A Generation Method for the Discussion Process Model during Research Progress Using Transitions of Dialog Acts

Yoko Nishihara
College of Information Science and Engineering
Ritsumeikan University
Shiga, Japan
nisihara@fc.ritsumei.ac.jp

Seiya Tsuji
College of Information Science and Engineering
Ritsumeikan University
Shiga, Japan

Wataru Sunayama
School of Engineering
The University of Shiga Prefecture
Shiga, Japan

Ryosuke Yamanishi
Faculty of Informatics
Kansai University
Osaka, Japan

Shiho Imashiro
Institute for Organizational Behavior Research
Recruit Management Solutions Co., Ltd.
Tokyo, Japan

***Abstract*—People in business and academic fields work in cooperation rather than alone. They may discuss their progress with others, like co-workers and supervisors, to help them obtain the best results, and sometimes, they may feel that such discussions are not conducted well. However, people do not evaluate the quality of each discussion on every occasion because it is tough work for them, and they usually do not have enough time for that. In the process of evaluating discussions, people might look back on their discussions and make a plan to have an improved discussion next time. This paper proposes a generation method for a discussion process model during research progress. First, discussions are recorded to generate transcripts in which each line has a speaker name and his/her utterance. Then, the transcripts are classified manually into high-quality and low-quality discussion groups. Next, dialog acts are assigned to utterances as labels. The labels of dialog acts are originally designed for discussion analysis for research progress. After the labeling, transitions of dialog acts with a high appearance rate are extracted. The transitions are connected if the same dialog act is in both transitions to make a network of dialog acts. The network represents a model of the discussion process. A model for a high-quality discussion group is compared with a model for a low-quality discussion group. By investigating dialog acts and transitions found only in one group, suggestions for low-quality discussions to high-quality discussions would be found. We used discussions between a supervisor and a student who was studying for a degree at a university. We generated models for high-quality and low-quality discussion groups by the proposed method and revealed suggestions for low-quality discussions to high-quality discussions. Our contributions are summarized in two points: (1) we proposed a new method to generate models that represent the discussion process, and (2) we found suggestions for low-quality discussions to high-quality discussions using the models obtained.**

***Index Terms*—Discussion Analysis; Discussion Mining; Discussion Improvement; Transitions of Dialog Acts**

I. Introduction

People in business and academic fields work in cooperation rather than alone. They must discuss their working progress with others, like co-workers and supervisors, to help them obtain the best results. Such people can exchange their opinions and advise each other. These discussions make people not only understand what others think but also ensure that members of the team agree with their work.

Discussions are sometimes not conducted well. This happens because discussions have a time limit, and people often fail to arrive at a common understanding because of a difference in their thinking styles. However, people do not evaluate the quality of each discussion every time because it is tough work for them; they usually do not have enough time for that. In the process of evaluating discussions, people might look back on their discussions and make a plan to have an improved discussion next time. Discussion summarization utilizing the text and acoustic information of the discussion might be a support for the evaluation. However, few of the current summarization methods could detect problems in discussions or suggestions for those.

This paper proposes a generation method for a discussion process model during research progress to find suggestions for low-quality discussions to high-quality discussions. The proposed method divides discussions into two groups: high-quality and low-quality discussion groups. The method makes models for the discussion process in each group using transitions of dialog acts. By comparing the two models, discussants can find suggestions for low-quality discussions to high-quality discussions. The method uses discussions between a supervisor and a student preparing for a graduation thesis. Note that we define a high-quality discussion to be a discussion in which both the student and his/her corresponding supervisor understand their research progress. We believe that transitions of dialog acts represent the quality of discussion. In our previous paper [1], the basic method was proposed and preliminary analysis results were discussed. The contribution of this paper

is summarized in two points. First, we improved the basic method into a new method to generate a model that illustrates the discussion process. Second, we found suggestions for low-quality discussions to high-quality discussions using the models.

II. Related Work

The objectives of conversation and discussion analysis research are classified into two types. One of them is to acquire knowledge to develop conversational agents. The other is to acquire knowledge to improve conversations and discussions. The smooth interaction between conversational agents and users requires that users feel that the agents understand them. Thus, informing methods of information effectively that is understood by conversational agents from users' utterances have been proposed [2]–[4]. Conversational breakdowns between conversational agents and users would decrease users' satisfaction, trust, and willingness to continue using the agents [5], [6]. Methods for addressing conversational breakdowns in task-oriented dialogs have been proposed [7], [8]. These studies analyzed conversations between agents and users to develop conversational agents. Though we analyze discussions between humans, the findings are utilized to improve discussions rather than develop conversational agents.

Some approaches utilize verbal and non-verbal information to analyze conversations between humans. Methods using non-verbal information utilize gaze information, hand gestures, and acoustic information [9]. Gaze information is related to turn-taking in conversations [10]. It can estimate the degree of engagement in a conversation [11]. The effect of hand gestures on understanding a conversation is studied [12]. These studies analyze conversations using non-verbal information to obtain findings. Conversely, the proposed method here analyzes discussions using verbal information to obtain suggestions. Conversations in real life are conducted by using both verbal and non-verbal media. The present study specifically targets language features to find limitations to the suggestions.

The proposed method uses transcripts of discussions during research progress to analyze the discussion process. Zehnalova et al. and Inches et al. postulated methods for analyzing discussion topics and context using words in transcripts [13], [14]. Nishihara et al.'s method utilized dialog acts of utterances in transcripts for the analysis [15]. Since topics and context of discussions during research progress would be changed, it makes little sense to compare topics and context between the same or different discussants. Therefore, we propose a new method to compare discussions using dialog acts; the types, rates and transitions are utilized.

In linguistics and, particularly, in natural language understanding, a dialog act is an utterance, in the context of conversational dialog, that serves a function in the dialog. Conversation analysis often uses the dialog acts to label utterances [16]–[18]. The set of dialog acts is generally composed for every target conversations because frequent dialog acts and the granularity depend on the conversations. Quinn et al. posited their original dialog acts *Declarative, Interrogative, and Imperative* to analyze conversations between nursing support agents and users [19]. Germesin et al. posited two types of dialog acts *Agreement and Non-agreement* to analyze discussions on consensus building [20]. Qadir et al. developed four types of dialog acts *Commissives, Directives, Expressives, and Representatives* [21], and Lampert et al. set two types of dialog acts *Request and Non-request* [22]. According to the previous studies, we also design a new set of dialog acts for analyzing discussions during research progress. The dialog acts are obtained from actual utterances in the discussions we analyze. The new set of dialog acts would be used to generate a model for the discussion process.

