Automatic Extraction of Task Statements from Structured Meeting Content

Katashi Nagao, Kei Inoue, Naoya Morita and Shigeki Matsubara
Department of Media Science, Graduate School of Information Science, Nagoya University, Nagoya, Japan

Keywords: Discussion Mining, Discussion Structure, Task Statement, Automatic Extraction, Probability Model.

Abstract: We previously developed a discussion mining system that records face-to-face meetings in detail, analyzes their content, and conducts knowledge discovery. Looking back on past discussion content by browsing documents, such as minutes, is an effective means for conducting future activities. In meetings at which some research topics are regularly discussed, such as seminars in laboratories, the presenters are required to discuss future issues by checking urgent matters from the discussion records. We call statements including advice or requests proposed at previous meetings "task statements" and propose a method for automatically extracting them. With this method, based on certain semantic attributes and linguistic characteristics of statements, a probabilistic model is created using the maximum entropy method. A statement is judged whether it is a task statement according to its probability. A seminar-based experiment validated the effectiveness of the proposed extraction method.

1 INTRODUCTION

Evidence-based research, such as research on life-logging (Sellen and Whittaker, 2010) and big data applications (Mayer-Schönberger and Cukier, 2013), has been receiving much attention and has led to the proposal of techniques for improving the quality of life by storing and analyzing data on daily activities in large quantities. These types of techniques have been applied in the education sector, but a crucial problem remains to be overcome: it is generally difficult to record intellectual activities and accumulate and analyze such data on a large scale. Since this kind of data is not possible to compress in a manner, such as taking the average, it is necessary to maintain the original data as the instances of cases. Such human intellectual-activity data should be treated as big data in the near future.

The aim of the study was to develop an environment in which the skills of students are empowered by analysis of abundant discussion data. We have developed a "discussion mining" system (Nagao et al., 2005) that generates meeting minutes linked to videos and audio data of the discussions. It also creates metadata for use in clarifying the semantic structure of the discussions. Statements made in meetings are classified into two types: "start-up," which means the statement starts a discussion of a new topic, and "follow-up," which means the statement continues the current topic of the discussion. The discussions are then segmented into discussion chunks corresponding to topics on the basis of the statement type. A discussion chunk is a set of statements that are semantically associated with each other.

Looking back and reconsidering the content of past discussions by browsing the recorded meeting content is an effective means for efficiently conducting future activities. In meetings at which some research topics are regularly discussed, such as seminars in laboratories, the presenters are required to discuss future issues by checking urgent matters from the structured discussion content.

In this paper, we call statements including advice or requests proposed at previous meetings "task statements" and propose a method for automatically extracting them. With this method, based on certain semantic attributes and linguistic characteristics of statements, a probabilistic model is created using the maximum entropy method (Wu, 1997).

We first discuss related work then describe our mining system. We then explain the proposed automatic extraction method of task statements from structured meeting content and describe our evaluation of this proposed method through a statistical hypothesis test.

Nagao, K., Inoue, K., Morita, N. and Matsubara, S..
Automatic Extraction of Task Statements from Structured Meeting Content.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 1: KDIR*, pages 307-315
ISBN: 978-989-758-158-8

2 RELATED WORK

There is not much research on the method of extracting useful information from the minutes of face-to-face meetings. The reason is that it is very costly to record all statements in meetings as text and to maintain them to analyze using a machine learning technique. We have solved this problem by the development and deployment of a variety of specialized tools.

To extract useful information, such as advice and requests, by using the records of the communication in online forums has been widely studied. Since it is common in terms of performing information extraction from the text of communication records, we describe the relevance of these studies to ours in this section.

Extraction of request representations from public comments embedded in online meetings organized by government agencies has been conducted (Kanayama and Nasukawa, 2008). Our proposed method covers not only requests by participants but also the presenter's responses and comments on future tasks.

