A Proposal of Activation Mechanism for User Communication based on User Behavior Analysis on Wedding Community Sites

Toshinori Hayashi*, Yuanyuan Wang†, Yukiko Kawai‡, and Kazutoshi Sumiya*
*Kwansei Gakuin University,
2–1 Gakuen, Sanda, Hyogo 669–1337, Japan
Email: den82687@kwansei.ac.jp, sumiya@kwansei.ac.jp
†Yamaguchi University,
2–16–1 Tokiwadai, Ube, Yamaguchi 755–8611, Japan
Email: y.wang@yamaguchi-u.ac.jp
‡Kyoto Sangyo University,
Kamigamo Motoyama, Kita-ku, Kyoto 603–8555, Japan
Email: kawai@cc.kyoto-su.ac.jp

***Abstract*—In this paper, we present an active communication mechanism based on a user behavior analysis on wedding community sites. To this end, we propose a novel mechanism for activation of user communication that provides related comments and users by detecting knowledge and interests from archived comments; this information from a wedding community website evokes conversations among users. The proposed mechanism has three components: 1) profiling static user information such as users' age and location and active user information like her dynamic interest and intention to communicate, 2) detecting and recommending users who are likely to communicate with each other, and 3) recommending comments that may be of interest to a user. Through the proposed activation mechanism, users on a wedding community site can communicate with each other easily and efficiently. We discuss our proposed user characteristic extraction and user recommendation methods using actual user posts from a wedding community website.**

Keywords–user behavior analysis; wedding community site; communication.

I. Introduction

In recent years, research has been conducted using data from Social Networking Services (SNSs) [1][2]. It is important to collect as much data as possible from SNS community sites, such as Facebook, LINE, and other Q&A sites. However, such services that focus on data collection cannot promote user communication on community websites because of differences in values. In this paper, we focus on a wedding community site, and we aim to promote user communication by recommending appropriate users and comments.

Specifically, we propose a novel active communication mechanism that shares comments of users by considering their knowledge and interests by analyzing their behavior on community websites. To this end, we first extract all posts of each user and extract their feature words using the term frequency–inverse document frequency (*tf-idf*) method. Next, we calculate the similarities among users to detect appropriate users. Finally, we recommend their comments by generating links to them in posts (Fig. 1). To use this mechanism, users can communicate with other users that are recommended to them about wedding planning; furthermore, it promotes communication among users on a wedding community site.

The remainder of the paper is organized as follows. Section II provides an overview of our system and reviews related work. Section III explains how to recommend users and their comments on a wedding community site. Section IV illustrates the experimental results obtained using a real dataset from a wedding community site. Finally, Section V concludes the paper and outlines our future work.

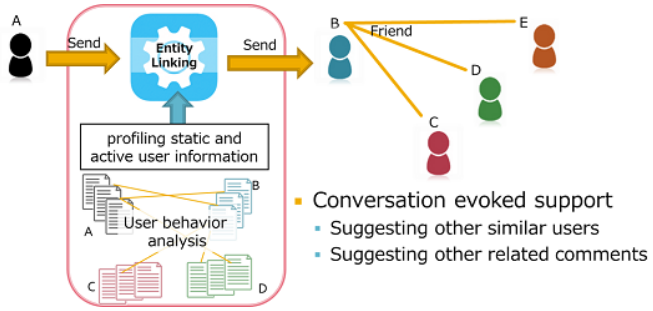

Figure 1. User and comment recommendations for activation of user communication based on a user behavior analysis.

II. System Overview and Related Work

A. Active Communication Mechanism

We present an active communication mechanism based on a user behavior analysis on wedding community sites. This mechanism has three steps: 1) user login information and user characteristic extraction, 2) user detection and recommendations, and 3) comment recommendations (Fig. 1).

To use this mechanism, users are required to install a toolbar (a browser plug-in) on an existing wedding community site in Japan. Wedding community sites are generally utilized by couples that plan to hold a wedding and are intended to assess a couples' needs regarding marriage. On this website, there are threads for wedding planning in different marriage statuses, and users can freely post their comments to each thread. The only way to communicate with other users is by replying to other users' comments on a thread. To improve replies, we propose a method that recommends both users and their comments by analyzing user behavior and their profile information on a wedding community site. Our goal is for our active communication mechanism to determine which users may want to communicate with other users.