As an approach for conversation analysis, both quantitative and qualitative analysis methods have been proposed [23], [24]. He et al. proposed a method for quantitative conversation analysis on consensus building [24]. The present study will also be a quantitative analysis because utterances are transformed into dialog acts, and their rates will be used for the analysis. A quantitative analysis would be influential in supporting discussion improvement because the analysis result will be quantitative so that the suggestions would be posed as concrete actions.

III. Proposed method

Fig. 1 shows the outline of the proposed method. First, discussions are recorded and classified into high-quality and low-quality discussion groups manually. Transcripts of each discussion are prepared manually. One line includes a speaker's name and an utterance. Each utterance is assigned labels for dialog acts. A matrix of the transition of dialog acts for each group is obtained. Each factor of the matrix has a transition rate between two different dialog acts. Transitions with a high rate are extracted from the matrix. The extracted transitions are connected if the same dialog act is included, and then a model for the discussion process is obtained. Two models for high-quality and low-quality discussion groups are obtained. By comparing the two models for high-quality and low-quality discussion groups, common and different points are investigated. It is posited that dialog acts and transitions included in the high-quality discussion group but not present in the low-quality would be suggestions for improving low-quality discussions to high-quality discussions. Discussants refer to the suggestions to have a better discussion next time.

A. Making transcripts of discussions

We posit that a research discussion should be conducted face-to-face. Research discussions are recorded by a voice recorder to generate transcripts. The proposed method uses transcripts of discussions to make a model for the discussion process. One line of a transcript includes a speaker's name and an utterance; an utterance includes fillers. The length of sound in a word (e.g., "weeeeeeeeell" is represented as just "well"), the length of silent time, and the information of laughing (e.g., "ha-ha-ha") are omitted. An utterance includes several sentences before turn-taking occurs. TABLE I exhibits an example of a part of a transcript.

TABLE I
EXAMPLE OF A ONE-TO-ONE DISCUSSION TRANSCRIPT. ONE LINE HAS A SPEAKER NAME AND AN UTTERANCE. THE THIRD COLUMN DISPLAYS ASSIGNED LABELS FOR DIALOG ACTS THAT ARE NOT INCLUDED IN A TRANSCRIPT.

Speaker	Utterance	Dialog Act
T	はい、じゃあ、えっと、よろしくお願いします。(Well, let's start the meeting.)	Greeting_T
T	えーっと、書き起こしが、まあ当然まだだと思ってるんですけども、えっと１人もらった？ (I'm sure that you have not made a transcription naturally. Did you get a recording data?)	Question_T
S	はい。(Yes.)	Answer_S
T	それは誰のん？(Who did you get it from?)	Question_T
S	岡村さんから。(From Riko.)	Answer_S
T	岡村さんからもらった、了解、了解。(You got it from Riko, OK.)	Understanding_T
T	それはなん分ぐらいのデータ？(How long was the data?)	Question_T
S	えっと、４５分ぐらいだったと思います。(Well, about 45minutes)	Answer_S
T	ながっ。(So long.)	Comment_T
S	長かったです。(It's so long.)	Repetition_S
T	すごい。(Awesome.)	Comment_T
S	今回は長かったって言ってたので。(She said the meeting was so long.)	Explanation_S, Report_S
T	あー、大変、大変、わかった、そしたらそれを書き起こしを是非、頑張ってください。(It will be tough work. Please make the transcript though it is tough.)	Request_T, Suggestion_T
S	はい。(Sure.)	Agreement_S
T	で、書き起こすときのポイントがあって、コンピュータで最後言葉を処理することになるので、単語の書き方を揃える。(There are some points in making transcripts. Please use the same words for the same objects because the transcripts would be processed by computer programs.)	Explanation_T, Suggestion_T
S	平仮名とか片仮名とかそういうことですか？(Do you mean to use the same representation for one object?)	Question_S
T	そういうことです、必ず、文の終わりに丸をつける（はい。）っていうのをしてってもらっていいでしょうか？ (That's right. Please add periods every end of a sentence.)	Answer_T

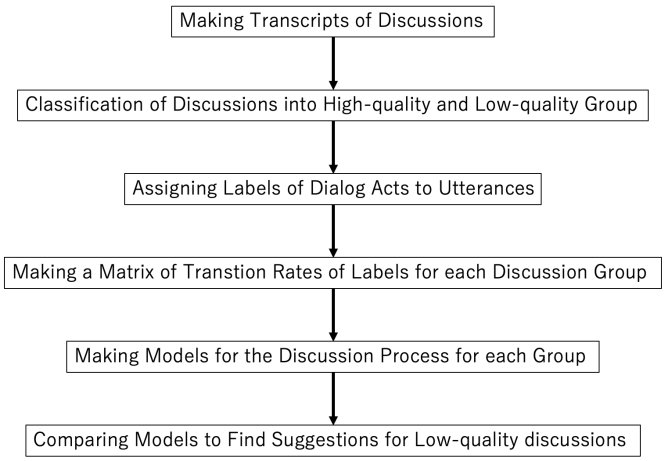

Fig. 1. Flowchart of the proposed method.

TABLE II
LABELS FOR DIALOG ACTS. T DENOTES A SUPERVISOR, WHILE S DENOTES A STUDENT. THERE ARE 62 TYPES OF LABEL; 31 LABELS FOR A SUPERVISOR'S UTTERANCE AND THE REST 31 LABELS FOR A STUDENT'S.

Greeting_T, Confirmation_T, Question_T, Answer_T, Agreement_T, Repetition_T, Explanation_T, Opinion_T, Admiration_T, Suggestion_T, Understanding_T, Topic shifting_T, Report_T, Degression_T, Soliloquy_T, Nodding_T, Request_T, Planning_T, Denial_T, Filler_T, Consultation_T, Response_T, Comment_T, Advice_T, Indication_T, Correction_T, Wondering_T, Surprise_T, Acknowledgement_T, Chatting_T, Additional comment_T
Greeting_S, Confirmation_S, Question_S, Answer_S, Agreement_S, Repetition_S, Explanation_S, Opinion_S, Admiration_S, Suggestion_S, Understanding_S, Topic shifting_S, Report_S, Degression_S, Soliloquy_S, Nodding_S, Request_S, Planning_S, Denial_S, Filler_S, Consultation_S, Response_S, Comment_S, Advice_S, Indication_S, Correction_S, Wondering_S, Surprise_S, Acknowledgement_S, Chatting_S, Additional comment_S

B. Labeling Dialog Acts to Utterances

Each utterance in a transcript is labeled with dialog acts. Though automatic labeling methods have been proposed [15], [25], the labeling accuracy is not enough to generate accurate models. Therefore, each utterance is labeled manually.

