Topic Visualization for Understanding Research Paper in Collaborative Discussion

Masato Aoki, Yuki Hayashi, Tomoko Kojiri, and Toyohide Watanabe

Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan
{maoki,yhayashi,kojiri,watanabe}@watanabe.ss.is.nagoya-u.ac.jp

Abstract. Our objective is to construct a collaborative discussion support system for understanding the research paper in the research group by visualizing the contents of the discussion. When reading a research paper, it is essential not only to understand the contents of the paper but also to obtain the related knowledge. Our system evaluates topics in discussion from the viewpoints of similarity between a topic and the paper and that between topics. Then, the system visualizes topics as circles around chapter nodes whose distances from their target chapter nodes and other topic nodes represent the similarities. By observing the visualized topics, to derive topics from various perspectives is encouraged.

Keywords: understanding research paper collaborative discussion, discussion visualization.

1 Introduction

When we read the related research papers, it is essential not only to understand the contents of the paper but also to obtain the related knowledge. However, each of us has limited knowledge and may not be able to obtain enough knowledge by himself/herself. One solution for this problem is to discuss the contents of the paper in a group to obtain various knowledge. Through discussion with others, related knowledge that other participants have can be acquired. Moreover, knowledge of different perspectives may be derived through the discussion. There are many researches for supporting the discussion in a distributed environment. By these researches, participants are able to communicate with each other without limitations of time and location. Therefore, this research focuses on a collaborative discussion in a research group for obtaining related knowledge of the research paper in a distributed environment. In order to appropriately assess the paper, participants should discuss various research stages in the paper such as the background, objective and method. However, participants cannot always discuss the paper effectively because they sometimes focus on the topics from limited perspective. If they can notice the situation of discussion timely, they can grasp a diversity of discussed topics. That is, if they discuss the paper only from limited perspectives, they may be able to derive new topics from different

R. Setchi et al. (Eds.): KES 2010, Part IV, LNAI 6279, pp. 153–162, 2010.

perspectives. Our objective is to construct a collaborative discussion support system for promoting the effective discussion by visualizing the diversity of the discussed topics. In our research, topics are defined as groups of utterances whose target is a certain part of the paper, such as chapter, section and words.

For promoting the reflection of the finished discussion, there are many researches which visualize the structure of the discussion after the discussion has been finished. Conklin et al. constructed a discussion support system for detecting inconsistencies and less discussed topics by representing the relationship between utterances for problem-solving[1]. The system divides participants' utterances into four types(Issue, Position, Argument, Other) and represents them by nodes with different attributes. In addition, the relations between utterances are represented by labeled links, such as *generalizes*, *specializes* and *replaces*. In order to organize effective discussion, available relations are defined according to the type of the target node. When participants make utterances, they need to input the types of utterances and the relations to the target utterances. Therefore, participants cannot conduct free discussion since their utterances are controlled by the types of target utterances.

Kojiri et al. proposed a system which visualizes the structure of an ongoing discussion[2]. In order to extract and present important utterances in topics, the number of utterances of the topics and the uttered time are considered. This system can indicate the current important utterances, but cannot promote an effective discussion. Instead of extracting only specific utterances of the topics, it is necessary to evaluate the discussion by considering all the utterances.

Some researches try to activate the discussion by showing activeness of participants. Viegas et al. developed the system called Chat Circles which expresses the utterances of each participant in resizable circles[3]. Active participants are displayed as large circles since the sizes of the circles become smaller depending on the time to say nothing. In addition, the system has a function for browsing the chat log by arranged bars in chronological order whose lengths vary depending on the number of words of each utterance. Erickson et al. proposed the visualization method of representing activeness of discussion by positions of circles which correspond to participants[4]. Each circle is placed on a common circle which shows their workspaces and the center of the circle means high activity level. Xiong et al. constructed the system which represents the activeness of each participant by using the metaphor of a flower[6]. In order to distinguish individual participants and to comprehend the overall interaction, each participant is represented as a flower whose length of the stem expresses the length of his/her login time of each participant. Moreover, the petals of each participant represent proposed topics, and the responses from other participants are displayed as small circles at the distal end of the target topic. Tat et al. focused on the social interaction centered around one participant, and supported self exploration of one's own chat history[5]. They constructed the system to display each chat history of the participants as a line of circles(utterances) in the direction corresponding to the participants. For the purpose of understanding conversational tones, this system estimates the emotions of each participant by used emoticons and represents the