A wedding community site is not a "Question & Answer site"; rather, it is a website where users can share their positive opinions and experiences about weddings. The proposed

 ISBN: 978-1-61208-548-7

system will recommend other users who have had similar situations or values of marriage to evoke communication between users. This system can also be used on other community websites; however, since the proposed system is considered on a wedding community site, it uses static information entered by a user during their initial user registration regarding their ideal wedding ceremony.

Fig. 1 shows the overview of our proposed mechanism. After a user posts, the mechanism analyzes the user behavior and recommends other users by calculating the similarities between them.

B. Related Work

Issac et al. [3] noted that communication is important to discuss different topics and work with others as a group. They mentioned that communication makes people more willing to contribute to society. Moreover, it is also effective for communication on websites, not only face-to-face communication. Ellison et al. [4] focused on SNS communities. According to these studies, communicating with others on SNSs makes more people feel happy.

In our previous work [5], users communicated with each other when they searched for web pages. In this work, we extend our previous work to recommend users and comments based on the link generation for a wedding community site. Although several automatic link generation methods for websites have been studied [6][7], they have primarily focused on web pages for knowledge support only; they did not consider communication among users. To address this deficiency, our proposed method recommends users to evoke communication. Other studies that have recommended analyzing user behavior on news sites [8] did not consider the relationships between users. In this paper, we first extract user posts to analyze user behavior and detect users to recommend by extracting the relationships between users.

Akihiro et al. [9] conducted an experiment for active communication in e-lectures through a chat system. However, it did not work very well because it was a burden for students to chat with others during the lectures. In this paper, we propose a new active communication mechanism by recommending appropriate users for different marriage statuses of users.

III. Active Communication Mechanism for Wedding Sites

A. User Behavior Analysis on a Wedding Community Site

To evoke communication among users, our active communication mechanism recommends users and their comments by analyzing user behavior on a wedding community site. According to our previous work [5], users can help other users when they search for the same web pages. Furthermore, in general, users communicate with each other easily when they are in similar statuses or situations. Therefore, in order to recommend users, we analyze to make 3 profiles based on aspects of wedding community site (see Fig. 2); in particular, we consider the axes of "Static Profile Information", "Marriage Status", and "Active Profile Information".

1) User Login Information Extraction: We extract user login information by acquiring user registration information on a wedding community site that users input upon site registration. Users input information such as their ages, places where they live, and marriage status. We divide the user login information as user static profile information and marriage status.

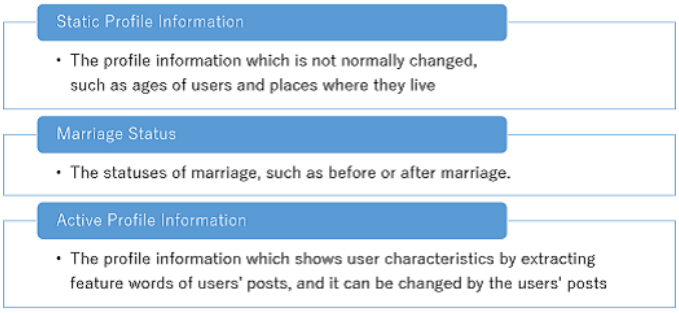

Figure 2. Profiling based on user's aspects.

2) User Characteristic Extraction: We extract user characteristics by extracting all posts of each user. Next, we calculate the term frequency and document frequency based on the *tf-idf* method; specifically, we use the following formulas:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}, \quad (1)$$

$$idf_i = \log \frac{|D|}{df_i}, \quad (2)$$

where $n_{i,j}$ denotes the term frequency of the word t_i in document d_j. In this work, d_j denotes the document that is integrated by all posts of one user. Therefore, the number of documents is equal to the number of users on the wedding community site. Furthermore, $\sum_k n_{k,j}$ denotes the sum of the term frequencies of all words in document d_j, and $|D|$ denotes the total number of documents, which is also equal to the number of users. Finally, df_i denotes the number of the documents that include the word t_i.

Based on the above, we use the obtained *tf-idf* values and feature words of each user to determine a users' active profile information.