Though label sets of dialog acts also have been proposed [19]–[22], [26], most of the sets of labels have their target discussions and conversations. We design a new set of labels for our target discussions, that is, discussions for research progress.

TABLE II shows designed labels for dialog acts. These labels are designed by referring to transcripts of discussions that will be used in experiments in Section IV. An utterance from a supervisor and that from a student will be distinguished by a subscript to a label (T and S). The proposed method uses $31 \times 2 = 62$ types of the label for dialog acts in total.

When labeling dialog acts to an utterance, the already appeared utterances are also considered because one utterance may not have enough information for labeling. If a student says "Yes" to a question from a supervisor, it means "Yes, you are right" labeled as "Answer_S." However, if a student says "Yes" to a proposal from a supervisor, it means "Yes, I will do it" labeled as "Agreement_S." The two instances of "Yes" are different types of utterance. Multiple labels may be put to an utterance because a single utterance may have several roles.

C. Making a matrix of transition rates of dialog acts

A matrix of transition rates of dialog acts is obtained for each group. The transition rate means a rate between labels for dialog acts. Suppose that ith utterance has a vector of labels

TABLE III
EXAMPLE OF A MATRIX OF TRANSITION RATES OF DIALOG ACTS. START DENOTES A DIALOG ACT IN THE STARTING POINT OF TRANSITION AND END REPRESENTS THAT IN THE ENDING POINT.

Start ＼ End	Greeting_T	Question_T	Understanding_T	Suggestion_T	Confirmation_T	Answer_S	Repetition_S	Agreement_S	Question_S	...
Greeting_T	0	0	0	0	0	0	0	0	0	...
Question_T	0	0	0	0	0	8.2%	0	3%	0	...
Understanding_T	0	0	0	0	0	0	0	0	0	...
Suggestion_T	0	0	0	0	0	0	0	1.4%	1.4%	...
Confirmation_T	0	0	0	0	0	3.1%	0	0.5%	0	...
Answer_S	0	5.4%	0	0.3%	0	0	0	0	0	...
Repetition_S	0	0	0	0	0	0	0	0	0	...
Agreement_S	0	3%	0	0.8%	0.7%	0	0	0	0	...
Question_S	0	0	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...

$L(i)$ and jth utterance has a vector of labels $L(j)$ ($j = i+1$). $L(i)$ is described by Eq. (1).

$$L(i) = \{l_n | 0 <= n <= 61\}, \quad (1)$$

where n is an index of label l_n which is equal to 0 or 1. If an nth label is put to an utterance, $l_n = 1$. Otherwise, $l_n = 0$. The frequency of transition from the label l_n in $L(i)$ to the label l_m in $L(j)$ is counted up if both l_n and l_m are not equal to zero. Let the frequency be $f_{n,m}$. Let the number of lines of the transcript be NL. $NL-1$ is the number of turn-taking in a discussion. The transition rate $r_{n,m}$ from the label l_n to the label l_m is calculated by Eq. (2).

$$r_{n,m} = \frac{f_{n,m}}{NL-1}, \quad (2)$$

where n and m are indices of labels. By assigning a rate $r_{n,m}$ to each cell, a matrix of transition rates is obtained. TABLE III illustrates an example of a part of the matrix.

D. Making a model for the discussion process

A model of the discussion process for each group is obtained by using the matrix of transition rates. Transitions of dialog acts with a rate more than the threshold TH are extracted from the matrix. The extracted transitions are connected if the same dialog act is included in two different transitions. The network of connected transitions will be a model for the discussion process.

E. Comparing models to find suggestions

The two models for high-quality and low-quality discussion groups are compared. The comparison is conducted by investigating dialog acts and transitions included in the models. First, common points in dialog acts and transitions between the two models are found. Then, different points between the two models are investigated. The different points might be suggestions for low-quality discussions to high-quality discussions.

IV. EXPERIMENT

We investigated whether the proposed method could generate a discussion process model corresponding to the quality of discussion and whether the proposed method could support finding suggestions for low-quality discussions to high-quality discussions. We would verify the former through the experiments in this section, and the latter in Section V.

A. Experimental hypotheses

We proposed two main hypotheses H1 and H2, each of which had two smaller hypotheses; in total there were four small hypotheses.

- H1a If the quality of the discussion is different, the dialog acts are different.
- H1b If the quality of the discussion is different, the transitions are different.
- H2a If the stage of the thesis is different, the dialog acts are different.
- H2b If the stage of the thesis is different, the transitions are different.

H1a hypothesized that if the quality of the discussion differs between the models, the dialog acts in each model are different. H1b was obtained by replacing dialog acts with transitions. If those hypotheses are proved, it will be verified that the proposed method can generate a model for the discussion process that corresponds with the quality of discussion.

H2a hypothesized that if the stage of the thesis differs between the models, dialog acts in each model are different. H2b was obtained by replacing dialog acts with transitions. Discussions during research progress should be held by supervisors and students regularly. The students will acquire the skill of discussion by practicing discussions repeatedly. The supervisors will also change the mode of explanation to match each student. If those hypotheses are proved, dialog acts and transitions only in a model of recent discussions might indicate changes obtained by the practices repeatedly.

B. Experimental procedures

The experiments were conducted as follows:

1) The experimenter collects discussion records and classifies them into high-quality and low-quality discussion groups. The experimenter collects discussions in two stages of the thesis (early and late stages).
2) The models for high-quality and low-quality discussion groups are obtained using the proposed method.
3) The hypotheses are tested from the four models obtained.

The experiments generated four discussion process models for:

1) a **high-quality** discussion group in the **early** stage of the thesis,
2) a **low-quality** discussion group in the **early** stage of the thesis,
3) a **high-quality** discussion group in the **late** stage of the thesis, and
4) a **low-quality** discussion group in the **late** stage of the thesis.

Hypothesis H1a was to be verified by comparing dialog acts between models 1) and 2), and between 3) and 4). Hypothesis H1b was to be verified by comparing transitions of dialog acts between models 1) and 2), and between 3) and 4). Hypothesis H2a was to be verified by comparing dialog acts between models 1) and 3), and between 2) and 4). Hypothesis H2b was to be verified by comparing transitions of dialog acts between models 1) and 3), and between 2) and 4).