Extracting contexts and answers of questions from the online travel forum “TripAdvisor” by using a structural support vector machine (SVM) was conducted (Yang, Cao and Lin, 2009). Since a target of this research was to assign the labels Context, Question, and Answer to each of the conversational sentences with the proposed method, it seems to be difficult to directly apply the method to task statement extraction. Assuming that if there is a statement that indicates the emergence of a task statement, the proposed method may be applied to our task extraction problem.

A rule-based approach to information extraction from online discussion boards was studied (Sarencheh et al., 2010). Some discussion boards are created with software such as SMF, phpBB, and vBulletin. The authors of that study developed a rule-base that includes rules regarding the relationships between the discussion structure and article content formatted in HTML tags. Since these rules are customized for each forum creation software and several versions, the versatility of the proposed method is not high with this approach. To increase the accuracy of task statement extraction, it is conceivable to combine machine learning and rule-based approaches.

Qu and Liu (Qu and Liu, 2012) investigated sentence dependency tagging of question and answer (QA) threads in online forums. They defined the thread tagging task as a two-step process. In the first step, sentence types (they defined 13 types such as Problem, Answer, and Confirmation) are labelled. In the second step, dependencies between sentences are tagged. With our approach, discussions are tagged manually by speakers during a meeting in a very convenient way and there is no need to consider all statements and their relationships.

Wicaksono and Myaeng (Wicaksono and Myaeng, 2013) provided a methodology for extracting advice-revealing sentences from online travel forums. They identified three different types of features (i.e., syntactic, context, and sentence informativeness) and proposed a hidden Markov model (HMM)-based method for labelling sequential sentences. Their features are similar to ours. Since the structure of the discussion is determined in advance of information extraction, our approach is easier to use than extracting the advice sentences from general online forums.

3 DISCUSSION MINING SYSTEM

Our discussion mining system promotes knowledge discovery from the content of face-to-face meeting discussions. Based on the meeting environment shown in Figure 1, multimedia minutes are generated for meetings in real time semi-automatically and linked with audiovisual data. The discussions are structured using a personal device called a “discussion commander” that captures relevant information. The content created from this information is then viewed using a “discussion browser,” which provides a search function that enables users to browse the discussion details.

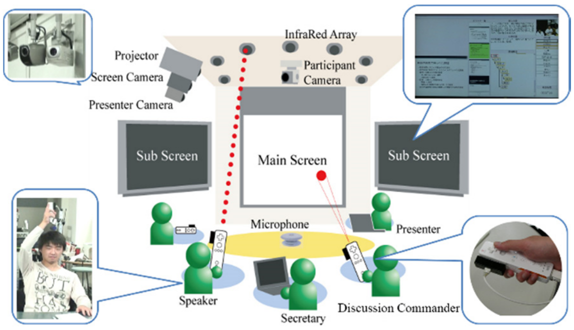

Figure 1: Overview of discussion mining system.

3.1 Recording and Structuring Discussions

Meeting discussions are automatically recorded, and the content is composed of structured multimedia data including text and video. The recorded meeting content is segmented on the basis of the discussion chunks. The segments are connected to visual and auditory data corresponding to the segmented meeting scenes.

Previous studies on structuring discussions include the issue-based information system (IBIS), graphical IBIS (gIBIS) (Conklin and Begeman, 1988), and argument diagramming of meeting conversations (Rienks, Heylen and van der Weijden, 2005), which take into account the structures of semantic discussions. However, such a semantic structure of discussions is still not at a practical level, and most studies on technology for generating discussion minutes have focused on devices, such as meeting recorders (Lee et al., 2002), for automatically recognizing auditory and visual data.

Our system uses natural language processing to not only support comprehension of the arguments in discussions but also form diversified perspectives using auditory and visual information in slides as well as other presentation media. It uses metadata to clarify the semantic structures of discussion content. Overall, our discussion mining system supports the creation of minutes of face-to-face meetings, records meeting scenes with cameras and microphones, and generates meta-information that relates elements in the meeting content.

