An Active Communication Mechanism based on User Behavior Analysis for a Wedding Community Site

Toshinori HAYASHI†, Yuanyuan WANG††, Yukiko KAWAI†††, and Kazutoshi SUMIYA†

† Kwansei Gakuin University,
2–1, Gakuen, Sanda-shi, Hyogo, 669–1546 Japan
†† Yamaguchi University,
2–16–1, Tokiwadai, Ube-shi, Yamaguchi, 755–8611 Japan
††† Kyoto Sangyou University,
Kamigamo Motoyama, Kita-ku Kyoto-shi, 603–8047 Japan
E-mail: †{anc93184,sumiya}@kwansei.ac.jp, ††y.wang@yamaguchi-u.ac.jp, †††kawai@cc.kyoto-su.ac.jp

Abstract This paper presents an active communication mechanism based on user behavior analysis on wedding community sites. For this, we propose a novel mechanism for activation of users' communication which provides related comments and users by detecting knowledge and interests from archived comments on a wedding community website to evoke conversation among users. This mechanism has three components, 1) extracting user login information and user characteristics of their posts to know what they need to communicate, 2) detecting and recommending users who are likely to communicate with each other, 3) recommending comments which users may be interested in. Through proposed activation mechanism, users on the wedding community site can communicate with each other easily and efficiently. In this paper, we discuss our proposed user characteristic extraction and user recommendation methods using actual user posts from a wedding community website.

Key words user behavior analysis, wedding community site, communication

1. Introduction

In recent years, several researches using data from Social Networking Service (SNS) has been done, it is important to collect much data from SNS community site, such as Facebook, LINE, and some Q&A sites. However, they focused on data collection, cannot promote user communication on community websites because of differences in values. In this paper, we aim to promote user communication focusing on a wedding community site by recommending appropriate users and their comments.

In this paper, we propose a novel active communication mechanism which share the comments of users by considering their knowledge and interests by analyzing their behavior on community websites. To achieve this, we first extract all posts of each user and extract their feature words by using *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 the posts (Fig. 1). To use this mechanism, users can communicate with other users who are recommended to them about wedding planning, and it will promote communication among users on the wedding community site.

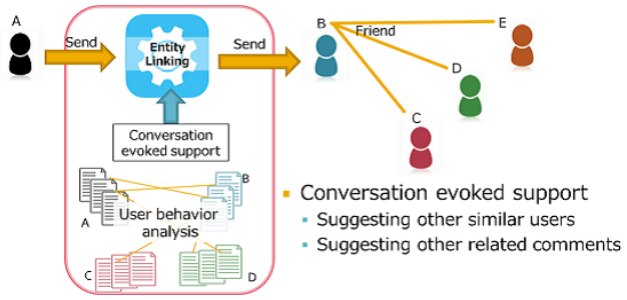

Figure 1 User and comments recommendation for activation of user communication based on user behavior analysis

The rest 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 by using a real dataset of a wedding community site. Finally, Section V concludes the paper and outlines our future work.

2. System Overview and Related Work

2.1 An Active Communication Mechanism

This paper presents an active communication mechanism

based on user behavior analysis on wedding community sites. This mechanism has 3 steps, 1) user login information and user characteristic extraction, 2) user detection and recommendation, 3) comment recommendation (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. There are some users logged in the wedding community site, when they plan to hold wedding. The website is created for observing users' needs about marriage. On this website, there are some threads for wedding planning in different marriage statuses, users can freely post the comments to each thread. There is only one was the way to communicate with other users by replying to other users' comments on the thread. To improve it, we propose the method to recommend both users and their comments by analyzing user behavior and their profile information on the wedding community site. The goal is for our active communication mechanism to figure out other users whom the users want to communicate with.

The wedding community site is not a "a Question & Answer site", is a website where users can share their positive opinions and experiences about wedding. The proposed system will recommend other users who have similar situation or values of marriage to evoke communication between users. This system can be used on other community websites. Since the proposed system is considered on a wedding community site, it uses the static information which the users registered by answering several questions about ideal wedding ceremony when they first create their accounts.

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

2.2 Related Work

Issac et al. [1] mentioned that communication is important to discuss things and to work with others as a group. They said that communication makes people willing to contribute for society on their study. It is effective for communication on websites. Ellison et al. [2] focused on the SNS communities. According to these studies, communicating with others on SNS makes more people feel happy.

In our previous work [3], users can communicate with each other when they search for web pages. In this work, we extend our previous work, to recommend users and their comments based on link generation for a wedding community site. Although several automatic link generation methods for websites have been studied [4], [5], they have focused on web pages for knowledge support only, they do not consider the communication among users. To solve it, our proposed method recommends users to evoke communication. Other studies about recommendation by analyzing user behavior on news sites [6], did not consider the relationships between users. In this paper, we first extract users' posts to analyze user behavior, and detect users to recommend by extracting the relationships between the users.

