i-Bee: the CSCL Assessment Tool for Making the Status of Discussion Visible

Toshio MOCHIZUKI*1,*2 Satoru FUJITANI*3 Shin-ichi HISAMATSU*4
Kazaru YAEGASHI*5 Tomoko NAGATA*6 Jun NAKAHARA*7
Toshihisa NISHIMORI*7 Mariko SUZUKI*8 Hiroshi KATO*1,*7

*1 Information Science and Technology Center, Kobe University, Japan
*2 Department of Cyber Society and Culture, The Graduate University for Advanced Studies, Japan
*3 Faculty of Business Administration, Mejiro University, Japan
*4 Graduate School of Media and Governance, Keio University
*5 Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Japan
*6 Hyogo University of Teacher Education, Japan
*7 National Institute of Multimedia Education (NIME), Japan
*8 Faculty of Education, Shiga University, Japan
E-mail: *1,*2tmochi@kobe-u.ac.jp, *3fujitani@mejiro.ac.jp, *4bm@sfc.wide.ad.jp,
*5kazaru@interlink.or.jp, *6tnagata@life.hyogo-u.ac.jp, *7jun@nakahara-lab.net,
*7nisimori@nime.ac.jp, *8mariko@sue.shiga-u.ac.jp, *1,*7hkato@nime.ac.jp

Abstract: This paper describes a CSCL (Computer Supported Collaborative Learning) assessment tool that aims at visualizing discussions among learners participating on electronic forums using a text-mining technique. The software, named "i-Bee" (Bulletin board EnvisionEr), is able to visually show the co-occurrence relations of topic keywords of each learner using Correspondence Analysis. Furthermore, i-Bee can show how actively each learner discusses; how each learner contributes to online discussions; and the frequency the learners use each keyword. In this paper, the authors report the rationale, the basic algorithm, and the implementation of i-Bee.

1. Introduction

Assessment and evaluation of collaborative learning has become an important research topic in recent years - not only for teachers who must to grade the learners, but also for the learners themselves. Unlike traditional concepts of educational evaluation, the learner's ability in collaborative learning emerges socially and therefore should be evaluated from a social constructivism way - the ability becomes visible through interaction between individual learners and the circumstances including artifacts and social factors (Kato, 2001). This viewpoint is different from the traditional “knowledge transfer” perspective. From this point of view, learners have to approach an authentic task (e.g. authentic problem-solving) in their learning, and the assessment or evaluation of their learning should be conducted with their all visualized abilities in the context of the learning (in situ).

In addition, qualitative assessment has increased its importance as a factor in educational evaluation in these days. Qualitative assessment puts higher value on evaluating a learning environment rather than a “learner’s skill”. Learners evaluate their own learning activities with their teachers using qualitative materials produced through their learning process, and reflect upon and re-design their activities and learning environment with their teachers. From a social constructivism perspective, it is very important for students to assess their learning when planning their upcoming study (Shaklee, et al., 1997). In collaborative learning, in order for the learners to make their learning substantial, they must monitor the conditions of their discussions, learning processes, and human relations to improve their learning community and to plan their upcoming study.

Messages in electronic forums are useful to assess collaborative learning based on social constructivism, because the messages are qualitatively visualized resources of interaction among learners in collaborative learning; messages in learners' discussion reflect their learning in the context of the learning (in situ) (Pea, 1993; Palincsar, 1998). Qualitative assessment of communication between and among learners in the records of CSCL has always consisted of a content analysis of all messages to detect any substantial change in them (Chi, et al., 1994; Oshima,
1997; Hmelo-Silver, 2003). However, assessing these records for each individual learner has created a tremendous burden on the investigators and is not practical.

In light of this, some researchers have tried to extract keywords (Simoff, 1999) and abstract of messages (Fujitani and Akahori, 2000) from the discussions using the quantitative method. However, certain problems still remain in these studies: (1) Due to employing the probabilistic method to show the co-occurrence relations, the sentences were too short in general to contain enough information for use in a collaborative learning context, raising the question of reliability; (2) as a result of presenting only the summaries, they do not go as far as to indicate the contribution of individual learners to the discussion, so that it is of little help in assessing individual learners, although the overall message can be understood; and (3) this method might be of some use in helping learners who did not participate in the discussion from the start to grasp the situation, but it is not clear what merit it has for those who actively participated.