In addition, the system graphically displays the structure of the discussions to facilitate understanding of the meeting content; therefore, improving the effectiveness of statements made during the discussions. The discussion commander has several functions for facilitating discussions, including one for pointing to and/or highlighting certain areas in presentation slides and one for underlining text in the slides displayed on the main screen.

Each statement is one of two types: “start-up” and “follow-up.” The “start-up” type is assigned to a statement that introduces a new topic, while the “follow-up” type is assigned to a statement that is on the same topic as the previous statement (i.e., it inherits the predecessor’s topic). Each discussion chunk begins with a start-up statement, as shown in Figure 2. Speakers are required to manually associate their statements with these attribute types with their discussion commanders when they start speaking during a meeting.

Real-time visualization of the discussion structure and visual referents (pointed texts and images) facilitate the current discussion. Moreover, the discussion structures can be modified by changing the parent nodes of the follow-up statements and by referring again to previous visual referents. A participant can perform these modifications by

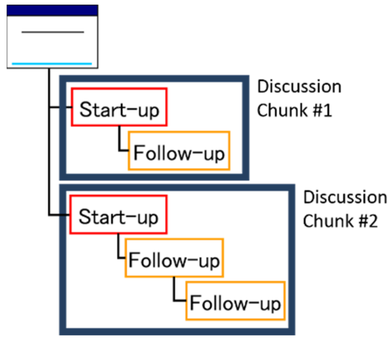

Figure 2: Discussion structure.

using his or her discussion commander. The participant can also use the discussion commander for marking the current statement by pressing the marking button. When these buttons are pressed, the system records who pressed the button and the target statement. Presenters mark the statement that they want to check later during the meeting and retrieve the marked statements by using the discussion browser mentioned in the next subsection.

3.2 Discussion Browser

The information accumulated with the discussion mining system is presented synchronously in the discussion browser with the timeline of the corresponding meeting, as shown in Figure 3. It consists of a video view, slide view, discussion view, search menu, and layered seek bar. The discussion browser provides a function for searching and browsing details about the discussions. For example, a participant can refer to a certain portion of a preceding discussion by doing a search using keywords or speaker names then browsing the details of the statements in the search results.

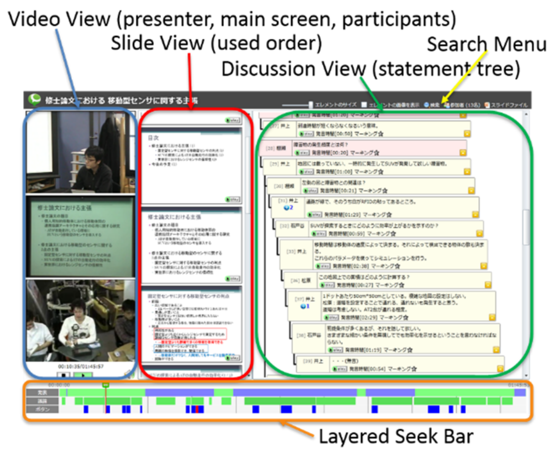

Figure 3: Discussion browser interface.

People who did not participate in the meeting can search and browse the important meeting elements displayed in the layered seek bar by searching for statements in discussions that were marked using a discussion commander or by surveying the frequency distributions of keywords.

The discussion browser has the following five components:

(1) Video view
The video view provides recorded videos of the meeting, including the participants, presenter, and screen.

(2) Slide view
Thumbnail images of presentation slides used in the meeting are listed in this view. The images are placed in the list in the order in which they are displayed on the main screen.

(3) Discussion view
The discussion view consists of text forms in which the content of the minutes primarily constitute information input by a secretary and relationship links, which visualize the structure of the discussions.

(4) Search menu
Three types of search queries are available in the search menu: speaker name, search target (either the content of a slide, a statement, or both), and keywords. The search results are shown in the layered seek bar (matched elements in the timeline are highlighted) and in the discussion view (discussions where the matched elements appear are highlighted).

(5) Layered seek bar
The elements comprising the meeting content are displayed in the layered seek bar. Various bars are generated depending on the element type.