emotions by colors of translucent planes in the directions corresponding to the participants. In addition, this system changes the color strengths of the circles according to the number of characters of utterances. The time filtering function allows participants to filter conversations by selecting a certain time and participants can understand the activeness of the discussion at the time. However, in these researches, participants cannot always grasp the goodness of the discussion which indicates varieties of perspectives derived in their discussion. The quality of the discussion is not represented only by active participants and active time. In order to facilitate effective discussion, the goodness of the discussion should be visualized in the interface.

In this research, topics are evaluated and visualized from the perspective of obtaining related knowledge for the purpose of promoting discussion. To obtain related knowledge is to acquire opinions from various perspectives for each phase of the research. In order to obtain additional related knowledge, it is desirable to discuss various topics from diverse perspectives. Thus, every topic in the discussion is evaluated from the viewpoint of similarity between the topic and the paper and that between topics. By organizing topics around a chapter, participants can understand the situation of the discussion intuitively and are encouraged to discuss diverse topics voluntarily.

2 Approach

The target paper of discussion is research paper in engineering. We assume that the participants of discussion are researchers who are interested in the target research field of the paper.

A discussion is generally classified into *creative discussion* and *problem-solving discussion*. In *creative discussion*, participants do not always have a clear goal, but intend to find various perspectives regarding the discussion theme. In *problem-solving discussion*, participants aim to reach a specific goal. The discussion for obtaining the related knowledge of the paper is *creative discussion* since a clear goal does not exist. Moreover, *creative discussion* is classified into *focused discussion* and *global discussion* depending on the location of the paper which participants need to acquire knowledge from. The purpose of *focused discussion* is the deep understanding of certain parts of the paper such as a technology and the assumed environment in the paper. In *global discussion*, participants desire to acquire the opinions about any part of the paper from various perspectives. Since each chapter of the paper corresponds to each stage of the research, to obtain comprehensive knowledge of the paper leads participants to consider various aspects of the paper such as background, objective, solution and evaluation. Therefore, in this research, we aim to support *creative* and *global discussion*.

Since each chapter of the paper is individually made up of the meaningful research aspects, it is important to discuss all the chapters. In addition, if there are many topics from the same perspective, participants evaluate the paper based on only limited viewpoint. Therefore, various topics should be discussed. In *creative discussion*, developed topics which are associated with the paper are required.

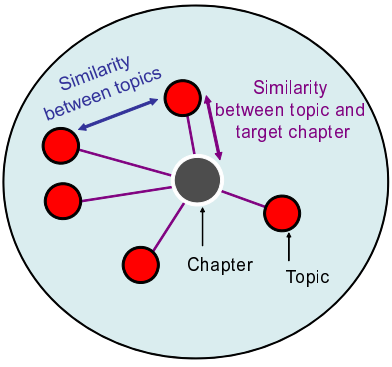

Fig. 1. Concept of visualization

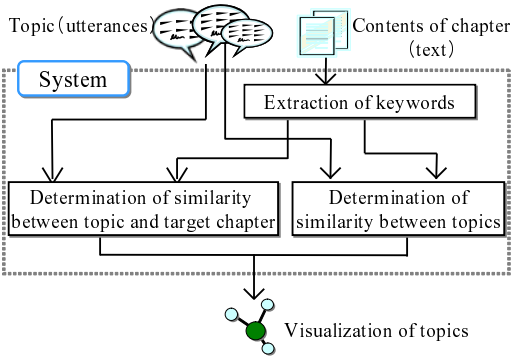

Fig. 2. Processing steps

Topics such as chitchats which do not directly relate to the contents of papers do not contain the information to evaluate the papers. Thus, in this research, every topic in the discussion is evaluated from viewpoint of the similarity between a topic and its target chapter and that between topics.

