Evidence-Based Education: Case Study of Educational Data Acquisition and Reuse

Katashi NAGAO
Graduate School of Informatics, Nagoya University
Nagoya, Japan

Naoya MORITA
Graduate School of Information Science, Nagoya University
Nagoya, Japan

Shigeki OHIRA
Information Technology Center, Nagoya University
Nagoya, Japan

ABSTRACT

There must be as many concrete indicators as possible in education, which will become signposts. People will not be confident about their learning and will become confused with tenuous instruction. It is necessary to clarify what they can do and what kinds of abilities they can improve. This paper describes a case of evidence-based education that acquires educational data from students' study activities and not only uses the data to enable instructors to check the students' levels of understanding but also improve their levels of performance. Our previous research called discussion mining was specifically used to collect various data on meetings (statements and their relationships, presentation materials such as slides, audio and video, and participants' evaluations of statements). This paper focuses on student presentations and discussions in laboratory seminars that are closely related to their research activities in writing their theses. We propose a system that supports tasks to be achieved in research activities and a machine-learning method to make the system sustainable for long-term operation by automatically extracting essential tasks. We conducted participant-based experiments that involved students and computer-simulation-based experiments to evaluate how efficiently our proposed machine-learning method updated the task extraction model. We confirmed from the participant-based experiments that informing responsible students of tasks that were automatically extracted on the system we developed improved their awareness of the tasks. Here, we also explain improvements in extraction accuracy and reductions in labeling costs with our method and how we confirmed its effectiveness through computer simulations.

Keywords: Evidence-Based Education, Educational Data Mining, Machine Learning, PDCA Cycle, Discussion Mining

1. INTRODUCTION

One issue with students' research activities in their theses at Japanese university research laboratories (labs) is that students are expected to start research as soon as they have been assigned to a lab, even though they generally have no familiarity with the research process. This poses difficulties for students as they often have little experience in the various elements required for successful research, such as long-term planning for one's project and successfully addressing and managing the various issues that can occur in the research process.

We now live in an information society, and information systems have been developed that assist humans in various areas. It is not surprising that education is one of these areas as well. Some specific types of educational-support systems that are in use are referred to as "e-learning", which records learning progress and evaluates work with an e-portfolio; another type is called Massive Open Online Courses (MOOCs), which are implemented via the Internet, and anyone can sign up to attend such courses. These systems all make use of IT systems. However, "data mining" systems that enable analysis of the accumulated data are still currently being developed [1]. The primary reason for this is that reusing data accumulated from human intellectual activities generally has a large cost associated with it. Additionally, the various types of educational-support systems that have been developed were not designed for data mining.

Of the various types of intellectual activities, discussion plays an important role in human society as it allows groups to arrange their thoughts, which helps in the resolution of problems and the development of policies. Seminars are important for university research labs as they provide periodic opportunities for the exchange of ideas with regards to research and a way to discuss issues that may occur in a student's future research. However, the particular issues that are thought to arise are not recorded in detail, which makes it difficult to use the information for practical purposes. This is why our lab is in the process of developing and using "discussion mining" (DM) technology [2], which records all metadata that we develop in the process of our research lab's interview-style presentations in great detail. This system helps us immensely in reflecting on the content of discussions that occur during our meetings. The act of reflecting on various issues is an important activity as it allows one to change his/her goals as he/she progresses.

It takes time to educate people, but their abilities to conduct intellectual activities, such as those in research, need to be firmly acquired based on long-term perspectives. We must provide clear instruction to provide signposts. Students will lose confidence in their studies with speculative guidance. The technique we developed is useful for clarifying what to do and what kinds of abilities it improves. "Evidence-based education" will be made possible by using such a data-oriented learning mechanism. Learning analytics (LA) is an emerging field in which sophisticated analytic tools are used to improve learning and education [3]. It draws from, and is closely tied to, a series of other fields of study including business intelligence, web analytics, academic analytics, and educational data mining. While LA covers the measurement, collection, analysis and reporting of data about learners and their contexts, detailed data on discussions at meetings and activities based on them have not been accumulated and analyzed in LA.

However, there is a risk that there can be too much information recorded that one can get buried in the details. This is why we focused on discussions at seminars that were closely related to the research activities of our students and carried out data mining on the content of discussion and used a type of machine-learning that is often referred to as active learning [4]. The combination of these two methods provided a support system to our students that helped them to address issues as they arose and propose practical long-term planning. We carried out tests that involved actual college students with regards to graduation and the completion of their research to test the effectiveness of the system. We also conducted a computational experiment with regards to increasing the sampling accuracy of our active-learning method.

2. EDUCATIONAL DATA ACQUISITION: EXTRACTION OF TASK STATEMENTS FROM MEETING MINUTES

2.1. Recording and Structuring Discussions

The DM system that was previously described promotes knowledge discovery from the content of face-to-face meeting discussions. Multimedia minutes are semi-automatically generated from meetings in real time and linked to audiovisual data based on the meeting environment in Figure 1. 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.

