the remarks are based on the context, it is not necessarily the response to the previous text. To address this, in this study, we take a moving average and weigh the previous k utterances to calculate the distance of the distributed representation considering the context In Eq. (1), w_i is weights that are defined in Eq. (2). $distance_{n,i}$ is the L2 distance of the distributed representation between nth text and ith text in utterance logs. dis_w represents the weighted distance to the nth text. In the case study of Section 4, we set $k = 5$.

$$dis_w = \sum_{i=n-k}^{n-1} w_i \times distance_{n,i} \tag{1}$$

$$w_i = \{i + 1 - (n - k)\}/k \tag{2}$$

4 CASE STUDY

To verify the validity and effectiveness of the proposed assessment method, we present a case study of the modeling of the design and evaluation process of solder joints [16,17] conducted by a company.

4.1 Overview of the Modeling Process

In the target process, the engineer designed a method for evaluating the fatigue durability of solder joints under combined loads such as thermal cycling, vibration, and shock. In the modeling, the engineer was interviewed, and the interviewer elicited his/her knowledge, interpreted it in the interviewer's way, and described it in a mind-mapping tool according to the process model stated in Section 3.1 that had been standardized in the company [18]. We captured a computer screen on which the interviewer handled and operated the tool and recorded the entire modeling process on video. For example, operations on the tool included dividing the extracted information into activities and describing their attribute values. We conducted two interview-style modeling cases. In the first case, the interviewer was an interview expert, and in the second case, the interviewer was a novice. According to the interviewers and interviewee, there was a clear difference in the results depending on the experience of the interviewers. In this study, we clarify the differences in experience through comparison and evaluate the conversation in modeling.

4.2 Captured the Modeling Log

After the modeling processes were completed, the transcription tool converted the audio data into text data that summarizes the timestamp, speaker, and speaking content at the time of utterance. We defined the brief remarks made by another speaker in the middle of the dialogue as noise and removed them from the text data in advance. Then, we manually extracted text data summarizing the tool operations from the screen recordings, and we classified the utterance logs and operation logs into RT and DOT.

4.2.1 Comparison of Two Cases

First, we qualitatively compare the process models created by the two cases. Figure 2 shows a part of a simplified version of the model. The interview expert decomposed the target process into six activities, e.g., setting the theme and referring to previous studies. However, the interview novice decomposed the target process into four activities, e.g., background and purpose of the study and research methods. Next, a comparison by modeling time revealed that the expert spent 113 minutes, and the novice spent 72 minutes. In addition, we compared the number of characters in spoken words, and the number of utterances turns. Comparing them, the number of utterance characters is roughly proportional to the interview time, but the expert interviewer case had three times more turns than the novice interviewer case. This means that the expert repeatedly exchanged short sentences.