Figure 1 shows the concept of visualizing topics in the discussion. The visualization should make participants aware of the situation of the discussion. In our system, the topics which are collections of utterances are extracted and placed as "topic node" around "chapter node" which indicates the contents of the target chapter. The similarities between topics and chapters are calculated based on keywords of topics in the paper. The similarity between a topic and its target chapter is represented by the distance between a topic node and a chapter node. The similarity among topics is represented by the distance among topic nodes. Participants are able to be aware of the situation of their discussion by distribution of topic nodes around the chapter nodes. If many topic nodes exist near the chapter node or in the same direction, participants are urged to derive developed topics from other perspectives. In addition, the chapter which lacks the discussion is able to be grasped by the distribution of the topic nodes for each chapter.

Figure 2 shows the processing steps for visualizing the topics. Currently, we focus on the text-based discussion using the chat. The system gets the keywords of chapters in advance from the texts of each chapter. The similarity between a topic and its target chapter is calculated based on the keywords contained in the topic. The similarity between topics is regarded as the difference between the target locations in the chapter. Sentences represent the context of the paper. Similar or related sentences are written in near locations. Therefore, the similarity between topics is calculated by the distances between target locations. The system determines the locations of the topics and displays them based on the calculated degrees of the similarity between a topic and its target chapter and the target location of the topics.

3 Topic Visualization Method

3.1 Extraction of Keywords in Chapter

The chapter contents can be expressed as a set of keywords. In this research, keywords are defined as characteristic words(noun) in each chapter. Such words appear frequently in the chapter and do not appear in a whole paper. In order to acquire keywords for each chapter, important degrees of each word for the chapters are calculated by Equation 1. $value(c, a)$ represents the importance degree of word a in chapter c. $count(c, a)$ is the number of word a which appears in chapter c and $N(c)$ is the total number of words in chapter c. The important degree of the word becomes large if the word is used frequently in the chapter and becomes small if the word appears in the whole paper. Our system extracts words whose important degrees are larger than the threshold as keywords.

$$value(c, a) = \frac{count(c, a)}{N(c)} \times \log\left(\frac{\sum_i N(i)}{\sum_i count(i, a)}\right) \tag{1}$$

3.2 Expression of Similarity between Topic and Chapter

It is desirable that the topic which has strong connection to contents of the target chapter is placed close to the chapter node. The distance between a topic node and the chapter node is defined as Equation 2. $distance(c, t)$ represents the distance of a topic t from target chapter c. $\sum_i value(c, i)$ is the sum of important degrees of all keywords in chapter c. $relation(c, t)$ is the degree of the similarity between topic t and its target chapter c. According to the equation, the distance of the topic which has a large degree of the similarity becomes small as shown in Figure 3.

The similarity between a topic and its target chapter is expressed by the ratio of keywords included in a topic. Thus, $relation(c, t)$ is defined as Equation 3. $W(t)$ is the total number of words in topic t. $chaptIn(c, t)$ is the number of keywords of target chapter c contained in topic t. α is a constant number that coordinates the effect of $\sum_{i \in t \cap c} value(c, i)$ and the value takes from 0 to $\sum_i value(c, i)$. As α gets larger, the effect of $chaptIn(c, t)$ becomes greater. Based on the equation, the degree of the topic that includes large number of important keywords of the chapter becomes large.

$$distance(c, t) = \sum_i value(c, i) - relation(c, t) \tag{2}$$

$$relation(c, t) = \frac{chaptIn(c, t)}{W(t)} \times (\alpha + \sum_{i \in t \cap c} value(c, i)) \tag{3}$$