B. User Detection and Recommendation

1) User Detection: We detect users based on the similarities of "Active Profile Information" between users by using the cosine similarity as follows:

$$Sim(\overrightarrow{x}, \overrightarrow{y}) = \frac{\sum_{i=1}^{|V|} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{|V|}(x_i)^2} \cdot \sqrt{\sum_{i=1}^{|V|}(y_i)^2}}, \quad (3)$$

where $\overrightarrow{x}$ denotes the feature vector of user x, and $\overrightarrow{y}$ denotes the feature vector of user y; $|V|$ is the number of dimensions of the feature vector.

"Marriage Status" is an absolute value, such as "before marriage" or "after marriage"; therefore, it will not change based on other users. However, "Static Profile Information" and "Active Profile Information" are relative values; they will change depending on each user.

 ISBN: 978-1-61208-548-7

TABLE I. FIVE USER PATTERNS FOR RECOMMENDATION.

Pattern	User (Who)	Marriage Status (to who)	Static Profile Information	Active Profile Information	Purpose
1	After marriage	Before	Neutral	Similar	Give advice
2	After marriage	After	Neutral	Similar	Share
3	Before marriage	Before	Similar	Different	Reference
4	Before marriage	Before	Neutral	Similar	Share
5	Before marriage	After	Neutral	Similar	Get advice

TABLE II. RECOMMENDATION SITUATION FOR EACH USER PATTERN.

Pattern	Purpose	When	How
1	Give advice	Links are generated in the comments	○○ needs some advice from you
2	Share	After Login	○○ is on the same status as you
3	Reference	Links are generated in the comments	You can refer to ○○
4	Share	After Login	○○ is on the same status as you
5	Get advice	Links are generated in the comments	○○ can be a good adviser for you

2) User Recommendation: We recommend users to communicate with others by considering users who have similar situations; such users may easily relate and share their experiences or advice. Based on the three axes described in the previous subsection, we classify five useful patterns of users on a wedding community site (see Table I).

We detect the user that is most similar to each user for Patterns 1, 2, 4, and 5; moreover, we detect the user that is most different from each user for Pattern 3. Based on the above procedure, we propose recommendations to users.

C. Comment Recommendation

1) Comment Extraction: In the previous subsection, we explained how to detect users and make recommendations to stimulate communication on a wedding community site. To recommend user comments, we calculate the most related comments from the recommended users that are derived using Eq. (3).

2) Recommendation Interface: Our active communication mechanism recommends users or user comments in different scenarios; we refer to each user pattern in Table II.

This mechanism has two methods of recommending users. The first method recommends users in the comments by generating links to them. The second method recommends users on the top page after login.

For the first method, the interface of recommendation for Patterns 1, 3, and 5, the mechanism generates links in the comments. To generate links in the comments after users have posted, we attach the links of user information or their comments to related words by extracting user characteristics (feature words).

In the second method, the interface of recommendation for Patterns 2 and 4, the mechanism presents users on the top page of the website after login. This mechanism also recommends users on the top page that are likely to share similar experiences. We assume that users prefer to see more users on the top page than in the links generated in the comments.

IV. EVALUATION

In this section, we first extract the actual data from a wedding community site to verify the user characteristic extraction method by extracting feature words of all posts for each user. Second, we detect similar users by comparing the cosine similarity with collaborative filtering.

A. Experiment 1: Verification of User Characteristic Extraction

To evaluate our user characteristic extraction, we extracted feature words of all posts for each user. We compared three methods as follows:

1) *tf*
2) *tf-idf* (*df* = all of users)
3) *tf-idf* (*df* = the users before or after marriage)

We extracted 7,728 terms from 588 user posts.

Table III shows the top-15 feature words for users A, B, C, and D for each method. Bold words denote that feature words are related to these users. We found that many feature words are proper nouns for Methods 2) and 3) such as “Fish paste” and “Limousine”. However, for Method 1), we found common words that all users often use, i.e., there are no effective words that can be considered feature words. We determined that

 ISBN: 978-1-61208-548-7

TABLE III. TOP-15 FEATURE WORDS OF USERS A, B, C, AND D.