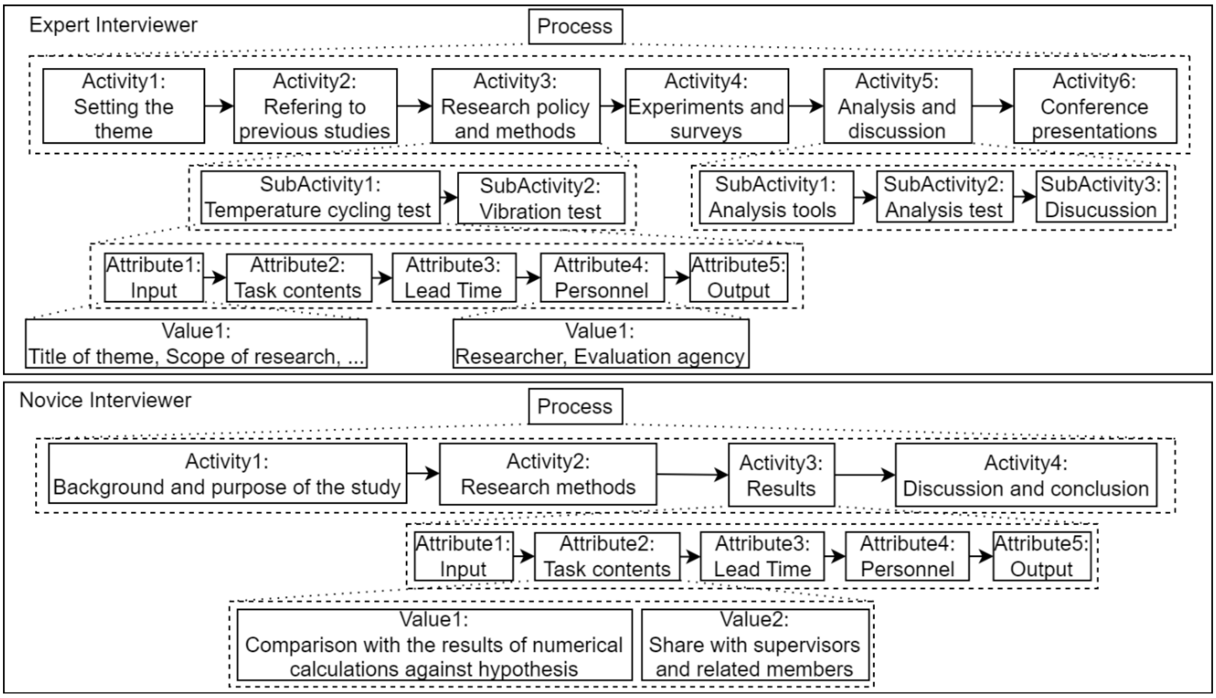

Figure 2. Overview of the model of the engineering design process in 2 interview cases. Subactivities, attributes, and attribute values are shown partially.

4.2.2 Comparison in Design Operation Type

We counted the number of DOTs. In the expert interviewer case, the number of DOT2: Set activities is larger than in the novice interviewer case, and the model is created hierarchically by dividing it into groups. The number of DOT5 is the largest for both cases among all DOTs, which shows that the description of attribute values is the most essential operation. For further analysis, Figure 3 shows the distribution of the time points where the operations were performed over the entire modeling time. Comparing DOT2 in two cases, the red bar (the interview expert) is widely distributed from time1 to time10, but the blue one (the interview novice) does not appear after time7. This result indicates that a novice interviewer finishes describing the upper levels of activity levels by the middle of the process and spends the rest of the time describing them in detail. Also, the wide distribution of red bars in all types shows that the expert is repeating the creation of the model across hierarchies.

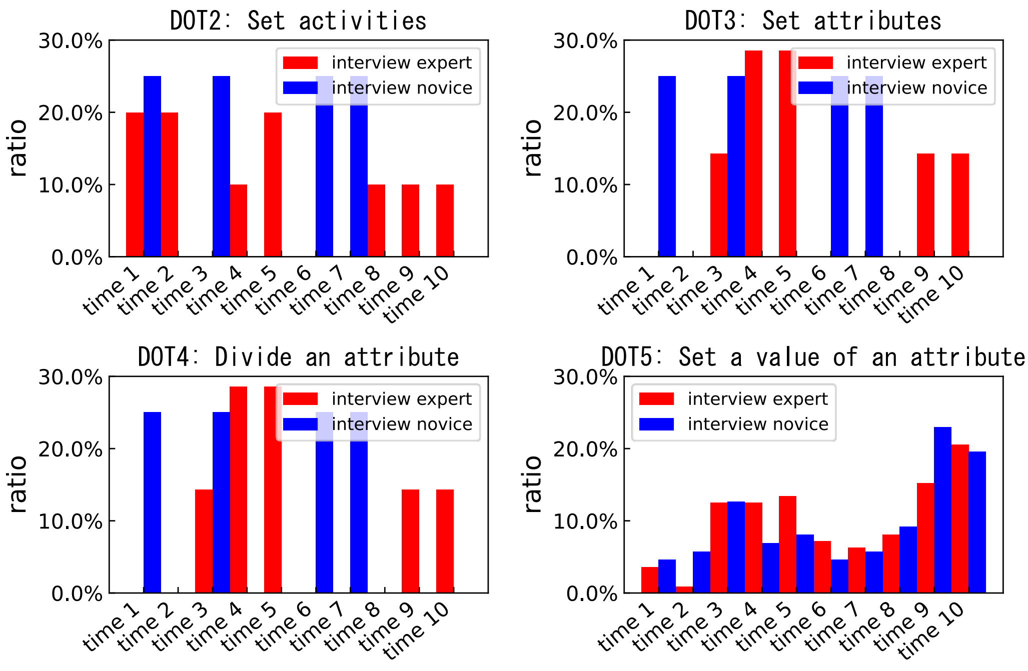

Figure 3. Appearance ratio of design operation type over time

4.2.3 Description of Argument Structure with gIBIS

We made the gIBIS representation of two cases based on the design operation template. Figure 4 shows a part of them. There were 581 nodes in the expert interviewer case and 27 nodes with the strikethrough red lines corresponding to the modified model representation. The novice case showed that there were 267 nodes, and the number of the strikethrough nodes was 11.

