Activation Mechanism for Recommending Appropriate Users and Comments 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 websites. To this end, we propose a novel mechanism for the activation of user communications, which suggests other users and their related comments by detecting knowledge and interests from archived comments. Such information on a community website evokes conversations among users. We focus on a wedding community website in this study. The proposed mechanism consists of the following three components: 1) the profiling of user login information, such as users' ages and locations, and extraction of user characteristics, such as their interests and intentions to communicate; 2) the detection and recommendation of users who are likely to communicate with each other; and 3) the recommendation of comments that may be of interest to a user. Through the proposed activation mechanism, users on a wedding community website can easily find other users who share similar experiences, and engage in active communication with them. We discuss our proposed user characteristic extraction and user recommendation methods using actual users' posts from a wedding community website, and also discuss applications for e-learning. In addition, we evaluated the activation of user communications using our proposed system through interviews.**

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

I. INTRODUCTION

In recent years, research has been conducted using data from social networking services (SNSs). In this context, it is important to collect as much data as possible from SNS community websites, such as Facebook, LINE, and other Q&A sites. However, such services that focus on data collection cannot activate user communications on community websites, because of differences in human values. In this paper, we focus on conversations on a wedding community website. Moreover, we focus on the problem that users cannot find appropriate users to communicate with. We aim to evoke user communications by recommending appropriate users and comments, considering human values. Our proposed system analyzes users' situations from user login information and user preferences by extracting user characteristics, because we assume that human values consist of situations and preferences. This research constitutes an extension to the work in [1], which we presented in Venice in April 2017.

Wedding community websites are community sites for sharing and obtaining information regarding wedding planning, and are completely distinct from dating websites, the purpose of which is to meet a partner for dating. We focus on a wedding community site that shares the same concept as other kinds of bulletin board systems (BBSs). On this community site, users can post their opinions and experiences on a conversation thread created by an administrator. Each conversation thread has questions that ask about weddings as discussion themes, such as "how did you choose your wedding location?" Thus, users post their opinions and experiences in response to these questions, and exchange information. However, conversations between users are not active in these threads, because it is difficult for users to find others to communicate with. Our proposed system solves this problem by finding appropriate users based on user login information and characteristics.

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

We also propose an active communication mechanism for e-learning, in order to confirm the possibility of applying our proposed system to other community websites. According to some studies regarding online discussions [2][3], active communications among users on e-learning is an effective mechanism for students to learn knowledge efficiently. To extend the effect of our proposed method, we aim to apply it to e-learning.

To evaluate this system, we use the actual data from a wedding community website. This is sorted and processed to provide recommendations to users for evaluation, in order to verify whether this mechanism is successful. Then, we interview five users to ask how helpful the proposed system is. Although we could not interview many actual users, we performed a qualitative evaluation through these interviews.

The remainder of this paper is organized as follows. Section

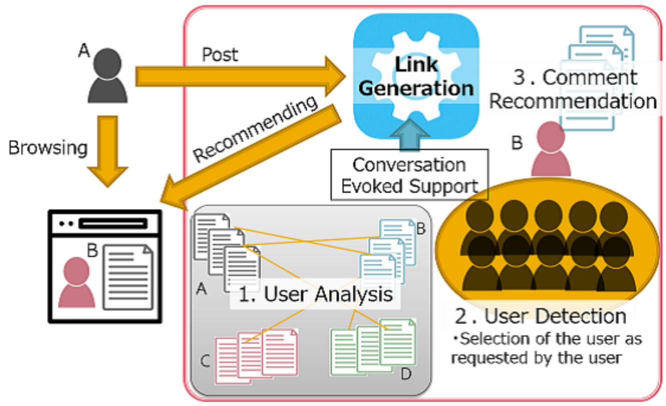

Figure 1. User and comment recommendations for the activation of user communications based on a user behavior analysis.

II provides an overview of our system and reviews related work. Section III explains how users and their comments are recommended on a wedding community website. Section IV explains the application of our proposed methods to e-learning. Section V presents the experimental results obtained using a real dataset from a wedding community website. Finally, Section VII concludes the paper, and Section VI outlines our future work.