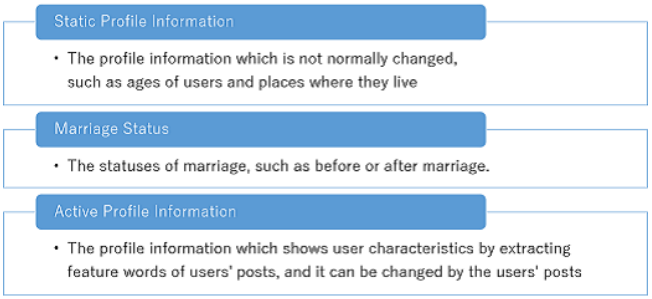

Figure 2 3 axises of user behavior

Akihiro et al. [7] conducted an experiment for activate 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' users.

3. Active Communication Mechanism for a Wedding site

3.1 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 [3], we knew that 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 status or situation. Therefore, in order to recommend users, we analyze the user behavior by using 3 axises (see Fig. 2). These are "Static Profile Information", "Marriage Status", and "Active Profile Information".

3.1.1 User Login Information Extraction

We extract user login information by acquiring the user registration information on the wedding community site that users input the first time they register on the website. Since all users are required to register their information at first, when they use the wedding community site, such as their ages, places where they live, and marriage status. We divide the user login information as users' static profile information and marriage status.

3.1.2 User Characteristic Extraction

We extract user characteristics by extracting all posts of

Table 1 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 advices
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 advices

Table 2 Recommendation situation for each User Pattern

Pattern	Purpose	When	How
1	Give advises	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 advices	Links are generated in the comments	○○ can be a good adviser for you

each user. Next, we calculate term frequency and document frequency based on the *tf-idf* method by the following formulae:

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

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

Here, $n_{i,j}$ denotes the term frequency of a word t_i in the document d_j. In this work, d_j denotes the document which is integrated by all posts of one user. Therefore, the number of documents as the number of users on the wedding community site. $\sum_k n_{k,j}$ denotes the sum of term frequency of all words in the document d_j. $|D|$ denotes the number of all documents, which also means the number of users. df_i denotes the number of the documents which include the word t_i.

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

3.2 User Detection and Recommendation

3.2.1 User Detection

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

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

Here, $\overrightarrow{x}$ 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 the absolute value, such as "before marriage" or "after marriage". Therefore, it will not change against any users. However, "Static Profile Information" and "Active Profile Information" are the relative values. it will change depending on each user.

3.2.2 User Recommendation

We recommend users by considering users who have similar situation to communicate with each other easily to share their experiences or advices. Based on 3 axises we described in the previous subsection, we classify five useful patterns of users on the wedding community site (see Table 1).

We detect the most similar user of each user for patterns 1, 2, 4, 5, and we detect the most different user of each user for pattern 3. Based on the above, these detected users will be recommended.

3.3 Comment Recommendation

3.3.1 Comment Extraction

In previous subsection, we explained how to detect users and recommend them to stimulate communication on the wedding community site. To recommend users' comments, we calculate the most related comments of the recommended users who are derived by using Eq. (3)

3.3.2 Recommendation Interface

Our active communication mechanism recommends users or users' comments in different scenes refer to each user pattern in Table 2.

This mechanism has two ways to recommend users. the first one is recommend users in the comments by generating links to them. The second one is recommend users on the top page after login.

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

In the second one, the interface of recommendation for patterns 2 and 4, this mechanism presents users on the top page on the website after they log in. This mechanism also recommends users whom they can share the experiences on the top page. We assume that the users would like to see more users on the top page than the links generated in the comments.

4. Evaluation

In this section, we first extract the actual data from a wedding community site to verify the user characteristic ex-

Table 3 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, aftter, Because, To, did, Et al., That	sister, Wedding ceremony, **Earthquake disaster, Fukushima**, Bata, Irreplace, Chaya, sister, Attendance, Column, heart, Family, Safety, Stop, name	Wedding ceremony, sister, **Earthquake disaster**, Bata, Attendance, heart, **Fukushima**, Chaya, Irreplace, 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

traction method by extracting feature words of all posts of each user. Secondly, we detect similar users by comparing the cosine similarity with the collaborative filtering.

4.1 Experiment 1: Verification of User Characteristic Extraction

To evaluate our user characteristic extraction, we extracted feature words of all posts of each user. We compared 3 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 posts of users.