C. Used data

We used 16 transcripts of discussions between a supervisor and a student in our laboratory. Half of them (eight out of 16) were collected in the early stage of the thesis (around April in 2019), while the rest was collected in the late stage (around October in 2019, after six months). Two supervisors and eight students (four males and four females) were in our laboratory, and each of the supervisors had four students. The students were 21-22 years old and studying in the College of Information Science and Engineering. The transcripts of discussions were read by the first and second authors. The two authors classified the transcripts into two groups: a high-quality and a low-quality discussion groups.

TABLE IV shows the details of the eight transcripts of discussions held in the early stage of the thesis: the length of discussion, the number of utterances from a supervisor, and that from a student are described. The average length of the discussions was 26min and 58s. The average number of utterances from a supervisor was 81, whereas that from a student was 71.5. The transcripts with IDs 1-4 were in the high-quality discussion group, while those with IDs 5-8 were in the low-quality discussion group. TABLE V shows the details of the eight transcripts of discussions held in the late stage of the thesis. The average length of the discussions was 25min and 17s. The average number of utterances from a supervisor was 86.3, whereas that from a student was 81.0. The transcripts with IDs 9-12 were in the high-quality discussion group, while those with IDs 13-16 were in the opposite group.

Compared with the discussions in the early stage of the thesis, the length was shortened by approximately 1 min, the number of utterances from a supervisor increased (five utterances), and the number from a student also increased (10 utterances). The increase in the number of utterances, even in less discussion time, denotes that the discussions in the late stage were more active. The number of utterances from a student was more significant than that from a supervisor, which indicated that the students might acquire the skill of discussion through repeated practice.

TABLE IV
USED EIGHT TRANSCRIPTS OF THE DISCUSSION HELD IN THE **EARLY** STAGE OF THE THESIS. THE LENGTH OF DISCUSSION, THE NUMBERS OF UTTERANCES FROM SUPERVISORS AND STUDENTS ARE PRESENTED. IDS 1-4 ARE IN HIGH-QUALITY DISCUSSION GROUP, WHEREAS IDS 5-8 ARE IN LOW-QUALITY DISCUSSION GROUP.

Discussion ID	Length (min : sec)	# of utterances from a supervisor	# of utterances from a student
1	48:02	139	127
2	24:29	50	43
3	19:20	50	44
4	44:11	151	145
5	20:38	53	50
6	28:55	100	76
7	17:25	67	58
8	12:42	38	29
Average	26:58	81.0	71.5

TABLE V
USED EIGHT TRANSCRIPTS OF THE DISCUSSION HELD IN THE **LATE** STAGE OF THE THESIS. THE LENGTH OF DISCUSSION, THE NUMBERS OF UTTERANCES FROM SUPERVISORS AND STUDENTS ARE PRESENTED. IDS 9-12 ARE IN HIGH-QUALITY DISCUSSION GROUP, WHEREAS IDS 13-16 ARE IN LOW-QUALITY DISCUSSION GROUP.

Discussion ID	Length (min : sec)	# of utterances from a supervisor	# of utterances from a student
9	32:02	135	130
10	19:57	48	46
11	24:47	89	82
12	16:38	68	65
13	22:05	88	82
14	30:43	94	85
15	24:37	76	73
16	31:23	92	85
Average	25:17	86.3	81.0

D. Experimental results

First, we report frequently appearing labels for dialog act in discussions of high-quality and low-quality groups. TABLE VI and TABLE VII exhibit labels for dialog acts that appeared more than 10 times in each group (high-quality/low-quality in the early/late stage). Though there were common points in appeared dialog acts, the different points were also found.

Next, we report characteristic transitions found in high-quality and low-quality discussion groups. TABLE VIII and TABLE IX exhibit transitions frequently appearing in each group (the appearance rate was more than 1.7% which was the threshold TH [1]). Though there were common points in appearing transitions of dialog acts, different points were also found.

Figs. 2, 3, 4, and 5 illustrate the models obtained for the discussion process. Fig. 2 is for a high-quality discussion group and Fig. 3 is for a low-quality discussion group in the early stage. Fig. 4 is for a high-quality discussion group and Fig. 5 is for a low-quality discussion group in the late stage.

[1]The rate of the used transitions was almost 40% for all transitions found in the transcripts.

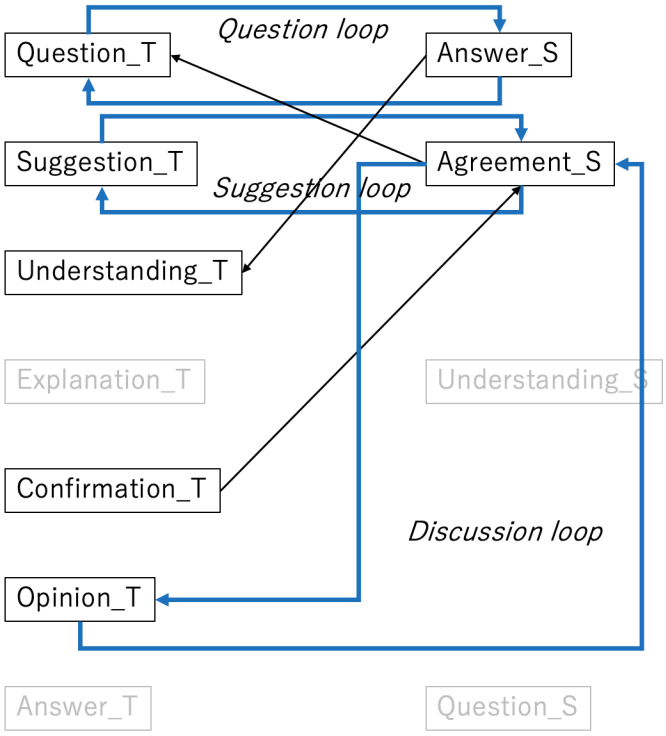

Fig. 2. **High-quality** discussion process model in the **early** stage.