Two kinds of tags are applied to statements; the first is a tag called a start-up to introduce new topics, and the second is a tag called a follow-up to continue the topic already being discussed. It is necessary for the follow-up statement to clarify from which statement it is continuing. Every participant inputs metadata about his/her speech using his/her "discussion commander", as outlined in Figure 1.

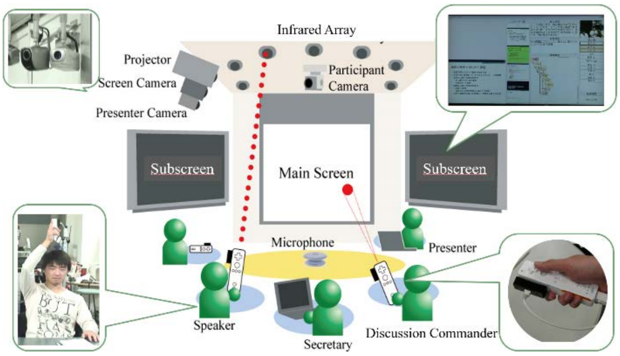


Figure 1: Overview of discussion mining system.

Participants who specifically ask questions or make comments on new topics assign start-up tags to their statements. Also, if they want to speak in more detail on topics related to the immediately preceding statement, they provide a follow-up tag. Furthermore, the system records pointer designates the location/time for the slide and information on the button for or against the statement during the presentation and during the question and answer session. Marking information on important statements is also recorded. Meeting discussions are automatically recorded, and the content is composed of structured multimedia data that include text and video. The recorded meeting content is segmented on the basis of discussion chunks. The segments are connected to visual and auditory data that correspond to the segmented meeting scenes in Figure 2.

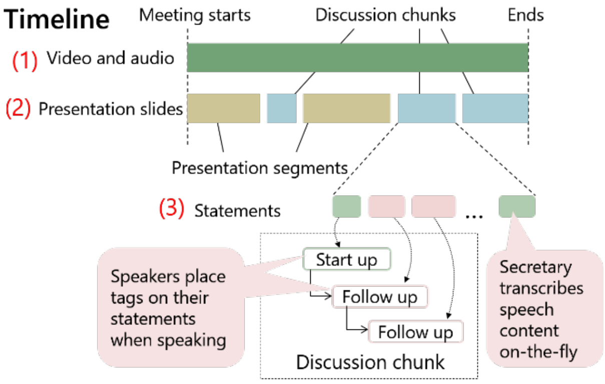


Figure 2: Structured meeting content.

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

We developed a machine-learning method of statistically determining whether the statements were about future tasks (i.e., task statements) [5]. Some attributes including linguistic characteristics, structures of discussions, and speaker information were used to create a probabilistic model.

2.2. Model of Task Statements

A task statement can include any of three types of 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 the problems that should be solved.

3. Tasks not yet carried out before the meeting; sometimes the presenter has already identified such tasks.

Candidates’ task statements are fragments of a discussion chunk that was described earlier. A typical discussion chunk is created from one or more questions and comments from 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 3 outlines an example of candidates for task statements.

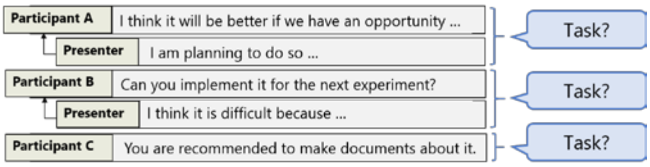

Participant A
I think it will be better if we have an opportunity ...
Presenter
I am planning to do so ...
Task?
Participant B
Can you implement it for the next experiment?
Presenter
I think it is difficult because ...
Task?
Participant C
You are recommended to make documents about it.
Task?

Figure 3: Candidates for task statements.

The method generates a probability model by using logistic regression analysis by using correct data that were manually created from past meeting content. The method calculates the probabilities for individual candidates for 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 4 outlines the overall process for extracting task statements.

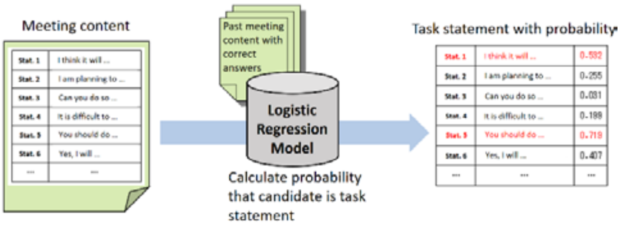


Figure 4: Overall process of extraction.

Ten-fold cross-validation was applied to the extracted results to confirm the effectiveness of the proposed method. The data used for verification included 42 types of meeting content and 1,637 groups of statements (candidates). Each presenter created correct data for task statements for each type of meeting content. We confirmed the effectiveness of the proposed method in terms of high precision (index for extraction accuracy), recall (index for extraction leakage), and the F-measure (or F1 score that is the harmonic mean of precision and recall).