From the above reasons, we have developed a CSCL assessment tool for qualitatively assessment, which maps the co-occurrence relations between the learners and the keywords; in other words, how many times each learner uses each keyword in his/her message. The map plots the overall status of discussion in the learning community, which shows the learners’ commitments in the group discussion. This paper introduces the methods to analyze the communication relationships; to implement the system.

2. Visualization of the discussion between individuals and overall the community

Many recent studies in CSCL have focused on visualizing learner activities in CSCL. Nakahara has indicated the difficulty for the learners to understand status of discussion in CSCL, and he has developed visualization tool that visualizes the status of the interaction on the top screen of mobile phone (Nakahara, et al., 2003). Other researchers have tried to visualize the activeness of electronic forums (e.g. Yamauchi, et al., 2002) and social networks in the community (e.g. Martinez, et al., 2003) by way of confirming the status of community on CSCL. However, up until now very few prior researches focused on visualizing contents of discussion among the learners. Puntambekar and Luckin (2003) have indicated that it must be proved worthwhile to make discussion content visible to the participants and have them learn by reflecting on the process. In this paper, we propose a visualization tool using the method of text-mining to assess discourse between the learners on electronic discussion forums.

Application of Text-mining Technique

In recent years, many progresses have been made in the field of text-mining. A number of methods have emerged for extracting applicable keywords from the text data. And multivariate analyses such as multivariable dimension scale (MDS) and Correspondence Analysis (CA) are generally used for visualizing the relationship of individual keywords to the whole (Blasius and Greenacre, 1998; Watanabe, 2001).

CA is a graphically descriptive method that facilitates understanding the relationship intuitively by presenting two or more discrete variables in a complex data matrix. When the matrix is based on the frequency of each keyword to each entity, frequently co-occurred variables are distributed in proximity to each other so that clusters (of keywords) can be grasped in the mapping thus generated.

Since this method does not depend on any statistical assumptions, it is possible to apply it even to a small size of statistical data.

Visualizing Discussion Using CA

In the discourse analysis proposed in this paper, \( n \) learners, who discuss concerning a relevant number of \( m \) keywords totaling \( n \times m \) for a cross-tab of \( N \); CA yields a mapping of a row vector of \( F \) and a column vector of \( G \). In other words, the generalized singular value decomposition (SVD) of the matrix \( P \) which is the relative frequency matrix of \( N \).

\[
P = A \ D_\mu \ B^T
\]

yields a left generalized singular vector \( A \) and a right generalized singular vector \( B \). And by using them

\[
F = D_r^{-1} \ A \ D_\mu
\]
\[
G = D_c^{-1} \ B \ D_\mu
\]

result the standardized principal coordinates \( F, G \), which construct a mapping (Greenacre, 1984).
Here $D_\mu$ is the diagonal matrix which makes the generalized singular value diagonal vector, $D_R$ the diagonal matrix which makes matrix $P$ the diagonal vector, and $D_C$ the diagonal matrix of the sum of the columns of matrix $P$. And $F$ corresponds to coordinates of learners and $G$ corresponds to those of keywords.

**The Meaning of the Mapping Generated by Analysis**

Generally, when CA is conducted using the relative frequency matrix $P$, $F$ and $G$ are distributed in proximity to each other if a coordinate of $F$ and that of $G$ have strong co-occurrence relation. In contrast, if a coordinate of $F$ and that of $G$ do not have co-occurrence relation, they are distributed far from each other. In addition, relatively high value in the matrix $N$ represents a coordinate located near by the original point, but relatively low value represents a coordinate located far from the original point.

CA actually conducts multivariate analysis so that the principal location explained above effects mutually and the coordinates are totally organized in a balanced manner. So it is believed that (1) distribution of coordinates indicates the co-occurrence relation between each learner and each keyword in his/her messages; (2) the total data of (1) represents of the topics in the discussions. Hence CA can show the status of overall discussion in the electronic forums and each learner’s involvement in the content of discussion.

We already conducted a pilot study to examine appropriateness of CA to visualize the discussion and to examine the effectiveness of the mapping for the learner’s self-assessment. The result showed that it could be appropriate and be effective for the learners to aware and reflect of the discussion (Mochizuki et al., 2003).