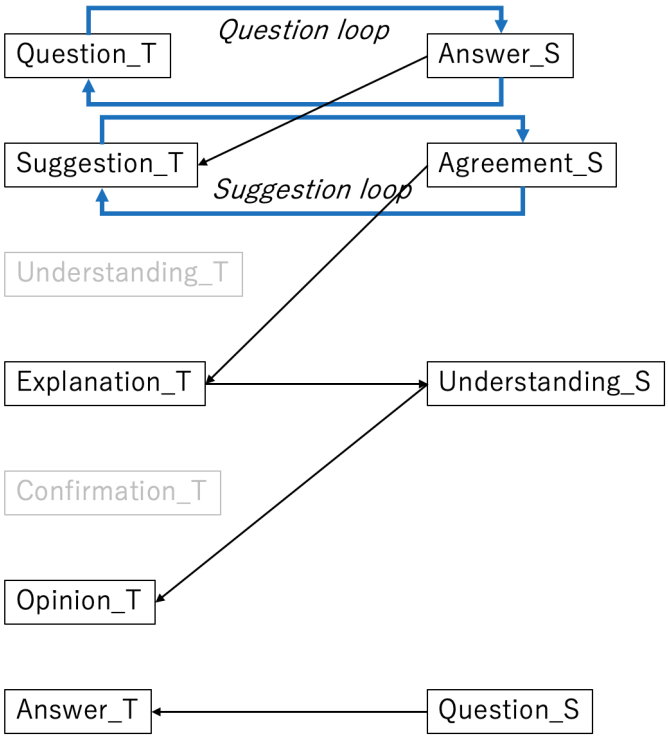

Fig. 3. **Low-quality** discussion process model in the **early** stage.

TABLE VI
LABELS OF DIALOG ACT APPEARING MORE THAN 10 TIMES IN EACH GROUP OF DISCUSSIONS HELD IN THE **EARLY** STAGE. THE LEFT-SIDE IS FOR HIGH-QUALITY DISCUSSION GROUP AND THE RIGHT-SIDE IS FOR LOW-QUALITY DISCUSSION GROUP. THE UPPER-HALF IS FROM SUPERVISORS AND THE BOTTOM-HALF IS FROM STUDENTS.

High-quality		Low-quality	
Dialog act	Frequency	Dialog act	Frequency
Suggestion_T:	53	Question_T	121
Question_T	51	Opinion_T	70
Opinion_T	33	Suggestion_T	34
Explanation_T	29	Confirmation_T	33
Answer_T	24	Explanation_T	26
Confirmation_T	13	Understanding_T	19
Understanding_T	10	Agreement_T	18
Agreement_T	10	Indication_T	17
		Answer_T	16
		Chatting_T	12
		Advice_T	12
Agreement_S	50	Answer_S	121
Answer_S	48	Agreement_S	96
Understanding_S	28	Understanding_S	34
Question_S	25	Opinion_S	20
Report_S	20	Question_S	19
Confirmation_S	12	Report_S	17
Opinion_S	10	Repetition_S	17
		Chatting_S	11
		Wondering_S	11

TABLE VII
LABELS OF DIALOG ACT APPEARING MORE THAN 10 TIMES IN EACH GROUP OF DISCUSSIONS HELD IN THE **LATE** STAGE. THE LEFT-SIDE IS FOR HIGH-QUALITY DISCUSSION GROUP AND THE RIGHT-SIDE IS FOR LOW-QUALITY DISCUSSION GROUP. THE UPPER-HALF IS FROM SUPERVISORS AND THE BOTTOM-HALF IS FROM STUDENTS.

High-quality		Low-quality	
Dialog act	Frequency	Dialog act	Frequency
Question_T	87	Opinion_T	63
Opinion_T	68	Explanation_T	48
Suggestion_T	44	Question_T	46
Explanation_T	41	Understanding_T	42
Confirmation_T	36	Confirmation_T	40
Request_T	18	Suggestion_T	35
Agreement_T	16	Agreement_T	28
Understanding_T	15	Answer_T	14
Wondering_T	12	Indication_T	11
Repetition_T	11	Advice_T	10
Indication_T	10		
Agreement_S	110	Opinion_S	114
Answer_S	84	Answer_S	44
Understanding_S	34	Understanding_S	38
Report_S	31	Report_S	36
Nodding_S	24	Opinion_S	35
Opinion_S	17	Confirmation_S	19
		Nodding_S	15
		Wondering_S	13
		Question_S	12

E. Hypotheses Verification

We investigated the differences of dialog acts between the high-quality and low-quality discussion groups in the early stage to verify H1a. TABLE VI shows 22 types of dialog acts in total. Thirteen of them were common, whereas seven of them were found only in high-quality discussion group or low-quality discussion group. Then, we investigated the differences of dialog acts between the high-quality and low-

TABLE VIII
TRANSITIONS OF DIALOG ACTS APPEARING FREQUENTLY IN EACH GROUP OF DISCUSSIONS HELD IN THE **EARLY** STAGE. THE LEFT-SIDE IS FOR HIGH-QUALITY DISCUSSION GROUP AND THE RIGHT-SIDE IS FOR LOW-QUALITY DISCUSSION GROUP.

High-quality		Low-quality	
Transition	Rate	Transition	Rate
Question_T→ Answer_S	8.2%	Question_T→Answer_S	12.4%
Suggestion_T→Agreement_S	4.4%	Answer_S→Question_T	5.8%
Agreement_S→Suggestion_T	3.8%	Opinion_T→Opinion_S	3.8%
Question_S→Answer_T	3.4%	Agreement_S→Opinion_T	3.7%
Answer_S→Question_T	3.1%	Suggestion_T→Agreement_S	2.7%
Explanation_T→Understanding_S	2.3%	Confirmation_T→Agreement_S	2.2%
Understanding_S→Opinion_T	1.9%	Agreement_S→Question_T	2.0%
Agreement_S→Explanation_T	1.7%	Answer_S→Understanding_T	1.7%
Answer_S→Suggestion_T	1.7%	Agreement_S→Suggestion_T	1.7%

TABLE IX
TRANSITIONS OF DIALOG ACTS APPEARING FREQUENTLY IN EACH GROUP OF DISCUSSIONS HELD IN THE **LATE** STAGE. THE LEFT-SIDE IS FOR HIGH-QUALITY DISCUSSION GROUP AND THE RIGHT-SIDE IS FOR LOW-QUALITY DISCUSSION GROUP.