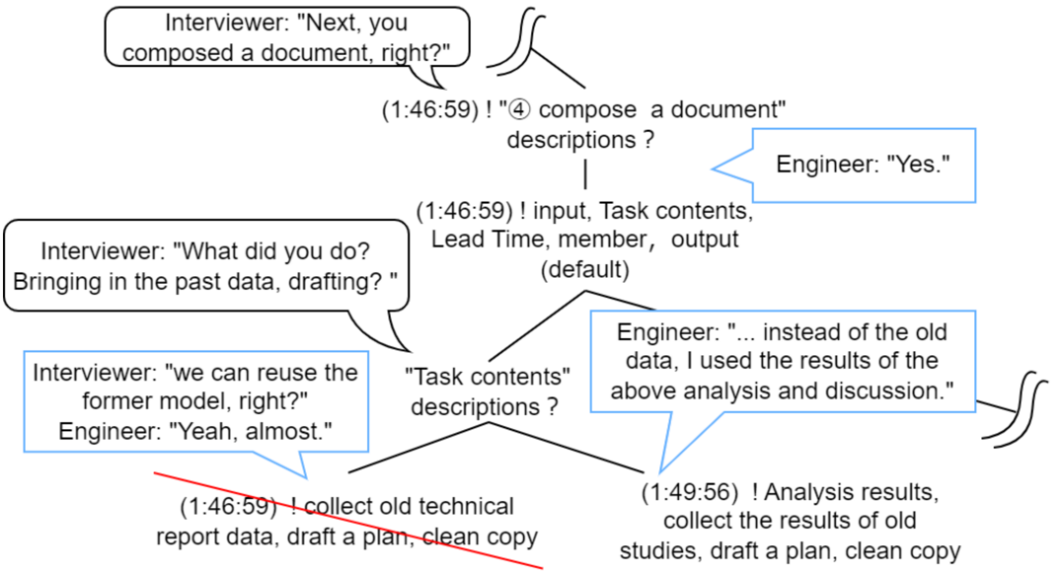

Figure 4. Argument structure representation in interview expert case with gIBIS

4.2.4 Comparison in Remarks Type

Next, we classified all remarks in the interview into RT, mainly in units of utterance turns. Table 3 and Table 4 show the breakdowns of the results. Both interviewers frequently used RTB1, RTB2, and RTB3, and the engineer frequently used RTA1, RTA2, RTA3, and RTA5. We examined the occurrence ratio of RTB2 and RTB3 over the entire interview time. It was found that the expert asked clear questions from the beginning of the modeling, and actively asked questions about the validity of the interpretation at the end of them. This result corresponds to the result in Section 4.2.2 and shows that the expert postpones asking deeper about the content of the answers and prioritizes grasping the general content.

Table 3. Breakdowns of remarks type in Qs.

label	Remarks type	expert	novice
RTB1	F/U Rmk	95	35
RTB2	Spec. Qs	64	21
RTB3	Int. Qs	73	18
RTB4	Expl. Qs	11	4
RTB5	Dir. Qs	12	3
RTB6	Intro. Qs	15	5
RTB7	Surv. Qs	4	0
RTB8	Frm. Rmk	6	1
RTB9	Advice	11	2

Table 4. Breakdowns of remarks type in Ans.

label	Remarks type	expert	novice
RTA1	Open-end. Ans	100	48
RTA2	Clsd. Ans	73	14
RTA3	Spec. Rmk	39	8
RTA4	Int. Rmk	5	1
RTA5	F/U Rmk	61	18
RTA6	Frm. Rmk	4	0
RTA7	Advice	3	4
RTA8	Qs for Qs	4	0
RTA9	Rmks. to B.T.	13	6

4.3 Text Similarity by the Distributed Representation

To consider the context, the texts up to five previous target remarks were weighted based on Eq. (1), and the distance between the remarks was calculated. As a result, the distances varied about from 0.90 to 1.30, with no clear differences between the two cases. Therefore, we examined in detail the texts with particularly large distances. For example, the distance was increased by the new topic of the presentation of the results since the previous statements had talked about the lead time. This result supports that the evaluation of interviews using the distributed distance is effective.