The extracted results obtained for the task statements with the proposed method were precision of 75.8%, recall of 64.2%, and an F-measure of 69.5%. Since we used the logistic regression model as a classifier, we also confirmed that this method works better than other classifiers, such as support vector machines (SVMs) and naïve Bayes classifiers. The best values of the F-measures for the SVM and naïve Bayes were 69.0% for the former and 67.1% for the latter. We found that our method was slightly better that of other traditional classifiers.

2.3. Issues with Task-Statement Extraction

2.3.1. Cost of giving teacher signals

Generating an extraction model of task statements is based on supervised learning, where a machine learns from humans who provide a discrimination class of teacher signals to the machine, and the extraction results depend on the training data set on which machine learning was executed. Therefore, it is necessary to increase the amount of training data to improve extraction accuracy especially when the amount of usable data is relatively small, as it was in this research; however, the assignment of teacher signals to the statements of all minutes recorded by DM is very costly. In addition, it is preferable for the presenter who best understands the content of the presentation to be in charge to minimize the number of misjudgments from teacher signals, i.e., it is also a task that requires specific human knowledge. Teacher signals of task statements cannot easily be generated.

2.3.2. Feature changes of task statements over time

Another issue in supervised learning is feature changes in objects to be extracted over time. There is no problem if the characteristics of the extraction target are completely invariant, but as new students enter the lab and progress with research activities transforms each year, the characteristics of task statements to be extracted change over time.

The problem with feature changes over time has often been discussed. A spam-mail filter is a good example. The techniques of spam mail are becoming increasingly more sophisticated, and unless the discrimination model is updated, it cannot adapt to the characteristics of new spam mail, which and decreases the accuracy of discrimination [6].

One concern with text minutes in DM is that there are differences in wording due to there being different secretaries who manually input text. Since strict rules are not defined for secretaries, the wording depends on the discretion of each person in charge, and the degree of sentence summarization also differs. Since the language information included in statements mostly affects extraction accuracy in the task-statement-extraction model, it is necessary to focus on feature changes due to differences in wording.

2.3.3. Solution by active learning

The discriminant model should always be updated when the amount of data to be analyzed is increased. However, it is very difficult to label all data when there is an increasing amount of data, and labeling incurs large costs. We used active learning [4, 7, 8, 9] to solve this problem, which was used to attempt to improve extraction accuracy for the limited data set we obtained.

Active learning can be implemented as an algorithm in which a sample that makes the greatest contribution to updating a discrimination model from a large number of samples without teacher signals is selected by a machine, and a teacher signal is given to it to minimize human effort. This process efficiently updates models [4]. All statements in task-statement extraction in the minutes of a seminar correspond to samples without teacher signals; some groups of statements are selected using an active-learning method, and the selected statement group is assigned teacher signals by students who are presenters at the corresponding seminar. The research-activity-support system that will be described later encourages students to reflect on the seminar after their presentations and obtain feedback on the automatically selected task statements.

Applying this active-learning method to task extraction in this way both simultaneously solves problems of cost in teacher-signal assignment and feature changes in task statements over time.

3. EDUCATIONAL DATA REUSE: ENCOURAGEMENT OF TASK ACHIEVEMENT IN RESEARCH ACTIVITIES

3.1. Support for Task Achievement

It is not easy to steadily achieve various tasks that arise during work and research. Indeed, there are many arguments on how to manage task achievement that will lead to success, and many scheduling-support systems, such as Google Calendar, have been developed to support this.

Most conventional scheduling-support systems have only focused on the schedule management of a plan, but not on understanding how the established schedule is processed and what state it is in before going onto the next task. There is no support for the implementation of tasks that have causal relationships over the long term, such as setting guidelines. As a major issue with graduation and completed studies is addressed by carrying out individualized tasks that have been segmentalized in a long-term time series, the plan-do-check-act (PDCA) cycle in business execution is recommended for education as well. It is necessary to have a mechanism to support the smooth execution of a series of performance cycles of planning, execution, and evaluation of tasks.

In addition, although it is necessary to be aware of the existence of all tasks to be scheduled by students as a prerequisite for conventional efforts related to task-achievement support, we aimed at advanced support that included awareness of the tasks to be carried out.

3.2. Reflection on Discussions at Seminars

The seminars regularly held at university labs are meetings where opinions are exchanged on the content of research and they include remarks that will become future issues (i.e., task statements). As the discussions at seminars during research activities are generally not recorded in detail, they are difficult to use to discover problems, whereas DM enables the discovery of issues from such discussions by recording their content. However, the number of statements recorded at one seminar is enormous; therefore, it is very time consuming to be aware of tasks by individually checking all of them.

We previously proposed an automatic method of extracting task statements to solve this problem by using supervised learning with various metadata in DM and language information on statements as features. We applied this method to support students in being aware of the existence of tasks in this study and led them to completing subsequent tasks.