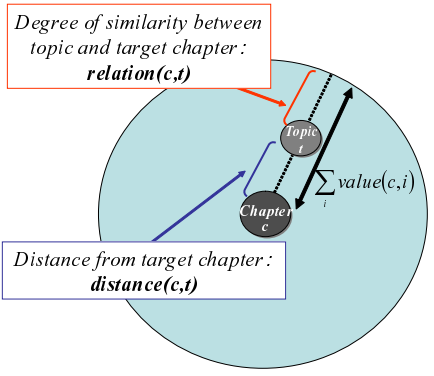

Fig. 3. Expression of similarity between chapter and topic

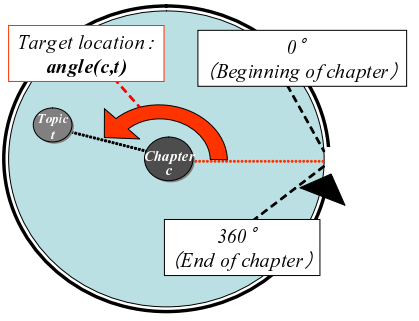

Fig. 4. Expression of target location

3.3 Expression of Similarity between Topics

The similarity between topics can be determined by the target locations in the chapter. We define the degree of the similarity between topics as the distance between the target locations of topics. The target location of a topic can be grasped by keywords of the chapter that it contains and is calculated by Equation 4. $position(c,t)$ represents the target location of topic t in target chapter c and its value takes from 0(beginning of the chapter) to 1(end of the chapter). $loc(c,i)$ indicates the appearance position of keyword i in chapter c. If keyword i appears in multiple locations, $loc(c,i)$ is set as the middle point of the appearance position of the keyword i. The target location of the topic is represented as the average of the appearance positions of all emerging keywords.

The angle of the topic node is determined according to the $position(c,t)$. The beginning of each chapter corresponds to 0° and the end of the chapter is 360° around the chapter node. The angle of the topic node is calculated by Equation 5. $angle(c,t)$ is an angle of topic t around chapter c. An angle is determined by multiplying 360° by a target location $position(c,t)$ as shown in Figure 4. Based on this expression, the beginning and the end of a chapter are placed near. It is often observed that main theme of the chapter is insisted in the first sentence and is summarized in the end. Thus, this expression is valid to some extent.

$$position(c,t) = \frac{\frac{1}{chaprIn(c,t)} \times \sum_{i \in t \cap c} loc(c,i)}{N(c)} \tag{4}$$