4 AUTOMATIC EXTRACTION OF TASK STATEMENTS

Remembering past discussion content helps us to seamlessly carry out future activities. For example, in laboratory seminars, presenters can remember suggestions and requests about their research activities from the discussion content recorded in detail. The meeting content contains useful information for the presenters, but it is burdensome to read the information. Necessary information is concealed in a large amount of statements, so it is not easy to find. It is problematic if past discussions are not being reviewed, even for other speakers not only presenters. Therefore, it is necessary to extract the information concerning unsolved issues from previous discussions. We call statements including future tasks "task statements."

Our proposed method determines whether the statements are about future tasks (i.e., task statements). Some attributes including linguistic characteristics, structures of discussions, and speaker information are used to create a probabilistic model.

4.1 Model of Task Statements

A task statement can include any of the following content:

1. Proposals, suggestions, or requests provided during the meeting
The presenter has determined that they should be considered.
2. Problems to be solved
The presenter has determined that they should be solved.
3. Tasks not yet carried out before the meeting
Sometimes the presenter has already noticed them.

Candidates of task statements are fragments of a discussion chunk, as mentioned earlier. A typical discussion chunk is made from one or more questions and comments of the meeting participants and the presenter's responses to them. A coherent piece of discussion content related to tasks consists of questions/comments and their responses. Thus, "participants' questions/comments + presenter's response" is a primary candidate and a target of retrieval. "Participants' questions/comments and no response" is a secondary candidate.
Figure 4 shows example candidates of task statements.

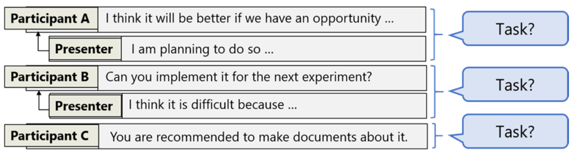

Figure 4: Candidates of task statements.

By using the correct data that were manually created from past meeting content, the method generates a probability model by using the maximum entropy method. For each candidate, the method calculates the probabilities of candidates of a task statement using the generated probabilistic model. A candidate whose probability value exceeds a certain threshold (e.g., 0.5) is extracted as a task statement.
Figure 5 shows the overall process of extracting task statements.

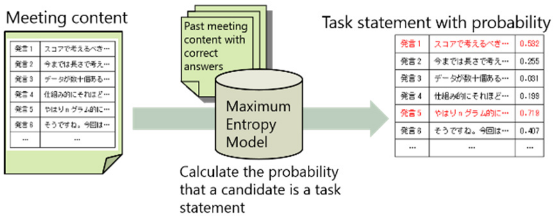

Figure 5: Overall process of extraction.

To discover the characteristics as clues to extracting task statements, some past meeting content was manually analyzed. The survey data included 11 types of meeting content and 598 groups of statements (candidates). Each presenter of the meeting manually selected task statements from each type of content.

As a result of manually extracting task statements of the survey data, 246 task statements were found corresponding to 41.1% of all candidates. By comparing the percentages of task statements, we analyzed the characteristics of the task statements.

For example, the attribute of speakers of statements and percentages of task statements to which the attribute contributed are listed in Table 1. Statements of teachers had a higher percentage of task statements overall. Therefore, the speaker attribute is helpful for calculating probabilities of task statements.

Table 1: Speaker attributes and percentages of task statements.

Participant attribute	Percentage of task statements
Teacher	52.9
Student	35.8
Total	41.1

As mentioned earlier, presenters use their discussion commanders to mark statements that they want to check later during the meetings. We investigated the effect of marking for discrimination of task statements by calculating the percentage of marked task statements of all task statements. The percentage of marked task statements was 73.4%, which was higher than that of the task statements for all candidates.

To examine whether there is a characteristic tendency in the number of characters in task statements, we obtained a distribution of the respective characters of a presenter's and participants' statements. We divided the number of characters into five groups and calculated the percentages of task statements in each group. In the participants' statements, the percentage of task statements increased when the number of characters increased. This is because when the participants were giving concrete requests and advice, the number of characters of their statements increased. On the other hand, in the presenter's statement, the number of characters of a higher percentage of task statements was 20 or less. The more characters there are the smaller percentage of task statements. It is believed that if the presenter accepts the requests or advice participants presented, his or her response would tend to be brief.