4. RESEARCH ACTIVITY SUPPORT SYSTEM

4.1. System Overview

4.1.1. PDCA cycle

The research-activity support system comprehensively supports research activities in general and conforms to the PDCA cycle recommended for general business operations [10].

The flow for task achievement is outlined in Figure 5, and the underlined parts are executed by the system. First, the user creates notes on a performance plan to achieve the tasks for the task statements that are extracted using the learned extraction model, and then manages the execution schedule with the scheduling component of the system. The user then carries out the tasks according to the planned schedule, is evaluated by other users (support for check step) based on the results added to the performance-plan notes, and considers the evaluations to improve subsequent activities.

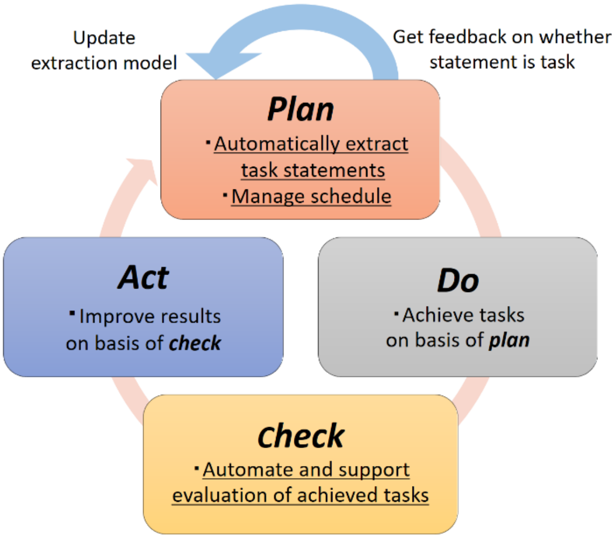


Figure 5: Flow of task achievement.

4.1.2. Presentation of tasks

We must begin by being aware of the problems that currently confront us to facilitate research activities. Task statements extracted from the proposed system are presented, as indicated in Figure 6.

Each statement that is presented is judged to be a task statement with a high degree of possibility, and it is necessary for the user to finally determine whether it should be achieved. Therefore, since the statement immediately after extraction has a blue icon marked "Task?", this icon can be clicked, and it can be determined whether the statement indicates a task to be achieved.

After the task is decided, the user creates a note citing the task statement and describes the details of the plan and results related to the task. The completion icon is displayed in the related task statement by setting the attribute of the note citing the statement to "Finished". The user can then easily be aware of which task is in the completed state.

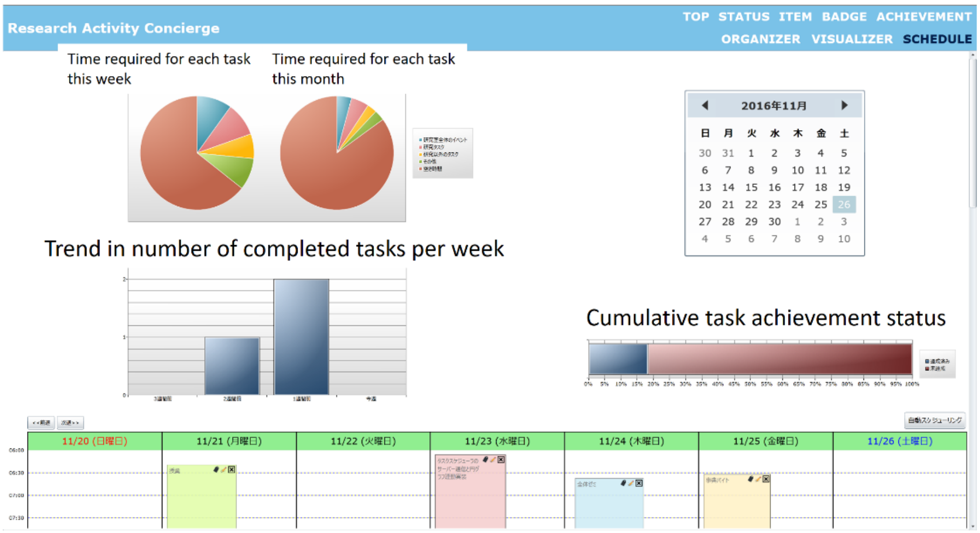
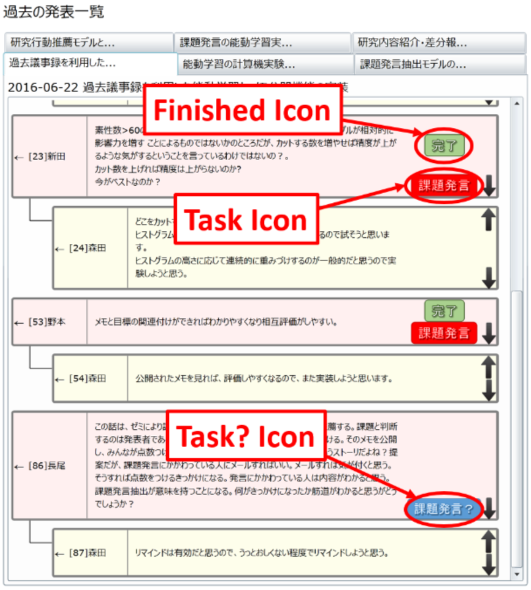