4.3.1 *The Distributed Representation Distance by Remarks Type*

Figure 5 shows the distributed representation distance by RT. The result shows that the distance between the text and the before of RTB6: Introductory question, is small for both cases. This is because a short reaction to the previous answer is inserted before asking the question of RTB6 to move on to the next topic. Compared to the two cases RTB5: Direct Questions in the expert case had more distance than the novice interviewer. The engineer was talking about sample creations for the experiment, but the interviewer suddenly asked a question about the type of solder. This increased the distance. In practice, the engineer seemed to have decided that there was no need to explain it. However, the interviewer's in-depth questions led to the process with which the interviewer and the engineer could communicate well. Therefore, direct questions can be effective even if the distance is large.

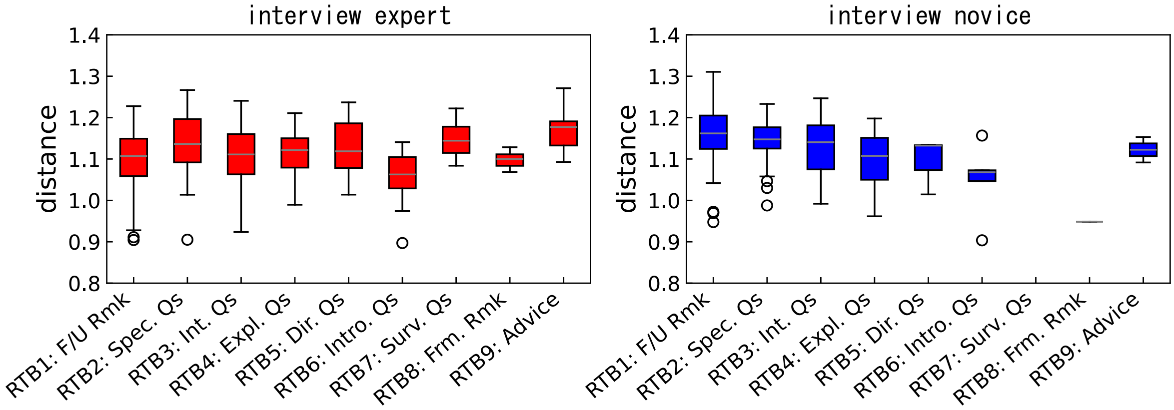

Figure 5. The distributed representation distance by remarks type

4.3.2 *Remarks Type and gIBIS*

We further consider the results by corresponding the defined RT to gIBIS. This allows us to determine useful statements that have been used in the modeling to elicit the rationale for modifying the process model representation. As mentioned in Section 4.2.3, the number of node modifications in gIBIS was 27 in the expert case. Of these, 14 discussion nodes described as the rationale for modifying the nodes, 11 RTA1: Open-ended answers, and 3 RTA3: Specifying remarks were adopted in the gIBIS description. In other words, in the process modeling, the basis for node modification is mainly based on these two types of statements, and it can be said that the questions that elicit these two types are important in the interview. Therefore, we examined the proportion of RT before RTA1 and RTA3. As a result, RTB1, RTB2, and RTB3 account for most of the questions before both RTA1 and RTA3. In other words, follow-up, specific, and interpretive questions were important remarks for modifying the process model.

5. DISCUSSION

The above case study evaluated the dialogue in modeling the engineering design process based on the gIBIS and the text distance. It was demonstrated that the interview can be objectively evaluated with the proposed method. Further, the proposed method clarified the difference in questioning between expert interviewers and novices.

The distributed representation of each text was determined using the interviewer and engineer's utterance turns as a single unit. By setting RT for each utterance turn, it was possible to correspond the distributed representation with RT. However, the length of the texts differed greatly in some places, causing outliers in the distance. It may be necessary to calculate the distance in units of sentences.