### 3. Development of i-Bee

We develop a CSCL assessment tool named i-Bee (Bulletin board EnvisionEr), to visualize a small group discussion on electronic forums using CA in real time. i-Bee can show activeness of the learners and the keywords, and the co-occurrence relation between them at a certain period. i-Bee can also visualize time-series change of these status as an animation.

**Outline and Design Concept**

We develop i-Bee as a plug-in tool that works with discussion forums of exCampus and its databases. exCampus is an e-Learning site module developed and distributed free by National Institute of Multimedia Education in Japan (Nakahara and Nishimori, 2003). exCampus covers many functions necessary to build an e-Learning site in a university- course management, learning management, interface for video streaming, discussion forums, etc.

When a learner logs in an electronic forum on exCampus, i-Bee pops up as an extra window (Fig.1). i-Bee displays participating learners (bees) and keywords (flowers) selected by teachers. Distribution of the bees and the flowers is based on the result of CA. Each bee and each flower are drawn with its name which represents what is described.

i-Bee is developed to understand the status on the forums, and to have learners reflect their attitude in the discussion based on the topics and opinions discussed. In order for learners to assess their discussion appropriately, it is necessary to design visualized image as learners can easily understand the overall image of the discussion and their involvement in the discussion.

In such a light, we adopt a metaphor of “bee and flower” to visualize the status of learners’ discussion. Based on the algorism of CA, strongly related elements are distributed as coordinates nearby each other. On the other hand, bee flies to flowers which she likes and wants to suck their nectar in the nature. Flowers need bees to distribute their pollen. Therefore we adopt the metaphor to explain the co-occurrence relation between learners and keywords in the discussion.

Thus the learners can take a view on content and status of the discussion in the forum, so it is expected that the learners can focus on some topics where they participate more active, plan their approach to other topics they have not interested in, or follow up members who cannot participate in the discussion well (Mochizuki, et al., 2003).
Visualization of Discussion Process

In order to promote the learners’ reflection on their discussion much more, we develop i-Bee to visualize previous status and process of change of the discussion. i-Bee visualizes trajectory of the learners’ coordinates from the unit time $t-1$ to $t$ as the opening animation of Flash movie (here $t$ is number of unit time which is counted from the beginning every a certain time length). Using the management tool, moderators, e.g. teacher and TAs, need to configure the unit of time appropriately in accord with learning activity; e.g. when the course is held once a week, the teacher may set the unit time to analyze one week.

The learners can also view further previous status at every unit of time from the beginning. The learners can see status of discussion as a snapshot at one week before, one unit of time before, one unit of time after, or one week after, clicking buttons within the window of i-Bee.

Visualization of Activeness

It is difficult for learners and the moderators to understand the status of discussion with only the simple coordinates of bees and flowers produced by CA, because it does not express recent activeness of the learners and the keywords (topics) in the discussion.

In order to visualize their activeness at a certain period, i-Bee draws bees and flowers at three levels (see Table 1): “Sleeping bee”, “normal flying bee”, and “active flying bee” are possible facial expressions of the learner’s recent participation level. “Bud of flower”, “flowering period”, and “full bloom” are those of recent appearance frequency of the keyword. i-Bee calculates each learner’s activeness as proportion of his/her messages within the recent unit time to its average per unit of time. In case of a certain keyword, i-Bee calculates the activeness as proportion of frequency of the keywords used by all learners within the recent unit time to its average per unit of time.

<table>
<thead>
<tr>
<th>Information Index</th>
<th>Target</th>
<th>Ways of Facial Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>What each learner talks</td>
<td>Coordinates calculated by CA</td>
<td>Distance between bees and flowers</td>
</tr>
<tr>
<td>Active of each learner</td>
<td>$i = \frac{\text{number of the learner's articles at a certain period}}{\text{average number of the learner's articles per a period}}$</td>
<td>“bee”</td>
</tr>
<tr>
<td>Active of each topic (keyword)</td>
<td>$i = \frac{\text{frequency of the keyword used by all learners at a certain period}}{\text{average frequency of the keywords used by all learners per a period}}$</td>
<td>“flower”</td>
</tr>
</tbody>
</table>
Cooperation with exCampus Discussion Forums

We develop i-Bee to cooperate with discussion forums of exCampus. The learners can search messages which contain a certain keyword drawn as a flower on i-Bee. Hence the learners can find some messages containing the keyword they are interested in, clicking on the flower. Therefore, i-Bee supports the learners to find interested articles from a lot of messages or surprising articles from active topics on the forums.