Research Activity Concierge
TOP STATUS ITEM BADGE ACHIEVEMENT
ORGANIZER VISUALIZER SCHEDULE
Time required for each task this week
Time required for each task this month
2016年11月
日 月 火 水 木 金 土
Trend in number of completed tasks per week
Cumulative task achievement status
11/20 (日曜日)
11/21 (月曜日)
11/22 (火曜日)
11/23 (水曜日)
11/24 (木曜日)
11/25 (金曜日)
11/26 (土曜日)
06:00
06:30
07:00
07:30

Figure 7: Task scheduler.


過去の発表一覧

2016-06-22 過去議事録

← [23]新田

素性数>60 ルが相対的に
影響力を増す ことによるものではないかのところだが、カットする数を増やせば精度が上が
るような気がするということを言っているわけではないの？。
カット数を上げれば精度は上がらないのか?
今がベストなのか？

← [24]森田

どこをカット
ヒストグラム るので試そうと思いま
す。
ヒストグラムの高さに応じて連続的に重みづけするのが一般的だと思うので実
験しようと思う。

← [53]野本

メモと目標の関連付けができればわかりやすくなり相互評価がしやすい。

← [54]森田

公開されたメモを見れば、評価しやすくなるので、また実装しようと思います。

← [86]長尾

この話は、ゼミにより 薦する。課題と判断
するのは発表者であ ける。そのメモを公開
し、みんなが点数つ うストーリだよね？提
案だが、課題発言にかかわっている人にメールすればいい。メールすれば 付くと思う。
そうすれば点数をつけるきっかけになる。発言にかかわっている人は内容がわかる 思う。
課題発言抽出が意味を持つことになる。何がきっかけになったか筋道がわかると思うがど
でしょうか？

← [87]森田

リマインドは有効だと思うので、うっとおしくない程度でリマインドしようと思う。

Figure 6: Presentation of tasks.

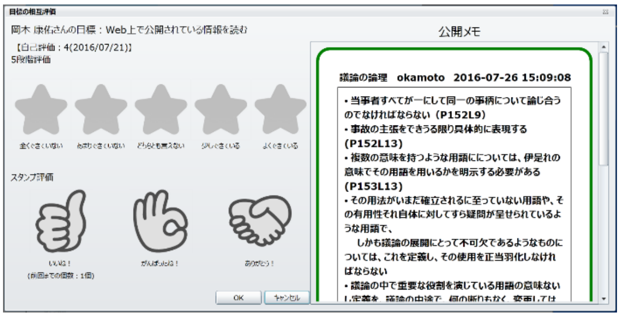
4.1.3. Time management of tasks

The state of awareness for students with little experience in research activities who only use the number of tasks is disadvantageous because of the uncertainty regarding execution. Therefore, after the existence of tasks is identified, each task should be well organized and scheduled to use time efficiently. We implemented a task scheduler (Figure 7) in the proposed system that could schedule the duration of task executions.

There is a graph at the top of the screen of the task scheduler that can roughly organize the proportion of information by type (e.g., surveillance, development, and experiment) of tasks and their situation with achievement. The user can arrange and update the time intervals of tasks scheduled to be executed on the time table at the bottom of the screen.

4.1.4. Evaluation of task-execution content

After the tasks are carried out, we should assess their



岡本 康佑さんの目標：Web上で公開されている情報を読む

【自己評価：4(2016/07/21)】

5段階評価

スタンプ評価

(前回までの個数：1個)

公開メモ

議論の論理 okamoto 2016-07-26 15:09:08

- 当事者すべてが一にして同一の事柄について論じ合うのでなければならない（P152L9）
- 事故の主張をできうる限り具体的に表現する(P152L13)
- 複数の意味を持つような用語にについては、伊是れの意味でその用語を用いるかを明示する必要がある(P153L13)
- その用法がいまだ確立されるに至っていない用語や、その有用性それ自体に対してすら疑問が呈せられているような用語で、
  しかも議論の展開にとって不可欠であるようなものについては、これを定義し、その使用を正当羽化しなければならない
- 議論の中で重要な役割を演じている用語の意味ないし定義を、議論の中途で、何の断りもなく、変更しては

Figure 8: Task-evaluation tool.

progress, comprehend the level of achievement, and use it as a source for future activities [11]. We implemented a function to evaluate task-execution content in the proposed system by obtaining other students' assessments of the content of execution by publishing task-execution notes in the lab.

We confirmed that mutual evaluation was reasonable by referring to the notes that described the content of task execution. However, since the task execution notes were primarily private, they are not supposed to be published. The sharing function of the notes provided an opportunity to receive a more accurate evaluation when there was motivation to tell others about the results of their task execution.