User \ Method	1)	2)	3)
A	of, a, ceremony, wedding ceremony, to, sister, I will, heart, family, after, because, to, did, et al., that	sister, wedding ceremony, **earthquake disaster, Fukushima**, bata, fireplace, chaya, sister, attendance, column, heart, family, safety, stop, name	wedding ceremony, sister, **Earthquake disaster**, bata, attendance, heart, **Fukushima**, chaya, fireplace, family, sister, column, 11, safety, influence
B	of, did, better, object, pull, a marriage, I will, he, now, a student, generation, learning, Toyama, now, chestnut	**fish paste**, Toyama, red snapper, gift, girlfriend, object, luck, a student, surprised, age, pull, mountain, form, chestnut, happiness	Toyama, red snapper, **fish paste**, object, gift, girlfriend, luck, a student, surprised, age, mountain, form, happiness, chestnut, woman
C	did, of, better, reach, day, that, friend, friends, ceremony, wedding ceremony, while, a, before, first, good	it seems intriguing, eve, **limousine**, the eve, first meeting, face to face, a van, friend, the other day, reach, move, the previous day, festival, the best	eve, it seems intriguing, **limousine**, first meeting, friend, face to face, the best, a van, move, the previous day, festival, the other day, Hawaii, fellow, reach
D	a, of, did, one, this, now, better, “”, to, about, place, et al., yo, filtration, meeting	reserved, **snow board**, lending, no, alternating current, table, **hair style**, comment, firing, male, rooftop, development, release, frank	reserved, **snow board**, alternating current, male, hair style, table, board, **BGM**, rooftop, firing, girlfriend, in Tokyo, development, comment

calculating with *idf* is a more effective way to extract feature words; however, there are no differences between Methods 2) and 3). The *idf* values imply how the words are generally used by many users; if the *idf* value is high, the word is rarely used among users, and similarly, if it is low, the word is common among users. Therefore, there are no differences between the posts of users before marriage and the posts of users after marriage. Thus, we considered different definitions of document groups, which are not limited to marriage status.

Our results suggest that in the future, we need to remove common words since some generally used words were identified using Methods 2) and 3).

The above discussion confirms that many feature words of users are effectively extracted using *tf-idf* methods, namely, Methods 2) and 3). To detect user characteristics with feature words, more advanced methods are required.

B. Experiment 2: Verification of User Detection

In our active communication mechanism, the similarities between users are the key point for recommending users. In the previous section, we described our classification scheme that classifies users based on similarities of three axes. In this way, we choose the most suitable users to promote communication.

To evaluate the similarities between users, we compared two calculation methods; the first method is the proposed method, specifically, the content-based recommendation method using the cosine similarity with active profile information, and the second method is the item-based recommendation method that uses collaborative filtering with static profile information and marriage status. As mentioned before, we calculated the cosine similarity based on user characteristics, which consist of feature words of each user. Therefore, each user has feature vectors of *tf-idf* values. In Experiment 1, Method 2) is the most useful method for extracting feature words. We also calculated the cosine similarity based on the feature words produced by Method 2). Collaborative filtering is also a method used to calculate similarities between users. This method calculates similarities using user login information as items of each user. It is mainly used to recommend other items to users according to the following formula:

$$Sim(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}. \quad (4)$$

TABLE IV. COSINE SIMILARITY AMONG 588 USERS.

value	#user combinations
0 - 0.1	154,132
0.1- 0.2	16,158
0.2 - 0.3	2,022
0.3 - 0.4	209
0.4 - 0.5	46
0.5 - 0.6	7
0.6 - 0.7	4
0.7 - 1.0	0

This equation calculates the similarity between users X and Y. On a wedding community site, users create individual accounts by answering questions about their wedding planning. For example, “Do you agree with a simple style marriage?” For each question, a user may choose from one of the following responses: ”Strongly disagree,” ”Disagree,” ”Neither disagree nor agree,” ”Agree,” or ”Strongly agree.” Each of these responses was assigned a numerical value ranging from 1 to 5, respectively, for calculation purposes. We then calculated the similarities using these numbers. Note that $\bar{x}$ and $\bar{y}$ denote the averages of the chosen answers, for example, if a user chose answers 1 to 5, the average value would be 3.

The users evaluated for our proposed user characteristic extraction are shown in Table III. For this evaluation, we calculated 172,578 combinations from 588 users; the value of the cosine similarity ranges between 0 and 1.