Implementation

Fig.2 shows the workflow of i-Bee. i-Bee needs a morpheme analysis system, e.g. “ChaSen” for Japanese text (Matsumoto et al., 2000), to count frequency of appeared words each from the text of the discussion under electronic forums.

For using i-Bee in a course, the moderators need to set keywords using the management tool, because the automatic keyword selection based on statistical analysis cannot choose appropriate words which represent the learners’ discussion (The management tool is only for the moderators. The moderators can modify settings of i-Bee- unit of time to organize frequency matrix, users whose articles are analyzed, users who use i-Bee, keywords selection, etc.). These keywords are stored in the condition database.

In the keyword database, frequency of keywords will be stored with the indexical information on the discussion, reflecting on the condition database. For appearance of each keyword, the database records:

- Speaker/Author of the message in a certain period
- Total frequency of each keyword of the message until a certain period

CA uses these data to draw graphical display of the discussion profiles using Ox. Ox is a formula processing environment which is an object-oriented matrix programming language with a comprehensive mathematical and statistical function library (Doomik, 2001).

Procedure of i-Bee is as follows: Firstly, the learners or the moderators open the visualizer (Fig.1) which is developed with Macromedia Flash MX, and the calculator orders the morpheme analysis system to count the appearance frequency of each word used by each learner till the moment, and store the frequency matrix in the database. Then, in order to show the status at a certain period and the previous status, CA calculates conjugated matrix of one at the time $t$ and the other at the previous period $t-1$, as mentioned before. That is, when $n(l, t, w)$ is accumulated frequency that the learner $l$ uses the keyword $w$ until the unit time $t$, $\mathbf{N}$ is organized as below.

$$
N_{\text{ill.w}} = \begin{bmatrix}
\[n(l, t-1, l) & \cdots & n(l, t-1, k) & \cdots & n(l, t-1, W) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
n(l, t-1, l) & \cdots & n(l, t-1, w) & \cdots & n(l, t-1, W) \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
n(L, t-1, l) & \cdots & n(L, t-1, w) & \cdots & n(L, t-1, W) \\
n(1, t, l) & \cdots & n(1, t, w) & \cdots & n(1, t, W) \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
n(l, t, l) & \cdots & n(l, t, w) & \cdots & n(l, t, W) \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
n(L, t, l) & \cdots & n(L, t, w) & \cdots & n(L, t, W)
\end{bmatrix}
$$

where $l = 1, \ldots, L, t = 1, \ldots, T, w = 1, \ldots, W$

Then the calculator orders Ox to analyze by CA. However, if a learner does not use any keywords or if a keyword is not used at all, the operation is conducted with a matrix which is omitted the corresponding row or line from $\mathbf{N}$, because the operation cannot be completed due to the zero-line or the zero-row. The analysis results some value of axis, and coordinates $\mathbf{F}$ and $\mathbf{G}$ are elected as the first and second axis of the result. The calculator transforms value of the coordinates to XML format, and the visualizer receives the data from the calculator.

The graphical display produced by CA shows the co-occurrence relation between participants and keywords. Participants, as learners, can think about not only his/her situation in the group, but also the flow of the discussion.
5. Summary

This paper describes the rationale and the implementation of a CSCL assessment tool which learners can assess their collaborative discussion. The tool named i-Bee analyzes messages in electronic forums on an e-Learning system and represents kinds of messages and learners’ contributions. We are now studying the effectiveness of i-Bee for the learner's self-assessment through an analysis of an actual discussion in a course.

Acknowledgement

A part of this research has received assistance of Grant-in-Aid for JSPS Fellows for Toshio Mochizuki, Grant-in-Aid for Exploratory Research (Subject No. 15650171 for Hiroshi Kato), and Grant-in-Aids for Young Scientists (B) (Subject No. 14780120 for Satoru Fujitani and Subject No. 14780126 for Jun Nakahara), from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References