When completing task execution, the system accesses the level of achievement of the research goals. The user can then publish related research notes and associate them with the previously shared notes to motivate others to update evaluations. The associated notes are displayed on the right of the evaluation window, as seen in Figure 8, when receiving an evaluation from another user. The user considers them as a supplemental resource to evaluate another's research results.

4.2. Improvements in Accuracy of Task-Statement Extraction

4.2.1. Active learning

Active learning is a technique frequently used in several fields such as natural language processing and biostatistics since expert knowledge concerning the assignment of teacher signals is required and the costs of data collection and teaching-signal extraction are very high. The technique is also suitable for improving the accuracy of task-statement extraction in which the cost of teacher-signal assignment is also very high.

A five-step procedure is repeated to apply the active-learning technique to DM.

1. Create a task statement extraction model using a set of statements with teacher signals as to whether or not it is a task statement
2. After a presentation has been made at a seminar, carry out task-statement extraction by using the recently created minutes and calculate the probability value that each statement is a task statement
3. Select the target set of statements to which the teacher signal is to be assigned by active learning
4. Display the results for the selected statement set and ask the system to provide teacher-signal feedback to the user
5. Add the statement set with teacher signals that have been obtained to the training data of machine learning

By updating the task-statement-extraction model by repeating these five procedures, it is not only possible to improve the accuracy of extraction by increasing the amount of training data but also to constantly adapt to feature changes of task statements over time.

4.2.2. Reuse of large numbers of past minutes

Information density is a representative algorithm of the active-learning technique that takes into account density on the feature space as a sampling reference [12]. The strategy with this algorithm is to consider that there is a large amount of information with higher density data; therefore, the benefits of providing a teacher signal are considered to be great.

Since information on teacher signals is not used for density calculation, it is possible to use a large number of samples without teacher signals for learning. As the DM project has been continuing for about 10 years, and the accumulated content of discussion is huge, it is very convenient to apply this method.

Although the practicality of information density is high when there is a large amount of noise data, it may adversely affect the weighting of model parameters. Morpheme information on statements is used in the task-statement-extraction model, and extremely long or short statements can be noise. For example, short statements such as "I will consider it" and "I do understand" that frequently appear in responses by presenters have fewer morphemes that are selected as modeling features, and very long statements that have many morphemes also tend to be noisy samples.

Our proposed method used a weighting algorithm in this research to cope with such problems that was based on a histogram of feature information. Since all the features used in the task-statement-extraction model are binary variables, we can define a histogram in which the number of features that have the value one in a sample is classified as a class. We call it a feature histogram. Figure 9 plots the feature histogram created for the past 492 meeting minutes.

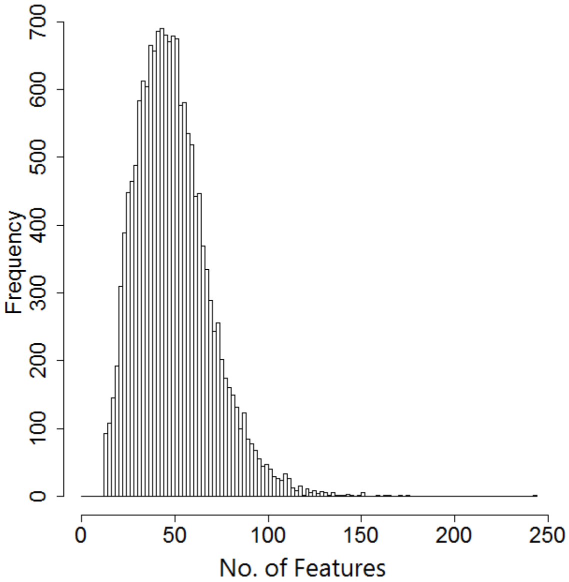


Figure 9: Feature histogram.

Our weighting algorithm was based on the ratio of frequency to the number of samples without teacher signals as a weight as:

$$\underset{x \in U}{\arg\max}\, \phi_x \times \left(\frac{1}{U} \sum_{i=1}^{U} sim(\boldsymbol{x}, \boldsymbol{x}_i) \times \frac{freq(bin(\boldsymbol{x}))}{|U|}\right). \qquad (1)$$

Here, ϕ_x is the score of sampling, U is the set of samples without teacher signals, and $sim(\boldsymbol{x}, \boldsymbol{x}_i)$ is the cosine similarity between $\boldsymbol{x}$ and $\boldsymbol{x}_i$. The $freq()$ is the frequency of a suggested bin and $bin(\boldsymbol{x})$ is the bin to which $\boldsymbol{x}$ belongs.

$\frac{freq(bin(x))}{|U|}$ is a new weighting and the other part of Formula (1) is the same as the calculation for information density. This weighting makes it possible to give a light weight to statements of extreme length that can have a negative effect on information density, and more effective sampling to improve extraction accuracy can be expected.