We also investigated certain types of sentences included in the task statements. In the participants' statements, the percentage of task statements was higher when sentences were in the present tense and in the declarative form (56.1%). This was due to the fact that a large amount of advice or requests were in the pattern of "should be ..." or "I want to …" In the presenter's statement, the percentage of task statements in the past tense and in the declarative form was low (29.2%). This is because when the presenter talked about future tasks, he or she did not tend to use sentences in the past tense. In addition, the percentage of task statements of the presenter in the past tense and in the interrogative form was 0%.

The details of the statistics of sentence types are presented in Table 2.

Table 2: Sentence types and percentages of task statements.

Sentence type characteristics of candidates	Percentage of task statements (Presenter)	Percentage of task statements (Participants)
Including declarative past tense sentence (DecPast)	29.2	42.3
Including declarative present tense sentence (DecPresent)	49.1	56.1
Including interrogative past tense sentence (IntPast)	0	32.4
Including interrogative present tense sentence (IntPresent)	51.6	35.5
No DecPast included	44.7	41.1
No DecPresent included	14.7	29.3
No IntPast included	42.4	41.7
No IntPresent included	41.7	49.4

The start time and duration of statements were also considered as characteristics to discriminate task statements. To determine the distribution of the start time of the participants' statements, we divided the entire meeting time into five intervals, each consisting of 20% of the meeting. We then determined the percentage of the task statements in each interval. At the 0-20% interval, the percentage of task statements was smaller. We assumed that this was because there were more questions than advice and requests in the early stages of the meetings. At the 20-40% and 80-100% intervals, the percentage of task statements was higher. That is, at the middle interval of the meeting, suggestions and advice about the purpose and approach were given, and at the final interval, future issues were presented as a summary of the entire meeting.

Morphemes and collocations of morphemes in statements are also important features. We generated a morpheme bigram of nouns, verbs, adjectives, and auxiliary verbs in the survey data by calculating the number of occurrences of the morphemes. We then determined a feature of morphemes and their bigrams of the statements if their occurrences exceeded certain thresholds. Specifically, the selected nouns had an occurrence percentage that was greater than or equal to 0.5% of all nouns. The selected verbs also had a percentage greater than or equal to 0.5% for all verbs. Morpheme bigrams were selected if their percentages were greater than 0.05% for the total morpheme bigrams. These selected morphemes and bigrams were used as features for discrimination of task statements.

Based on the above survey results, the following features were selected for creating a prediction model:

1 Attribute of presenter
2 Feature of participant's statement
  2.1 Start time and duration of statement
  2.2 Speaker type (teacher or student)
  2.3 Statement type (start-up or follow-up)
  2.4 Marking (0 or 1)
  2.5 Length (number of characters)
  2.6 Sentence types
  2.7 Morphemes and morpheme bigrams
  2.8 Response by presenter (0 or 1)
3 Feature of presenter's response
  3.1 Marking (0 or 1)
  3.2 Length (number of characters)
  3.3 Sentence types
  3.4 Morphemes and morpheme bigrams

For values of sentence type features, we used answers (0 or 1) to the following questions:

1. Does the statement include a sentence in the past tense and in the declarative form?
2. Does the statement include a sentence in the present tense and in the declarative form?
3. Does the statement include a sentence in the past tense and in the interrogative form?
4. Does the statement include a sentence in the present tense and in the interrogative form?
5. Does the statement include a sentence of the other type?

As mentioned earlier, a probabilistic model for extracting task statements is created using the maximum entropy method based on the above features. We used the Apache OpenNLP library (https://opennlp.apache.org/) for implementing this method.

Among the features used, morpheme, morpheme bigram, and sentence type are dependent on the language (in this case, Japanese). However, other languages, such as English, seem to have almost the same properties; therefore, it is necessary to analyze in detail.