Table 3 shows top-15 feature words of users A, B, C, and D by each method. Bold words denote the feature words are related to these users. We found that many feature words are proper nouns on the methods 2) and 3), such as “Fish paste” and “Limousine”. However, on the method 1), we found common words that all users often use, there are no effective words as feature words. We considered that calculating with *idf* is more effective way to extract feature words. However, there are no differences between the methods 2) and 3). The *idf* values imply how the words generally used by many users. If the value of *idf* is high, it implies the word is rare among the user. If it is low, it implies the word is common among them. Because of them, there are no differences between the posts of users before marriage and the posts of users after marriage. We considered that different definitions of document groups which are not only before and after marriage.

In the future, the results suggest that we need to remove the common words, since there still have some generally used words by using the methods 2) and 3).

As discussed above, we could confirm that many feature words of users are extracted effectively by using *tf-idf* methods 2) and 3) . To detect the user characteristics with feature words, more advanced methods are required.

4.2 Experiment 2: Verification of User Detection

In our active communication mechanism, the similarities between users is the key point for recommending users. In previous section, we described that we classify the users by the similarities of 3 axises to choose the most suitable users to promote communication.

To evaluate the similarities between users, we compared two calculation methods, the first is the proposed method as the content-based recommendation method by using cosine similarity with the active profile information, and the other one is the item-based recommendation method by using the collaborative filtering with the static profile information and marriage status. As mentioned before, we calculated the cosine similarity based on the user characteristics which consist of feature words of each user. Therefore, each user has the feature vectors by *tf-idf* values. In Experiment 1, the method 2) is the most useful method for extracting feature words. Then, we calculated the cosine similarity based on the feature words by the method 2). The collaborative filtering is also a method to calculate similarities between users. This

Table 4 The 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

method calculated the similarities by using the user login information as items of each user. Therefore, it mostly used for recommending other items to the users by the following formula:

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

Here, it calculates the similarity between users X and Y. On a wedding community site, the users create own accounts with answering questions about their wedding planning. For example, "Do you agree that the marriage with simple style?" They can choose one answer that is "Strongly disagree", "Disagree", "Neither disagree nor agree", "Agree" or "Strongly agree" for each question. We replaced the numbers 1 to 5 for those answers. Then, we calculated the similarities by using these numbers. $\bar{x}$ and $\bar{y}$ denote the average of the answer they chosen, which means it must be as "3" in this case as they chosen the answers 1 to 5.

The users that we evaluated for our proposed user characteristic extraction are shown in Table 3. On this evaluation, we calculated totally 172,578 combinations of 588 users, and the value of the cosine similarity should be between 0 to 1.

Table 4 shows the distribution of the 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. Because most users talk about different things for their wedding planning. In addition, some user combinations which has high cosine similarity.

Table 5 shows the distribution of results of the collaborative filtering. The value of the collaborative filtering should be between -1 to 1. On this method, the values was calculated by the answers of the questions about wedding planning when users create the accounts on this wedding community site. The high value means the users have similar wedding planning. On this evaluation, we calculated 435 combinations of 30 users. The average value of all combinations is 0.304. We confirm that many users have similar wedding planning.

Table 5 The 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
0.6 ∼ 0.7	44
0.7 ∼ 0.8	31
0.8 ∼ 0.9	26
0.9 ∼ 1.0	9

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

Fig. 3 shows the distribution of the cosine similarity and the collaborative filtering for 10 users, E, H, I, J, K, L, M, N, O, and P. Each dot means one user and it has two values: the cosine similarity with each user, and the collaborative filtering with each user. The vertical axis implies the values of the cosine similarity, and the horizontal axis implies the values of the collaborative filtering. We focused on two users, F and G for user E. Both of them has high cosine similarity which are over 0.6, but the values of the collaborative filtering are 0 and 0.54.

Firstly, we compared the posts of the users E and F. The post of the user E describes the cousins' impressive wedding with grooms' tears. On the other hand, the post of the user F describes how cousins' wedding was organized. Even though, common words were used on their posts, the meanings of the sentence and the topics were different.

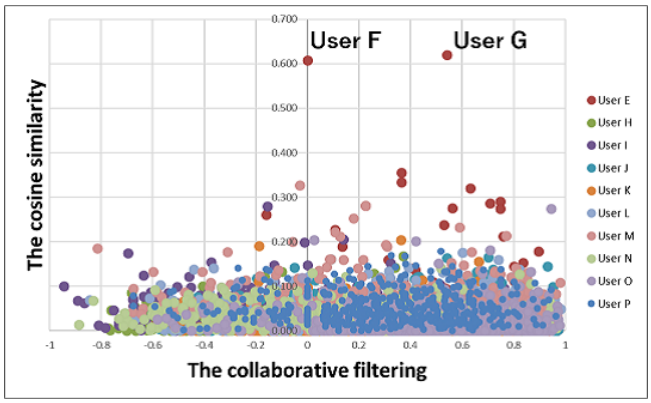

Figure 3 Distribution of the cosine similarity and the collaborative filtering E, H∼P