5. EVALUATION TESTS

5.1. Evaluation by System Operation

The breakdown for the students who participated was two 2nd year graduate students, three 1st year graduate students, and three undergraduate students. We randomly divided each group into an intervention group (with automatic task extraction and the presentation function) and a control group (without the function). The tests were again based on the spring (April to July) and autumn (October to December) semesters. Crossover comparison tests were conducted. The proposed system in Figure 9 represents the values in the intervention group, and the conventional system represents the values in the control group. The evaluation criterion was the task-awareness rate that indicated the extent of the task to be achieved.

The results are given in Figure 10. As a result of the t test for the task-awareness rate of the proposed system compared to the conventional system, the p value was 0.0481, and a significant difference was found at the significance level of 5%. Also, the differences were small for participants d and f whose task-awareness rates with the proposed system were lower than those with the conventional system. We confirmed that automatic task extraction and presentation with the proposed system enhanced awareness of tasks for students and contributed to their awareness of more tasks to be achieved.

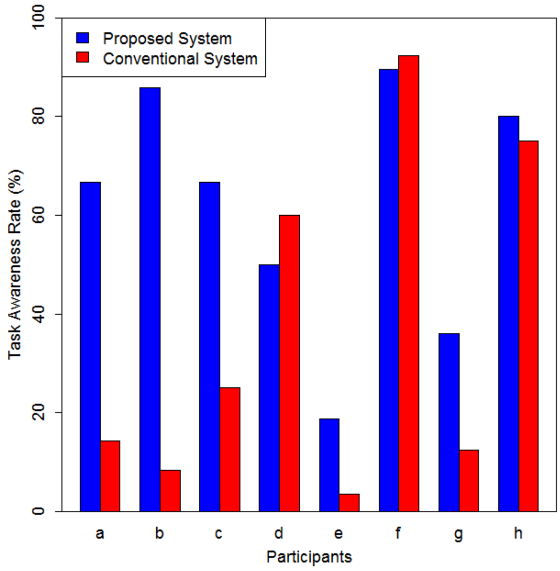


Figure 10: Task-awareness rate.

5.2. Computer Simulation Experiment on Improvements to Extraction Accuracy

Since teacher-signal feedback with the proposed system improved the accuracy of task-extraction, we performed a simulation of what kind of accuracy transition could be observed depending on different data-sampling methods for active learning for a situation in which ten statements (for full sampling of all statements) were added as training data when one minute of the meeting was created. The transitions in the value of the F-measure (or F1 score that is the harmonic mean of precision and recall) were compared with six methods in a seven-fold cross validation (Figure 11) for the data of the minutes recorded by the DM system (data group: 1,637).

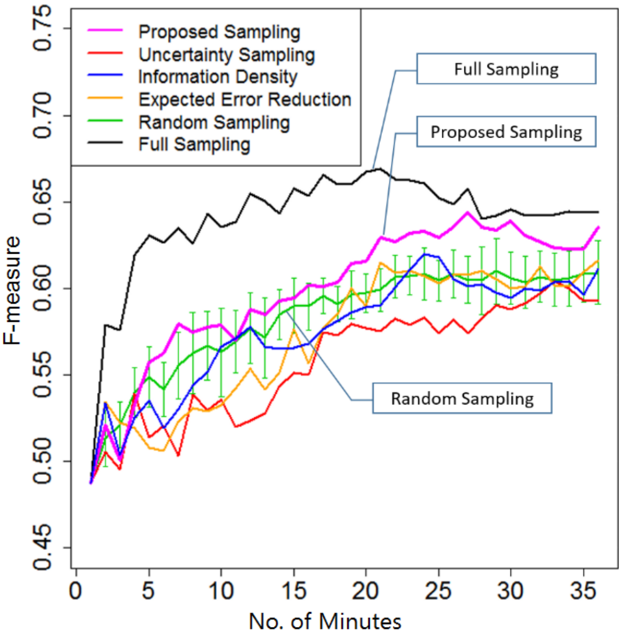


Figure 11: Comparison of accuracy transition.

- Proposed sampling: Formula (1) (proposed method)
- Uncertainty sampling: Conventional method [13]
- Information density: Conventional method [12]
- Expected error reduction: Conventional method [14]
- Random sampling: Statements are randomly selected for sampling
- Full sampling: All statements are selected to assign teacher signals

Full sampling indicates the limits of improving extraction accuracy as a reference value, where the amount of training data is about four times that of the other methods. The proposed method can maintain high extraction accuracy as a whole compared with the method excluding full sampling by reducing noise on the feature space, which has not been taken into account in the conventional methods.

6. CONCLUDING REMARKS

We proposed a research-activity-support system to facilitate the research activities of students in their theses at university laboratories. Since the tasks were basically confirmed and discovered at seminars when exchanging opinions, we focused on the mechanism of automatically extracting tasks from minutes and on assistance to smoothly execute the series of tasks.

We also proposed a machine-learning method of continuously updating the task-extraction model by active learning to maintain long-term system operation. The weighting algorithm was improved based on feature information in the application of active learning to minutes data because improvements to extraction accuracy with the conventional algorithms were not satisfactory.