4.2 Results of Task Statement Extraction

We give examples of task statements that were correctly extracted with the proposed method.

Example 1

Participant (regarding listeners' comments on a presentation rehearsal): Is it possible for presenters to ask their deep intensions of the comments?

Presenter: I think our system should deal with such situations.

The presenter expressed the intention to handle the requests from the participants. This task statement was not marked, so it was very difficult to find in the browsing of the meeting content.

Example 2

Participant (regarding self-driving cars): although a goal is to make vehicles run precisely according to their routes, I think it is difficult. So it is better to decide the acceptable range of the target route.

Presenter: We calculated an acceptable margin for each route, but there is a need to ascertain how far the vehicles deviated from the route.

The proposed method correctly extracted the description of the work the presenter should do and also the advice from the participant.

Example 3

Participant (regarding gamification): Because it is good that there is a sense of tension, I think it is better to reduce the goals and to achieve them repeatedly.

This statement is a proposal by the participant, but the presenter did not reply to it. Statements without a response from the presenter can also be extracted with the proposed method.

We also give examples of extraction failure.

Example 4

Participant (regarding document retrieval and summarization): Do you have any idea of summarization?

Presenter: No specific idea has been considered yet.

This is an example of statements that were not extracted despite a task statement. It can be considered to have failed in extracting because the concrete content of the task has not been stated.

Example 5

Participant (regarding user adaptation of an authoring tool): If you wanted to take the data as part of a rigorous evaluation, I think that you should have to do it exactly from start to end.

Presenter: I think that it is useless to say this now, I should do so.

This is an example of statements extracted as a task statement by mistake. Because it contains the auxiliary verb “should” indicating the meaning of duty and suitable, it can be considered to have been misclassified as a task statement.

By learning from these failures, we consider additional features if phrases such as “not considered” and “I think it should be …” are included in the candidates of task statements.

5 VERIFICATION OF EXTRACTION RESULTS

5.1 Experimental Results

To confirm the effectiveness of the proposed method, 10-fold cross-validation was applied to the extraction results. The data used for verification included 42 types of meeting content and 1,637 groups of statements (candidates). Each presenter created correct data of task statements in each type of meeting content as well as the survey data mentioned earlier. The data used for verification were totally different from the survey data.

We also compared the results of the proposed method to the extraction results of alternative methods that just select a set of statements that included any of the following features: (1) statements from teachers, (2) statements marked by presenters, (3) statements that have features (1) and (2).

We confirmed the effectiveness of the proposed method based on high precision (index for extraction accuracy), recall (index for extraction leakage), and F-measure (harmonic mean of precision and recall), as shown in Table 3.

Table 3: Experimental results.

	Proposed method	Marked or teachers' statements	Marked statements	Teachers' statements
Precision	0.758	0.543	0.689	0.500
Recall	0.642	0.441	0.239	0.326
F-measure	0.695	0.487	0.355	0.395

The extraction results of the task statements with the proposed method are as follows: precision was 75.8%, recall was 64.2%, and F-measure was 69.5%. On the other hand, the results of the three alternative extraction methods were as follows: selecting the statements that were marked by the presenter had the highest precision (68.9%), selecting the statements from teachers or statements that were marked by the presenter had the highest recall (44.1%) and F-measure (48.7%). The proposed method obtained the highest values compared to these other extraction methods.

Table 4 lists the results without certain features of the probabilistic model. The F-measure significantly decreased to 56.9% when the features of morphemes and morpheme bigrams were not used. Since the F-measures of the methods lacking any features were reduced, the validity of the features used with the proposed method was confirmed.

Table 4: Experimental results without features.

Feature	F-measure
All	0.695
No Presenter Attribute	0.679
No Start Time/Duration	0.689
No Speaker Attribute	0.694
No Statement Type	0.693
No Marking Info.	0.687
No Statement Length	0.693
No Sentence Type	0.686
No Morpheme and Morpheme bigram	0.569
No Response Info.	0.693

As mentioned earlier, the proposed method calculates the probabilities of candidates of a task statement using the generated probabilistic model. A candidate whose probability value exceeds a certain threshold is extracted as a task statement.