$$angle(c,t) = 360° \times position(c,t) \tag{5}$$

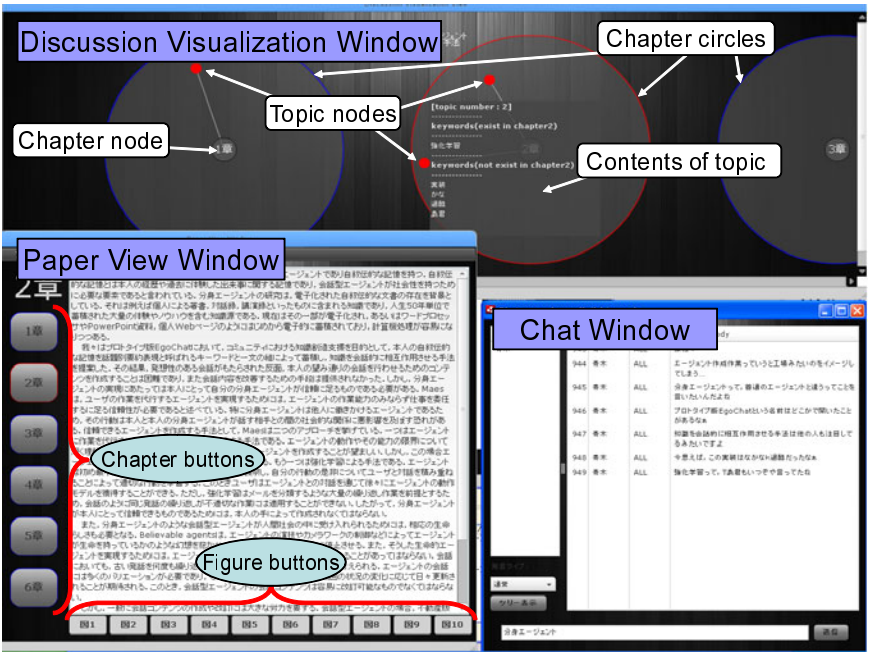

Fig. 5. Interface of prototype system

4 Prototype System

We have constructed a prototype system by embedding paper view mechanism and discussion visualization mechanism in the collaborative learning support system in our laboratory[7]. The interface of this system consists of three windows as shown in Figure 5. Participants make utterances using the chat window. Paper view window displays the contents of the paper to all participants. This window provides the same contents to all participants. By selecting the chapter button, the contents of the selected chapter appear in the windows. By selecting the figure button, the figure in the chapter is displayed in a separate window. When a chapter button is pushed, our system regards that the topic is changed and gets the utterances that compose the topic. Moreover, these utterances are analyzed by our visualization method and the result is sent to discussion visualization window.

In discussion visualization window, each chapter is viewed as a circle with its chapter node exists in the center of the circle. Topic nodes are represented by red circles around the chapter node. Information of words within the topic are shown by moving the mouse cursor over a topic node(Figure 6). In addition, the utterances of the topic are displayed by clicking a topic node(Figure 7). When a participant clicks the circumference of the chapter circles, the diverse keywords of the chapter are displayed whose angles correspond to the emerging location

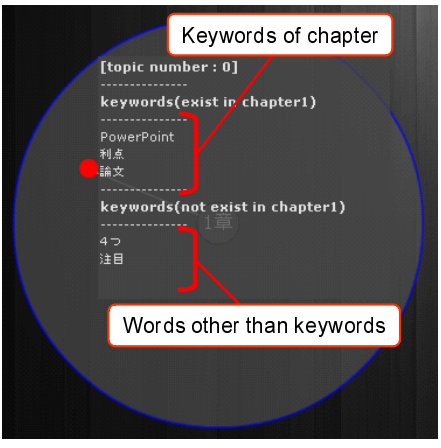

Fig. 6. Information of words in topic

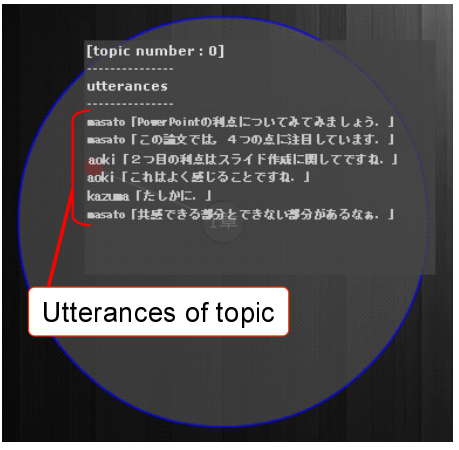

Fig. 7. Utterances of topic

in the chapter. Discussion of the specific keyword may be enforced by observing such keywords. By clicking the displayed keyword, the keyword is posted in the input area of the chat window.

5 Experiment

5.1 Experimental Setting

We evaluated the adequacy and effectiveness of the visualization method using the prototype system. In this experiment, maximum length of the distance between a chapter node and a topic node was normalized to 100, and α was set to 50.

Two groups(A and B) of four students in our laboratory were asked to discuss the research paper which consists of six chapters. The purpose of the discussion was to get the related knowledge of the paper from various perspectives. Each examinee was asked to read the paper and understand the contents in advance. If they wanted to check the contents during the discussion, they were asked to view their own papers and not to use the paper view window. The paper view window was used only for changing the target of the discussion topics. One examinee of each group was asked to determine the end of the discussion. In both groups, the discussion continued for more than one hour.

After the discussion, the examinees answered a questionnaire about the similarity between a topic and the chapter and that between topics. In addition, the examinees were asked to observe the discussion record of the other group that were organized for each topics. In order to evaluate the validity of calculated degrees of the similarity between a topic and its target chapter, the examinees were asked to divide the topics of the other group according to the relation to the chapter, slightly related ones and greatly related ones. For evaluating validity of calculated degrees of the similarity between topics, they were also asked to select the combination of similar topics of the other group with reasons.

5.2 Experimental Result

Table 1 shows the results of the average distances of greatly related or slightly related topics to the chapter calculated by the system. The average distance of all topics in group A is 73.14, and that in group B is 71.64. For both groups, the average distances of slightly related topics are greater than those of greatly related ones. Therefore, the system adequately expresses the similarity between a topic and the target chapter.