High-quality		Low-quality	
Transition	Rate	Transition	Rate
Question_T→Anwer_S	9.1%	Question_T→Answer_S	5.2%
Suggestion_T→Agreement_S	4.3%	Confirmation_T→Agreement_S	4.3%
Confirmation_T→Agreement_S	3.4%	Agreement_S→Opinion_T	3.7%
Answer_S→Question_T	3.3%	Opinion_T→Agreement_S	3.6%
Agreement_S→Opinion_T	3.0%	Suggestion_T→Agreement_S	3.1%
Opinion_T→Opinion_S	3.0%	Agreement_S→Suggestion_T	2.6%
Opinion_S→Suggestion_T	2.7%	Explanation_T→Understanding_S	2.5%
Opinion_S→Question_T	2.5%	Agreement_S→Confirmation_T	2.2%
Explanation_T→Understanding_S	2.2%	Report_S→Understanding_T	2.1%
Answer_S→Opinion_T	2.0%	Understanding_S→Explanation_T	2.0%
Answer_S→Confirmation_T	1.8%	Opinion_T→Understanding_S	1.9%
Request_T→Agreement_S	1.8%		
Agreement_S→Explanation_T	1.7%		

quality discussion groups in the late stage. TABLE VII shows 23 types of dialog acts in total. Thirteen of them were common, while six of them were found only in one discussion group. The results indicated that the dialog acts found in the discussion process depend on the quality of discussion. Hypothesis H1a was verified.

We investigated the differences of transitions between high-quality and low-quality discussion groups in the early stage to verify H1b. TABLE VIII shows 14 types of transition in total. Four of them were common, whereas 10 of them were found in only one group. Then, we investigated the differences in transitions between high-quality and low-quality discussion groups in the late stage. TABLE IX shows 16 types of transitions in total. Five of them were common, whereas 14 of them were found only in one group. The results indicated that the transitions found in the discussion process depend on the quality of discussion. Hypothesis H1b was verified.

We investigated the differences of dialog acts found in high-quality discussion groups between the early stage and the late stage to verify H2a. TABLE VI and TABLE VII show 20 types of dialog act in high-quality discussion group. Twelve of them were common, while eight of them were found only in one group. Then, we investigated the differences of dialog acts found in low-quality discussion groups between the early stage and the late stage. TABLE VI and TABLE VII show 23 types of dialog act in low-quality discussions. Sixteen of them were common, while seven of them were found in only one group. The results indicated that the dialog acts found in the discussion process depend on the stage of the thesis. Hypothesis H2a was verified.

We investigated the differences of transitions found in high-quality discussion groups between the early stage and the late stage to verify H2b. TABLE VIII and TABLE IX show 17 types of transition in high-quality discussion groups. Five of them were common, while 12 of them were found in only one group. Then, we investigated the differences of transitions found in low-quality discussion groups between the early stage and the late stage. TABLE VIII and TABLE IX show 15 types of transition in low-quality discussion groups. Six of them were common, while nine of them were found only in one group. The results indicated that the transitions found in the discussion process depend on the stage of the thesis. Hypothesis H2b was verified.

Since all the hypotheses were verified, we found that the proposed method could generate a model for the discussion process that corresponded to the quality of discussion and the stage of the thesis. The next section investigates suggestions for low-quality discussions to high-quality discussions comparing the four models.

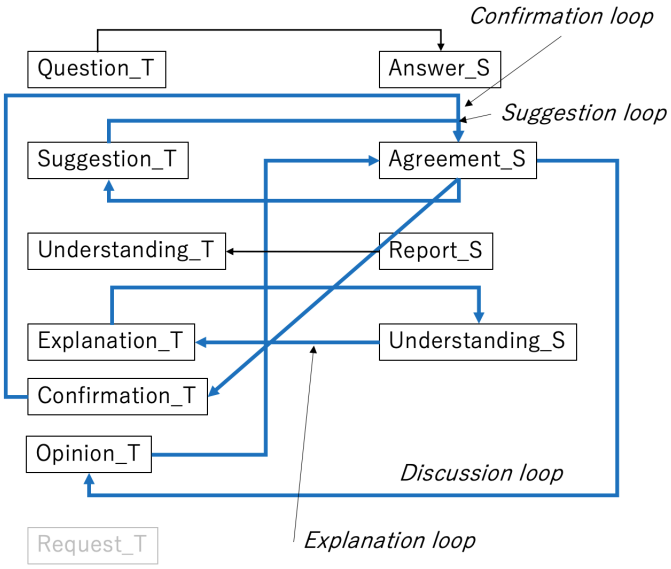

Fig. 4. **High-quality** discussion process model in the **late** stage.

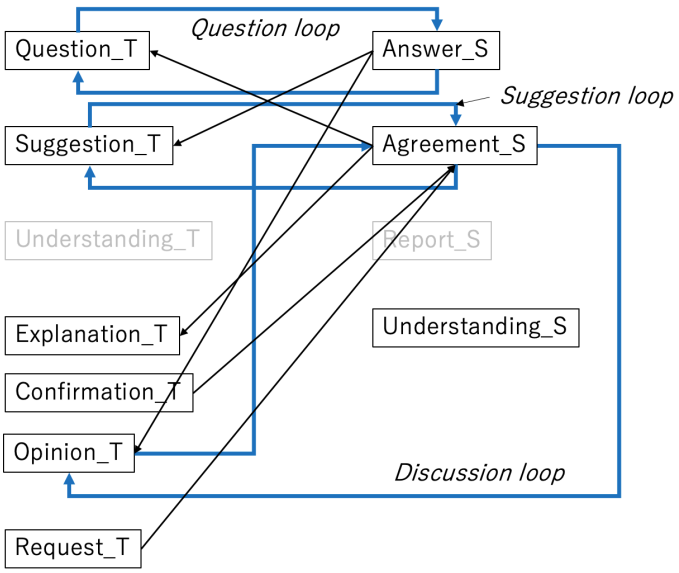

Fig. 5. **Low-quality** discussion process model in the **late** stage.

V. Discussion

We discuss the models obtained for the discussion process and investigate suggestions for low-quality discussions to high-quality discussions.

A. Difference between models in the early stage

First, we discuss the common points between high-quality and low-quality discussion groups in the early stage, referring to Figs. 2 and 3. Both models had the same loop of Question_T $\leftrightarrow$ Answer_S that illustrated a repetition of questioning by supervisors and answering by students. The loop could be interpreted as a process by which supervisors tried to make students understand their thesis and progress (we name it the *Question loop*). Moreover, both models had the same loop of Suggestion_T $\leftrightarrow$ Agreement_S that illustrated a repetition of suggestions by supervisors and agreements by students. The loop could be interpreted as a process by which supervisors tried to make their students agree on the objectives of their thesis (we name it the *Suggestion loop*).

The different points between the high-quality and low-quality discussion groups in the early stage were also found. The model for the high-quality discussion group had a loop of Opinion_T $\leftrightarrow$ Agreement_S that represented a repetition of supervisors giving opinions and students agreeing. The loop could be interpreted as a process by which both supervisors and students discussed their thesis (we name it the *Discussion loop*).