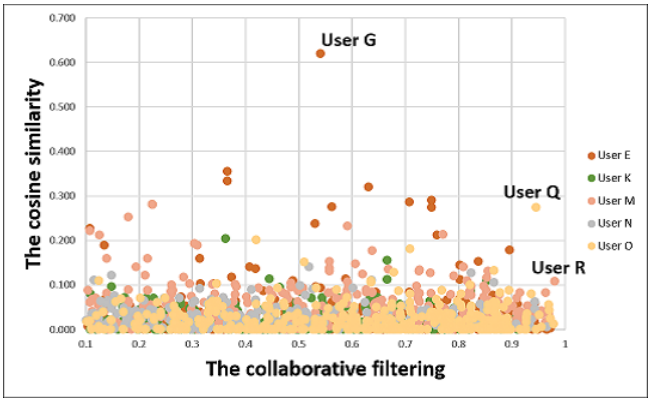

Figure 4 Distribution of the cosine similarity and the collaborative filtering E, K, M∼O

Secondly, we compared the posts of the users E and G. The post of the user E was mentioned above. The post of the user G describes the cousins' wedding with tears because of the letter that is about grand mother who was gone. These posts of them mentioned about the same type of wedding, the cousins' wedding with tears, even though the content of posts are a little different.

As a result, we found that only calculating the cosine similarity is not effective to detect the similar comment. However, we found that calculating both the cosine similarity and the collaborative filtering is effective. Therefore, these two methods can help us to detect users' similar comments to evoke communication among the users. Through this result, we still need to evaluate other situation of users with other users' axises, marriage status.

Fig. 4 shows the distribution of the cosine similarity and the collaborative filtering for 5 users, E, K, M, N, and O briefly. We could find several users who are especially similar to them such as users Q and R. As a future work, we are planning to propose methods for clustering with the cosine similarity and the collaborative filtering.

5. Conclusion

In this paper, we proposed an active communication mechanism for a wedding community site. This mechanism recommends 1) users who are supposed to evoke the communication, and 2) their comments. To detect users, this mechanism classifies all users in 3 axises, which are "Static Profile Information", "Marriage Status", and "Active Profile Information". Then, we calculated the similarities between users by using the cosine similarity. To extract comments which were posted on the wedding community site by the recommended users, this mechanism detects the most related comments. Finally, we evaluated the user characteristic extraction from posts by comparing *tf-idf* methods and evaluated the similarity calculation methods with the cosine similarity and the collaborative filtering.

In the future, we plan to enhance the proposed method based on the experimental results, and evaluate the effects of user recommendation. Furthermore, we will extract the relationships between users by constructing the matrix based on user behavior as our previous work. [8].

Acknowledgment

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

References

[1] Isaac, R. M. and Walker, "J. M., COMMUNICATION and FREE-RIDING BEHAVIOR: THE VOLUNTARY CONTRIBUTION MECHANISM," Economic Inquiry, 26: 585608, 1988.

[2] Ellison, N. B., Steinfield, C. and Lampe, C., "The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites," Journal of Computer-Mediated Communication, 12: pp. 1143–1168, 2007.

[3] Yuya Matsui, Yukiko Kawai, "Social Search System for Retrieval and Communication based on Networks of Pages and Users," Computer Software, 28(4), pp. 196–205, 2011.

[4] Dong Zhou, Mark Truran, Tim Brailsford, Helen Ashman, Amir Pourabdollah, "LLAMA-B: Automatic Hyperlink Authoring in the Blogosphere," Proc. of ACM HT ' 08, pp. 133–137, 2008.

[5] R. D'souza, A. Kulkarni, M. A. Imran, "Automatic Link Generation for Search Engine Optimization," International Journal of Information and Education Technology, Vol. 2, No. 4, pp. 401–403, 2015.

[6] J. Liu, P. Dolan, E. R. Pedersen, "Personalized news recommendation based on click behavior," Proc. of IUI 2010, pp. 31–40, 2010.

[7] Akihiro Hatanaka, Madoka Yuriyama, Hiroyuki Tarumi, Yahiko Kambayashi, "Experimental Chat Sessions to Encourage Active Communication in Lectures," Information Processing Society of Japan (IPSJ), Vol.45, 61-66, 2000.

[8] Yuanyuan Wang, Yukiko Kawai, Setsuko Miyamoto, Kazutoshi Sumiya, "An Automatic Scoring System for E-Reports based on Student Peer Evaluation using Groupware," DBSJ Journal, Vol. 13, No. 1, pp. 71–76, 2015.