As shown in Tables 3 and 4, RTA7, RTB6, and RTB8 appeared in the expert case, but not in the novice case. The utterances related to these types may not directly contribute to the modeling, but they are important for facilitating communication.

Thus, the proposed method made it possible to quantitatively evaluate the sensory differences between the two cases felt by the interviewers and the engineers. If we use this method to analyze more cases, we can expect to establish an effective interview method.

6. CONCLUSION

In this study, we proposed a method for clarifying the conversation structure and evaluating them in an interview to elicit engineering knowledge. By examining the number of occurrences and time series of

DOT and RT, the difference between expert interviewer and novice could be clarified, and interview methods could be evaluated. We also showed that it is possible to analyze the interviews by describing the argument structure with gIBIS utilizing the design operation template and calculating the distance for all the statements in the interview.

REFERENCES

[1] D. Baxter, J. Gao, K. Case, J. Harding, B. Young, S. Cochrane, S. Dani, An engineering design knowledge reuse methodology using process modeling, *Res. Eng. Des.*, 2007, 18, pp. 37-48.

[2] Angrosino, M., *Doing Ethnographic and Observational Research,* 2007 (Sage Publications, Thousand Oaks).

[3] Kawakita, J. *The Way of Thinking for Creative Development*, 1967 (Chuukousinnsyo, Tokyo) (In Japanese).

[4] Jana I. Saadi, Maria C. Yang, Generative Design: Reframing the Role of the Designer in Early-Stage Design Process, *J. Mech. Des.*, 2023, 145(4), 041411.

[5] Weizenbaum J., ELIZA-A computer program for the study of natural language communication between man and machine, *Commun. ACM.*, 1966, 9(1), pp. 36-45

[6] Winograd T., Understanding natural language, *Cognitive Psychol.*, 1972, 3(1), pp.1-191

[7] C. Wu, et al., *Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems*, 2019, arXiv:1905.08743.

[8] Sixing Wu, Ying Li, Dawei Zhang, Yang Zhou, and Zhounghai Wu, Diverse and Informative Dialogue Generation with Context-Specific Commonsense Knowledge Awareness, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, online, July 2020, pp.5811-5820

[9] Nomaguchi Y., Fujita K., Knowledge Representation Framework for Interactive Capture and Management of Reflection Process in Product Concepts Development, *Adv. Eng. Inform.*, 2013, 27(4), pp. 537-554.

[10] Nomaguchi Y., Morikawa H., Nishida R., and Fujita K., A Framework of Real-time Knowledge Capture and Formalization for Model-based Design with Spoken Annotation and Design Operations, *J. Comput. Inf. Sci. Eng.,* 2024. Accepted.

[11] ISO, *ISO/IEC/IEEE 31320-1:2012 Information Technology - Modeling Languages - Part 1: Syntax and Semantics for IDEF0*, 2012.

[12] Eppinger, S. D. and Browning, T. R., *Design Structure Matrix Methods and Applications*, 2012 (The MIT Press, Cambridge).

[13] Conklin J. and Begeman M.L., gIBIS: A Hypertext Tool for Exploratory Policy Discussion, *ACM T. on Inform. Syst.*, 1988, 6(4), pp. 303-331.

[14] Brinkmann S. and Kvale S., *Interviews: Learning the Craft of Qualitative Research Interviewing, 3rd ed.*, 2015 (Sage Publications, Thousand Oaks).

[15] Mikolov, T., *et al.*, *Efficient estimation of word representations in vector space*, 2013, arXiv:1301.3781.

[16] Omori T., Hirohata K., Mukai M., Takahashi H., Ogawa H., Assessment of Soder Joint Durability Under Combined Loading, *In the JIEP 23rd Annual Meeting*, 23(0), Yokohama, March 2009, pp.190-191 (In Japanese).

[17] Omori T., Hirohata K., Mukai M., Takahashi H., Ogawa H., Effect of Initial Temperature Cycle on BGA Solder Joint Durability Against Vibrational Loading. *In JSME 22nd Computational Mechanics Conference*, 2009.22(0), Kanazawa, October 2009, pp. 722–723 (In Japanese).

[18] Yoshida S., Sano K., Murao R., Database Realizing Efficient Use of Accumulated Expertise Converted into Explicit Knowledge to Improve Productivity, *Toshiba Review*, 2020, 75(1), pp. 30-33 (In Japanese).