The model for the high-quality discussion group had other significant transitions; Answering_S $\rightarrow$ Understanding_T that represented supervisors' understandings of students' answers and Confirmation_T $\rightarrow$ Agreement_S that illustrated students' agreements on supervisor's confirmations. The significant transitions indicated that both supervisors and students in high-quality discussion group could have constructive discussions based on students' understandings.

Meanwhile, the model for the low-quality discussion group had different transitions: Answer_S $\rightarrow$ Suggestion_T that represented supervisors' suggestions complying with students' progress, Understanding_S $\rightarrow$ Opinion_T that represented supervisors giving opinions for the students' understandings, and Question_S $\rightarrow$ Answer_T that represented supervisors' answering students' questions. Those transitions indicated that supervisors compensated for the lack of students' understanding.

B. Difference between models in the late stage

First, we discuss the common points between the high-quality and low-quality discussion groups in the late stage, referring to Fig. 4 and Fig. 5. Both models had *Discussion loop*s that were included only in the model for the high-quality discussion group in the early stage. The partial change indicated that both supervisors and students in the late stage came to discuss their thesis to some extent.

However, the model for the low-quality discussion group kept in three transitions Answering_S $\rightarrow$ Suggestion_T, Agreement_S $\rightarrow$ Explanation_T, and Explanation_T $\rightarrow$ Understading_S that were included in the model for the low-quality discussion group in the early stage. The remaining of the three transitions indicated that the students in the low-quality discussion group did not understand their thesis perfectly. The model was added to with new transitions Answering_S $\rightarrow$ Confirmation_T and Answering_S $\rightarrow$ Opinion_T that represented supervisors giving confirmations and opinions for students' answers. The two transitions indicated that supervisors could not understand or satisfy with the students' answers. Moreover, a transition of Request_T $\rightarrow$ Agreement_S also emerged that implied supervisors had trouble making their

students understand by explanation and moved to order them specific tasks to finish their thesis.

Meanwhile, the model for the high-quality discussion group had a new loop of Explanation_T ↔ Understanding_S that represented students' understandings of supervisors' detailed explanations (we name it the *Explanation loop*). The model had another new loop of Confirmation_T ↔ Agreement_S that represented a repetition of students' agreements for supervisors' confirmations (we name it the *Confirmation loop*). The model for the high-quality discussion group comprised the four loops of transitions. The four loops illustrated that supervisors and students discussed their specific topics in depth. Since more loops were found in the discussion process, we could estimate that the discussions had various topics with deep insights.

The model for the high-quality discussion group did not include a *Question loop* that was in the model for the high-quality discussion group in the early stage, because a transition of Answer_S → Question_T was disappeared, which meant that students started to answer supervisors' questions appropriately; therefore, it was not necessary to repeat questions. One more newly added transition was Report_S → Understanding_T that represented students reporting their progress and supervisors understanding them. The transitions indicated that students came to understand and conduct their thesis appropriately.

C. Suggestions for low-quality discussions to high-quality discussions

The above findings were summarized as suggestions for low-quality discussions to high-quality discussions. The discussants should drive their discussions to include the following transitions.

1) Supervisors conduct discussions so that students can understand their thesis. The suggestion is realized by removing one-way transitions in Fig. 3 except the transition of Explanation_T → Understanding_S.
2) Supervisors lead students to agree on supervisors' opinions based on the students' understanding. The suggestion is realized by adding a *Discussion loop*.
3) Both supervisors and students have discussions to understand each other's opinions. The suggestion is realized by conducting discussions that the five changes occur:
    a) Students answer supervisors' questions appropriately based on their understandings. The suggestion is realized by removing a *Question loop*.
    b) Supervisors do not explain unilaterally. They should check students' understandings and explain points one by one. The suggestion is realized by adding an *Explanation loop*.
    c) Supervisors do not confirm unilaterally. They should check students' agreements and confirm them one by one. The suggestion is realized by adding a *Confirmation loop*.
    d) Students should come to understand supervisors' opinions. They should not agree on everything without understanding.
    e) Students should report their progress in a way that can be understood by supervisors, which means students must conduct their thesis based on their understandings.

VI. Conclusion

This paper proposed a generation method for the discussion process model during research progress. First, discussions are recorded, then transcripts of them are made that have a speaker name and an utterance in each line. The transcripts are classified into high-quality and low-quality discussion groups. After classification, labels for dialog acts are assigned to each utterance. The labels for dialog acts are originally designed for discussion analysis for research progress. Frequently appearing transitions of dialog acts are extracted in each group. By connecting dialog acts that are common in two transitions, a network of dialog acts is generated. The network illustrates a model for the discussion process. The models for high-quality and low-quality discussion groups are compared. The differences between the two models should be the suggestions for low-quality discussions to high-quality discussions.

We experimented with the proposed method using eight discussions in the early stage of the thesis and another eight discussions in the late stage. We made two main hypotheses: if the quality of discussion is different, the dialog acts and the transitions in the models are also different (H1), and if the stage of the thesis is different, the dialog acts and the transitions in the models are also different (H2). We classified the 16 discussions into four groups considering the quality (high/low) and the stage (early/late), then generated models for each group. By comparing the four models, we proved that all hypotheses were valid. The experimental results verified that the proposed method generated models corresponding to the quality of discussion and the stage of the thesis. We investigated the four models and found the suggestions for low-quality discussions to high-quality discussions. The suggestions were summarized in three points: (1) Supervisors should conduct discussions so that students understand their thesis. (2) Supervisors should lead students to agree on their opinions based on the students' understanding. (3) Both supervisors and students have discussions to try to understand each other's opinions.

Future studies will endeavor to apply the findings to improve discussions. It is assumed that the way of conducting discussions depends on the research field and experiences. There might be limitations in the proposed method because the discussions used were about information science and the students were engaged in their first piece of research. We will make several improvements to the proposed method to cope with the differences in the research field and experiences.

ACKNOWLEDGMENT

This work was supported by Recruit Management Solutions Co., Ltd. Grant. We show our best appreciation.

REFERENCES

[1] S. Tsuji, Y. Nishihara, W. Sunayama, R. Yamanishi, and S. Imashiro, "Analysis method for one-to-one discussion process for research progress using transition probability of utterance types," in The Thirteenth International Conference on Advances in Computer-Human Interactions, 2020, pp. 210–213.

[2] D. J. Litman and S. Silliman, "Itspoke: An intelligent tutoring spoken dialogue system," in Demonstration Papers at HLT-NAACL 2004, ser. HLT-NAACL–Demonstrations '04. USA: Association for Computational Linguistics, 2004, p. 5–8.