We first set the threshold to 0.5. It is not guaranteed that this threshold value is optimal. Therefore, we re-evaluated the outputs of the system by lowering the threshold by 0.1 from 0.5. The results are listed in Table 5.

Table 5: Experimental results with different thresholds.

Threshold	Precision	Recall	F-measure
0.5	0.758	0.642	0.695
0.4	0.670	0.765	0.714
0.3	0.596	0.885	0.711
0.2	0.496	0.965	0.656
0.1	0.417	0.999	0.591

We found that the F-measure at a threshold of 0.4 was highest (71.4%). In the future, it should be conducted to extract task statements by setting a threshold to 0.4.

Since we used the maximum entropy method as a classifier, we also confirmed that this method works better than other classifiers. Table 6 shows the results of the SVM and naive Bayes classifiers. We used the “kernlab” package for the SVM and the “e1071” package for the naive Bayes of R language.

Table 6: Experimental results with alternative classifiers.

Classifier	Precision	Recall	F-measure	Difference
SVM	0.767	0.626	0.690	-0.024 ~ -0.005
Naive Bayes	0.656	0.686	0.671	-0.043 ~ -0.024

The difference column in this table shows the differences between the F-measures of the subject classifier and the maximum entropy method in the case of 0.4 (best performance) and 0.5 (initial setting) thresholds. We found that our method is slightly better that other traditional classifiers. While the performance of the maximum entropy method did not have a very significant advantage, the results obtained as probability values can contribute to flexible control of the presentation of results by using techniques such as sorting and filtering.

5.2 Permutation Test

Since comparison with simple baselines (i.e., teachers’ statements and marked statements) is not sufficient for proving the reliability of the proposed method, we require another technique for this proof.

As well as evaluating the performance of the proposed method, we also determined if the results are statistically reliable. Therefore, we conducted a statistical hypothesis test regarding the misclassification rate calculated from cross-validation. The statistical hypothesis test is a kind of contradiction to prove the significance by rejecting a hypothesis in which a complementary event of the hypotheses is to be clarified. Since the correctness of some results is generally difficult to prove directly, the concept based on this contradiction is used in the statistical hypothesis test.

In cross-validation, the misclassification rate is calculated by

$$\frac{\text{number of misclassifications (the candidates are misclassified into task statements)}}{\text{number of discriminations}}.$$

Even though this statistical hypothesis test is based on the misclassification rate and unknown null distribution, it is possible to estimate the null distribution in a nonparametric manner by using a permutation test (Good, 1994). In a permutation test, a sample of a label is repeated many times to be sorted randomly (here a label corresponds to whether it is a task statement), and a null distribution is virtually constituted. A ratio of statistics in this manner is produced for each permutation that becomes equal to or less than the value of the original test statistics. The ratio is called a p-value, which is a measure of the probability of events observed under the null hypothesis. When the p-value is less than the significance level that was set in advance, the observed events under the null hypothesis do not occur by chance, that is, the null hypothesis is rejected. Then, we use an alternative hypothesis in which the prediction model is statistically significant.

In this experiment, we set the significance level to 0.05 and conducted a permutation test from 1,000 iterations. The results are listed in Table 7. The misclassification rate was 0.2260, and the p-value for this was less than 0.001. It was confirmed that our probabilistic model of task statements is statistically significant below the level of significance. Since a p-value is calculated from 1000 iterations, its accuracy will rise in 0.001 increments. In other words, the actual p-value is also considered much less likely than 0.001.

Table 7: Results of permutation test.

Number of misclassifications	Number of discriminations	Misclassification rate	p-value
370	1637	0.2260	< 0.001

6 FUTURE WORK

Future work includes improvement in the accuracy of our proposed method and in the usability of our application system. We are considering the use of the sentence end representation of statements and planning to enhance the application system to automatically generate a summary statement indicating the content of the task from a set of statements and to send feedback of users’ quotation data of the task statements to the extraction module for modification of the probability model.