Table IV shows the distribution of results of the cosine similarity. The average value of all combinations is 0.045. We found that many results of user combinations are below 0.1. This can be attributed to the fact that most users talk about different topics related to their wedding planning. However, some user combinations induce a high cosine similarity.

Table V shows the distribution of results of collaborative filtering. The value of collaborative filtering should be between -1 and 1. For this method, the values are calculated based on the answers from the questions regarding wedding planning when users create accounts on the wedding community site. A high value implies the users have similar wedding planning ideas. For this evaluation, we calculated 435 combinations of 30 users. The average value of all combinations was 0.304, which confirms that many users have similar wedding planning tastes.

 ISBN: 978-1-61208-548-7

TABLE V. COLLABORATIVE FILTERING AMONG 30 USERS.

value	#user combinations
-1.0 - -0.9	0
-0.9 - -0.8	0
-0.8 - -0.7	2
-0.7 - -0.6	4
-0.6 - -0.5	5
-0.5 - -0.4	8
-0.4 - -0.3	8
-0.3 - -0.2	15
-0.2 - -0.1	23
-0.1 - 0	26
0 - 0.1	35
0.1- 0.2	39
0.2 - 0.3	40
0.3 - 0.4	38
0.4 - 0.5	41
0.5 - 0.6	41
0.3 - 0.7	44
0.7 - 0.8	31
0.8 - 0.9	26
0.9 - 1.0	9

Based on these results, we compared two similarity calculation methods. Here, we focused on user E, who has a high cosine similarity with other users and often posts on a wedding community site as a main user. We calculated all combinations with user E; therefore, there were a total of 588 values of the cosine similarity and 588 values of collaborative filtering.

Fig. 3 shows the distribution of the cosine similarity and collaborative filtering for 10 users, specifically, users E, H, I, J, K, L, M, N, O, and P. Each dot corresponds to one user and has two values: the cosine similarity with each user, and the collaborative filtering with each user. The vertical axis corresponds to the values of the cosine similarity, and the horizontal axis corresponds to the values of collaborative filtering. We focused on two users, specifically, F and G for user E. Both of these users have high cosine similarity values above 0.6, but their values of collaborative filtering are 0 and 0.54, respectively.

First, we compared the posts of users E and F. A post by user E describes their cousins' impressive wedding with the grooms' tears. On the other hand, a post of user F describes how their cousins' wedding was organized. Even though common words were used in their posts, the meanings of these sentences and their topics are different.

Second, we compared the posts of users E and G. The post from user E is the same post mentioned above. A post from user G describes their cousins' wedding with tears because of a letter about a grandmother who was gone. These posts both mention the same type of wedding and their cousins' weddings with tears, even though the content of these posts is slightly different.

As a result, we found that only calculating the cosine similarity is not effective to detect similar comments. However, we found that calculating both the cosine similarity and collaborative filtering are effective. Therefore, these two methods can help detect similar user comments to evoke communication among users. However, we still must evaluate other situations of users with other users' axes and marriage statuses.

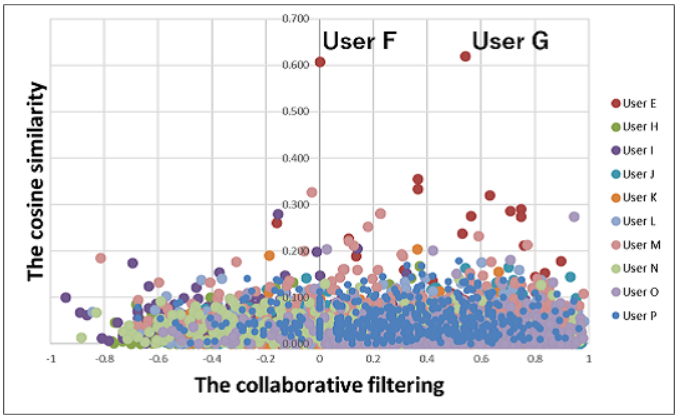

Figure 3. Distribution of the cosine similarity and collaborative filtering 1.