[3] A. Raux, B. Langner, D. Bohus, A. Black, and M. Eskenazi, "Let's go public! taking a spoken dialog system to the real world," in 9th European Conference on Speech Communication and Technology, 01 2005, pp. 885–888.

[4] K. Inoue, D. Lala, K. Yamamoto, S. Nakamura, K. Takanashi, and T. Kawahara, "An attentive listening system with android ERICA: Comparison of autonomous and WOZ interactions," in Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 1st virtual meeting: Association for Computational Linguistics, Jul. 2020, pp. 118–127. [Online]. Available: https://www.aclweb.org/anthology/2020.sigdial-1.15

[5] F. Bentley, C. Luvogt, M. Silverman, R. Wirasinghe, B. White, and D. Lottridge, "Understanding the long-term use of smart speaker assistants," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 2, no. 3, Sep. 2018. [Online]. Available: https://doi.org/10.1145/3264901

[6] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz, "Resilient chatbots: Repair strategy preferences for conversational breakdowns," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–12. [Online]. Available: https://doi.org/10.1145/3290605.3300484

[7] C.-H. Li, K. Chen, and Y.-J. Chang, "When there is no progress with a task-oriented chatbot: A conversation analysis," in Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, ser. MobileHCI '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3338286.3344407

[8] T. J.-J. Li, J. Chen, H. Xia, T. M. Mitchell, and B. A. Myers, "Multi-modal repairs of conversational breakdowns in task-oriented dialogs," in Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, ser. UIST '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1094–1107. [Online]. Available: https://doi.org/10.1145/3379337.3415820

[9] X. Liu, A. Xu, V. Sinha, and R. Akkiraju, "Voice of customer: A tone-based analysis system for online user engagement," in Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, ser. CHI EA '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–6. [Online]. Available: https://doi.org/10.1145/3170427.3188454

[10] K. Jokinen, H. Furukawa, M. Nishida, and S. Yamamoto, "Gaze and turn-taking behavior in casual conversational interactions," ACM Trans. Interact. Intell. Syst., vol. 3, no. 2, Aug. 2013. [Online]. Available: https://doi.org/10.1145/2499474.2499481

[11] R. Bednarik, S. Eivazi, and M. Hradis, "Gaze and conversational engagement in multiparty video conversation: An annotation scheme and classification of high and low levels of engagement," in Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction, ser. Gaze-In '12. New York, NY, USA: Association for Computing Machinery, 2012. [Online]. Available: https://doi.org/10.1145/2401836.2401846

[12] Y. Xiong, F. Quek, and D. McNeill, "Hand gesture symmetric behavior detection and analysis in natural conversation," in Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, ser. ICMI '02. USA: IEEE Computer Society, 2002, p. 179. [Online]. Available: https://doi.org/10.1109/ICMI.2002.1166989

[13] S. Zehnalova, Z. Horak, and M. Kudelka, "Email conversation network analysis: Work groups and teams in organizations," in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ser. ASONAM '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1262–1268. [Online]. Available: https://doi.org/10.1145/2808797.2808862

[14] G. Inches and F. Crestani, "Online conversation mining for author characterization and topic identification," in Proceedings of the 4th Workshop on Workshop for Ph.D. Students in Information and Knowledge Management, ser. PIKM '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 19–26. [Online]. Available: https://doi.org/10.1145/2065003.2065009

[15] Y. Nishihara and W. Sunayama, "Estimation of friendship and hierarchy from conversation records," Inf. Sci., vol. 179, no. 11, May 2009, p. 1592–1598. [Online]. Available: https://doi.org/10.1016/j.ins.2008.11.024

[16] G. A. Wang, H. J. Wang, J. Li, A. S. Abrahams, and W. Fan, "An analytical framework for understanding knowledge-sharing processes in online q&a communities," ACM Trans. Manage. Inf. Syst., vol. 5, no. 4, Dec. 2014. [Online]. Available: https://doi.org/10.1145/2629445

[17] S. Oraby, M. Bhuiyan, P. Gundecha, J. Mahmud, and R. Akkiraju, "Modeling and computational characterization of twitter customer service conversations," ACM Trans. Interact. Intell. Syst., vol. 9, no. 2–3, Mar. 2019. [Online]. Available: https://doi.org/10.1145/3213014

[18] A. Ahmadvand, J. I. Choi, and E. Agichtein, "Contextual dialogue act classification for open-domain conversational agents," in Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1273–1276. [Online]. Available: https://doi.org/10.1145/3331184.3331375

[19] K. Quinn and O. Zaïane, "Identifying questions & requests in conversation," ACM International Conference Proceeding Series, 08 2014.

[20] S. Germesin and T. Wilson, "Agreement detection in multiparty conversation," in Proceedings of the 2009 International Conference on Multimodal Interfaces, ser. ICMI-MLMI '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 7–14. [Online]. Available: https://doi.org/10.1145/1647314.1647319

[21] A. Qadir and E. Riloff, "Classifying sentences as speech acts in message board posts," in Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, UK.: Association for Computational Linguistics, Jul. 2011, pp. 748–758. [Online]. Available: https://www.aclweb.org/anthology/D11-1069

[22] A. Lampert, R. Dale, and C. Paris, "Detecting emails containing requests for action," in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Los Angeles, California: Association for Computational Linguistics, Jun. 2010, pp. 984–992. [Online]. Available: https://www.aclweb.org/anthology/N10-1142

[23] F. Cabrerizo, I. Pérez, and E. Herrera-Viedma, "Managing the consensus in group decision making in an unbalanced fuzzy linguistic context with incomplete information," Knowledge-Based Systems, vol. 23, no. 2, 2010, pp. 169 – 181. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705109001610

[24] N. He and O. Yoshie, "Conversation analysis based on interpersonal relationship in consensus building," in Proceedings of International Conference on Information Integration and Web-Based Applications & Services, ser. IIWAS '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 26–33. [Online]. Available: https://doi.org/10.1145/2539150.2539178

[25] Y. Hoshikawa and K. Wakabayashi, "Automatic extraction of discussion based on sentence type estimation," in Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, ser. CSCW '17 Companion. New York, NY, USA: Association for Computing Machinery, 2017, p. 203–206. [Online]. Available: https://doi.org/10.1145/3022198.3026336

[26] D. Jurafsky, E. Shriberg, B. Fox, and T. Curl, "Lexical, prosodic, and syntactic cues for dialog acts," in Discourse Relations and Discourse Markers, 1998. [Online]. Available: https://www.aclweb.org/anthology/W98-0319