We conducted a participant-based experiment on the proposed support system to evaluate the proposed method and a computer-simulation-based experiment on the proposed sampling method for active learning. Even though the number of subjects was small in the participant-based experiment, we confirmed that students in laboratories who actually participated in research at university and used the proposed system could more accurately comprehend the tasks to be achieved. The computer-simulation-based experiments confirmed that the proposed sampling of teacher-signal assignment for active learning could stably improve extraction accuracy compared to other sampling methods.

Future tasks include a larger scale participant-based experiment for a longer time period, quantitative evaluations on the achievement of tasks, and extension of the task-extraction model to handle linguistic-style changes in text minutes.

More detailed data will be obtained by conducting a large-scale subject experiment, which will lead to improvements in the system especially in terms of usability. The evaluation of the proposed system in this research for the second task mainly focused on the extent to which the user was able to be aware of the tasks to be achieved, but all such tasks were not necessarily attained. We should also quantitatively evaluate whether tasks have actually been achieved. However, since the amount of labor required for each task differed, it will be necessary to scrutinize a method of quantifying the achievement level. It goes without saying that the extraction of task statements for the third task increases as extraction accuracy increases, and the less the burden on the user decreases.

Active learning is one solution to improving accuracy, but there is a method of reviewing and improving the features of the extraction model. The features concerning the morphemes of the statement used in the current task-statement-extraction model were determined based on past preliminary survey data. Unlike other features, the features of morphemes are directly affected by temporal changes in features, so updating morphemes as features when the amount of data increases will be one method of model expansion. In addition, although only the surface features of language, such as morphemes and sentences, are used as linguistic features, improvements to extraction accuracy can be expected by using deep language features such as semantic information on sentences, such as word senses and rhetorical structures of statements.

ACKNOWLEDGMENTS

The authors would like to thank reviewers of this journal for their comments for improving this paper. The authors also thank members of Nagao Laboratory of Nagoya University for their efforts for managing systems and acquiring data.

REFERENCES

[1] R. S. J. D. Baker and G. Siemens, "Educational Data Mining and Learning Analytics", In K. Sawyer (Ed.), **The Cambridge Handbook of the Learning Sciences** (2nd Edition), Cambridge; New York: Cambridge University Press, 2014.

[2] K. Nagao, K. Kaji, D. Yamamoto, and H. Tomobe, "Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities", In **Proceedings of the Fifth PacificRim Conference on Multimedia**, pp. 522–531, 2004.

[3] B. K. Daniel (Ed.), **Big Data and Learning Analytics in Higher Education: Current Theory and Practice**, Springer International Publishing, 2017.

[4] B. Settles, "Active Learning Literature Survey", **Computer Sciences Technical Report** 1648, University of Wisconsin-Madison, 2010.

[5] K. Nagao, K. Inoue, N. Morita, and S. Matsubara, "Automatic Extraction of Task Statements from Structured Meeting Content", In **Proceedings of the 7th International Conference on Knowledge Discovery and Information Retrieval**, pp. 307–315, 2015.

[6] K. Georgala, A. Kosmopoulos, and G. Paliouras, "Spam Filtering: an Active Learning Approach using Incremental Clustering", In **Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics**, No. 23, 2014.

[7] H. Shimodaira, "Improving Predictive Inference under Covariate Shift by Weighting the Log-Likelihood Function", **Journal of Statistical Planning and Inference**, vol. 90, pp. 227–244, 2000.

[8] M. Sugiyama and M. Kawanabe, **Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation**, MIT Press, 2012.

[9] A. Liu, L. Reyzin, and B. D. Ziebart, "Shift-Pessimistic Active Learning using Robust Bias-Aware Prediction", In **Proceedings of the AAAI Conference on Artificial Intelligence**, 2015.

[10] T. Osone and K. Uota, "An Approach to Teaching Basic Information Education based on PDCA Cycle", **Business Review of Senshu University**, No. 100, pp. 1–14, 2015.

[11] F. Watanabe, Y. Mori, and C. Kogo, "Analyzing Learners' Subjective Evaluation of Peer Assessment in Japan Massive Open Online Courses", **Waseda Journal of Human Sciences**, Vol. 28, No. 2, pp. 237–245, 2015.

[12] B. Settles and M. Craven, "An Analysis of Active Learning Strategies for Sequence Labeling Tasks", In **Proceedings of the Conference on Empirical Methods in Natural Language Processing**, Association for Computational Linguistics, 2008.

[13] D. D. Lewis and W. A. Gale, "A Sequential Algorithm for Training Text Classifiers", In **Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval**, pp. 3-12, ACM/Springer, 1994.

[14] N. Roy and A. Mccallum, "Toward Optimal Active Learning through Monte Carlo Estimation of Error Reduction", In **Proceedings of the 18th International Conference on Machine Learning (ICML)**, pp. 441–448, 2001.