Future work also includes creating a more semantic structuring of discussions. In particular, we aim to develop a system that can automatically determine to what extent a discussion proceeds depending on the topic. For example, if the topic in a discussion chunk changes, the system should sub-divide the chunk accordingly and determine whether the previous topic is convergent.

Our previous study revealed that some follow-up statements were about a topic different from that of the start-up statement (Tsuchida, Ohira and Nagao, 2008). The discussion may thus become unsettled and be abandoned because the participants do not know whether the discussion on the previous topic reached a conclusion. We may be able to develop a mechanism that can automatically identify such unsolved topics and suggest that participants discuss them again.

7 CONCLUSIONS

We proposed an automatic extraction method of task statements from meeting content. With 10-fold cross-validation and permutation test, we evaluated the effectiveness and reliability of the proposed method. We also compared the results with those from alternative methods without certain features and confirmed the validity of the features used with the proposed method.

Although our discussion mining system is able to record face-to-face meetings in detail, analyze their content, and conduct knowledge discovery, it is unable to structure the discussions so that the topic of each discussion is classified. To overcome this problem, we aim to achieve more semantic structuring of discussions by deeply analyzing linguistic characteristics of statements and by applying certain machine learning techniques.

REFERENCES

Sellen, A. J. and Whittaker, S., 2010, Beyond Total Capture: A Constructive Critique of Lifelogging. *Commun. ACM*, vol. 53, no. 5, pp. 70–77.

Mayer-Schönberger, V., Cukier, K., 2013. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt.

Nagao, K., Kaji, K., Yamamoto D. and Tomobe, H., 2005. Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities, *Advances in Multimedia Information Processing – PCM 2004*, LNCS, vol. 3331, pp. 522–531. Springer.

Wu, N., 1997. *The Maximum Entropy Method*, Springer Series in Information Sciences, 32, Springer.

Kanayama, H. and Nasukawa, T., 2008. Textual Demand Analysis: Detection of User's Wants and Needs from Opinions, In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-2008)*, pp. 409–416.

Yang, W.-Y., Cao, Y. and Lin, C.-Y., 2009, A Structural Support Vector Method for Extracting Contexts and Answers of Questions from Online Forums, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 514–523.

Sarencheh, S., Potdar, V., Yeganeh, E. A. and Firoozeh, N., 2010, Semi-automatic Information Extraction from Discussion Boards with Applications for Anti-Spam Technology, In *Proceedings of ICCSA 2010*, LNCS, vol. 6017, pp. 370–382. Springer.

Qu, Z. and Liu, Y., 2012, Sentence Dependency Tagging in Online Question Answering Forums, In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pp. 554–562.

Wicaksono, A. F. and Myaeng, S.-H., 2013, Automatic Extraction of Advice-revealing Sentences for Advice Mining from Online Forums, In *Proceedings of the 7th International Conference on Knowledge Capture (K-CAP 2013)*.

Conklin, J. and Begeman, M. L., 1988. gIBIS: A Hypertext Tool for Exploratory Policy Discussion, *ACM Transactions on Information Systems (TOIS)*, vol. 6, no. 4, pp. 140–152.

Rienks, R., Heylen, D., and van der Weijden, 2005. Argument Diagramming of Meeting Conversations, In *Proceedings of the Multimodal Multiparty Meeting Processing Workshop at the 7th International Conference on Multimodal Interfaces (ICMI 2005)*.

Lee, D., Erol, B., Graham, J., Hull, J. and Murata, N., 2002. Portable Meeting Recorder, In *Proceedings of ACM Multimedia 2002*, pp. 493–502.

Good, P., 1994. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypothesis*, Springer.

Tsuchida, T., Ohira, S. and Nagao, K., 2008. Knowledge Activity Support System Based on Discussion Content, In *Proceedings of the Fourth International Conference on Collaboration Technologies*.