The average angles between similar topics are shown in Table 2. The average angle between all topics in group A is 90.62°, and that in group B is 70.59°. For both groups, the average angles between similar topics are smaller than that of all topics. Therefore, the system is able to place similar topics in the near location.

The result of the questionnaire about topic visualization method is shown in Table 3. For each question, 1 is the worst and 5 represents the best. In questions about the adequacy and effectiveness of visualizing topics*(a,b,c)*, answers were good. Therefore, the topic visualization method is appropriate for understanding the situation of the discussion. However, the results of questions about the effectiveness for triggering a new topic*(e,d)* indicate that the visualization did not lead to specific topics. Examinees commented that discussion topic changes according to the context, so it is difficult to change topics according to the keywords in the circumference of the circle. Therefore, the method for guiding a discussion topic needs to be considered so as to reflect the context of the discussion.

Table 1. Average distances between topic and target chapter

	Group A	Group B
Topics slightly related to chapter	83.40	77.05
Topics greatly related to chapter	66.32	69.54

Table 2. Average angles between topics

	Group A	Group B
Similar topics	56.97°	38.02°

Table 3. Average scores of each question

	Group A	Group B
a. Adequacy of distance between topic and chapter	3	3.5
b. Adequacy of distance between topics	3.75	4
c. Effectiveness for grasping situation of discussion	3.5	4
d. Effectiveness for selecting target chapter	1.75	2.5
e. Effectiveness for selecting target location in chapter	2.25	1.75

6 Conclusion

In this research, we proposed a system for supporting the collaborative discussion for obtaining the contents related to the research paper. The topics of the discussion are displayed for each chapter according to the similarity between a topic and the chapter and that between topics. The experimental results showed that the visualization of topics is appropriate for grasping the situation of the discussion, but does not contribute for leading a discussion for specific topics.

For our future work, we have to devise a method for leading effective discussion by showing the keywords according to the progress of discussion. Appropriate keywords for the next topic may relate to the previous topics. In order to select such keywords, the detection for keywords which are not discussed effectively in the previous topics and relate to the current topic needs to be developed.

Our collaborative discussion system focuses on a research activity of reading a research paper. The objective of reading the research paper is to make the position of own research clear by evaluating the capability of other researches. In order to help participants assess the paper after the discussion, it is desirable that the discussed topics are arranged from each participant's viewpoint. In our future research, we will help participants evaluate the paper using the results of the discussions.

References

1. Conklin, J., Begerman, M.: glBIS: A Hypertext Tool for Team Design Deliberation. In: Proc. of Hypertext 1987, pp. 247–251 (1987)
2. Kojiri, T., Yamaguchi, K., Watanabe, T.: Topic-tree Representation of Discussion Records in a Collaborative Learning Process. The Journal of Information and Systems in Education 5(1), 29–37 (2006)
3. Viegas, F.B., Donath, J.S.: Chat Circles. In: Proc. of SIGCHI 1999, pp. 9–16 (1999)
4. Erickson, T., Kellogg, W.A., Laff, M., Sussman, J., Wolf, T.V., Halverson, C.A., Edwards, D.: A Persistent Chat Space for Work Groups: The Design, Evaluation and Deployment of Loops. In: Proc. of the 6th ACM Conference on Designing Interactive Systems, pp. 331–340 (2006)
5. Tat, A., Carpendale, S.: CrystalChat: Visualizing Personal Chat History. In: Proc. of the 39th Annual Hawaii International Conference on System Sciences, vol. 3, pp. 58–68 (2006)
6. Xiong, R., Donath, J.: PeopleGarden: Creating Data Portraits for Users. In: Proc. of the 12th Annual ACM Symposium on User Interface Software and Technology, pp. 37–44 (1999)
7. Hayashi, Y., Kojiri, T., Watanabe, T.: Focus Support Interface Based on Actions for Collaborative Learning. International Journal of Neurocomputing 73, 669–675 (2010)