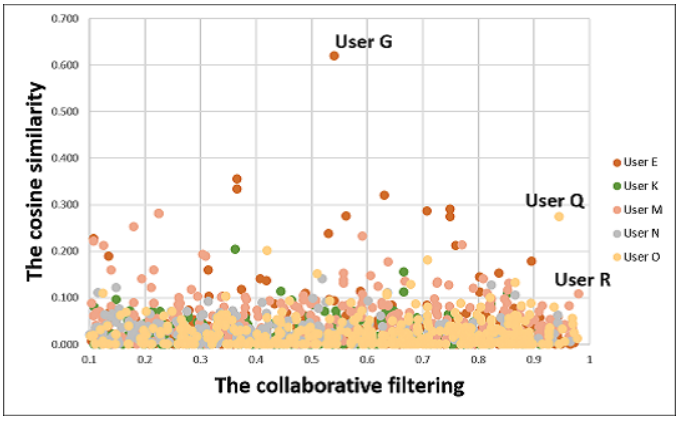

Figure 4. Distribution of the cosine similarity and collaborative filtering 2.

Fig. 4 briefly shows the distribution of the cosine similarity and collaborative filtering for users E, K, M, N, and O. We found several users that are especially similar to these users such as users Q and R. In future, we plan to propose methods for clustering with the cosine similarity and collaborative filtering.

V. CONCLUSION

In this paper, we proposed an active communication mechanism for a wedding community site. This mechanism recommended 1) users who may potentially evoke communication and 2) their comments. To detect users, this mechanism classified all users into three axes, specifically, "Static Profile Information," "Marriage Status," and "Active Profile Information." We then calculated the similarities between users using the cosine similarity. To extract comments that were posted on a wedding community site by recommended users, our mechanism detected the most related comments. Finally, we evaluated the user characteristic extraction from posts by comparing *tf-idf* methods and evaluated similarity calculation methods with the cosine similarity and collaborative filtering.

In the future, we plan to enhance the proposed method based on our experimental results and evaluate the effects of user recommendations. Furthermore, we plan to extract the relationships between users by constructing a matrix based on user behavior, as in our previous work [10].

 ISBN: 978-1-61208-548-7

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grant Numbers 15K00162 and 16H01722.

REFERENCES

[1] U. Pfeil, R. Arjan, and P. Zaphiris, "Age differences in online social networking–a study of user profiles and the social capital divide among teenagers and older users in myspace," Computers in Human Behavior, vol. 25, no. 3, 2009, pp. 643–654.

[2] M. D. Roblyer, M. McDaniel, M. Webb, J. Herman, and J. V. Witty, "Findings on facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites," The Internet and higher education, vol. 13, no. 3, 2010, pp. 134–140.

[3] R. Isaac and J. Walker, "Communication and free-riding behavior: The voluntary contribution mechanism," Economic Inquiry, vol. 26, no. 4, 1988, pp. 585–608.

[4] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook "friends:" social capital and college students' use of online social network sites," Journal of Computer-Mediated Communication, vol. 12, no. 4, 2007, pp. 1143–1168.

[5] Y. Mastsui and Y. Kawai, "Social search system for retrieval and communication based on networks of pages and users," Computer Software, vol. 28, no. 4, 2011, pp. 196–205.

[6] D. Zhou, M. Truran, T. Brailsford, H. Ashman, and A. Pourabdollah, "Llama-b: Automatic hyperlink authoring in the blogosphere," in Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia, ser. HT '08. New York, NY, USA: ACM, 2008, pp. 133–138. [Online]. Available: http://doi.acm.org/10.1145/1379092.1379119

[7] R. D. souza, A. Kulkarni, and I. A. Mirza, "Automatic link generation for search engine optimization," International Journal of Information and Education Technology, vol. 2, no. 4, 2012, pp. 401–403, ISSN: 2010-3689.

[8] J. Liu, E. Pedersen, and P. Dolan, "Personalized news recommendation based on click behavior," Proceedings of the 15th International Conference on Intelligent User Interfaces, 2010, pp. 31–40.

[9] H. Akihiro, Y. Madoka, T. Hiroyuki, and K. Yahiko, "Experimental chat sessions to encourage active communication in lectures," IPSJ SIG Notes, vol. 2000, no. 45, 2000, pp. 61–66.

[10] Y. Wang, Y. Kawai, S. Miyamoto, and K. Sumiya, "An automatic scoring system for e-reports based on student peer evaluation using groupware," DBSJ Journal, vol. 13, no. 1, 2015, pp. 71–76.

 ISBN: 978-1-61208-548-7