II. System Overview and Related Work

A. Active Communication Mechanism

We present an active communication mechanism based on user behavior analysis on wedding community websites. This mechanism consists of the following three steps (see Figure 1):

1) User login information and user characteristic extraction.
2) User detection and recommendations.
3) Comment recommendations.

To use this mechanism, users are required to install a toolbar (a browser plug-in) on an existing wedding community website in Japan. Wedding community websites are generally utilized by people who plan to hold a wedding, and are intended to assess a couples' requirements regarding the wedding. On the website we focus on in this study, there are conversation threads for wedding planning for various marriage statuses, and users can freely post their comments on each thread. The website addresses users' anxieties and troubles. For example, a user can create a post such as "I do not know what I should be careful about in planning for a wedding." Users can also share their experiences, such as impressions and enjoyments. For example, a user could post "I had an amazing wedding overseas."

The only way to communicate with other users is by posting comments as replies to other users' comments on threads. However, users rarely post replies on other users' comments, because they cannot find appropriate users to communicate with. To raise the number of replies, we propose a method that recommends both users and their comments by analyzing user behavior and profile information on a wedding community website to activate conversations.

A wedding community website is not a "question & answer" site; rather, it is a website where users can share their opinions and experiences regarding weddings. The proposed system will recommend other users who have been in similar situations or have shared values regarding marriage, in order to evoke communications between users. This system can also be adopted on other community websites. However, because the proposed system is considered on a wedding community website, it uses user login information entered by a user during their initial registration regarding their ideal wedding ceremony.

Figure 1 presents an overview of our proposed mechanism. After a user posts, the mechanism analyzes their behavior and recommends other users by calculating the similarities between them. After a user posts information, it will pass through the link generation system. This system will generate information to evoke conversations. First, it extracts all of the posts of users, and analyzes user behavior through feature words. Then, it detects recommended users by performing selections as requested by the user. To detect a recommended user, we categorize each user by three kinds of profile information. These methods are explained in Section III.

B. Related Work

Issac et al. [4] noted that communication is important for discussing various topics and working with others as a group. They mentioned that communication makes people more willing to contribute to society. Moreover, communication on websites is also effective, not only face-to-face communication. Ellison et al. [5] focused on SNS communities. L. H. Shaw et al. found that Internet communication decreases loneliness and depression significantly [6]. According to these studies, communicating with others on SNSs makes people feel happier.

In our previous work, users communicated with each other when searching for web pages [7]. In this work, we extend our previous work to recommend users and comments based on link generation for a wedding community site.

Knowledge extraction from online communities [8] have also been researched. Park et al. [9] confirmed that the knowledge users obtain through online communications is connected to their purchasing behavior. They proved that online communities provide users with important information regarding purchase behavior. Randhawa et al. [10] improved the utility of knowledge collaboration between organizations and online communities for open innovation (OI). There have been many studies concerning the use of data from online communities [11][12][13][14] to discover new knowledge.

Our proposed system activates communications on online community websites. Akihiro et al. [15] conducted an experiment concerning active communication in e-lectures through a chat system. However, this did not work very effectively, 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 to other users with various marriage statuses.

Some researchers [16][17][18] analyzed effects of conversation from online community. Hemmings-Jarrett et al. [19] measured the effect of online communications with a focus on political topics, comparing before and after the event. They assumed that the activity levels of users on online conversations were related to their opinion. However, this research was not intended to activate conversations.

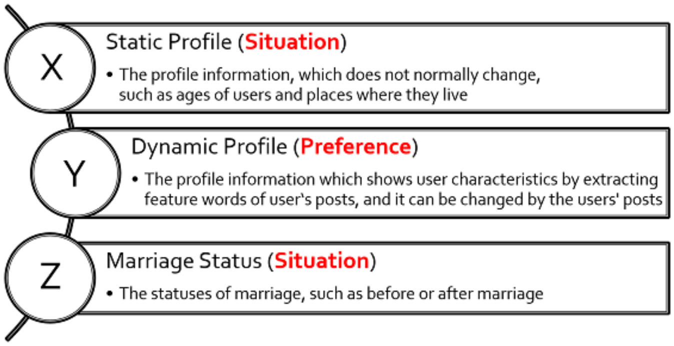

Figure 2. Profiling based on user aspects.

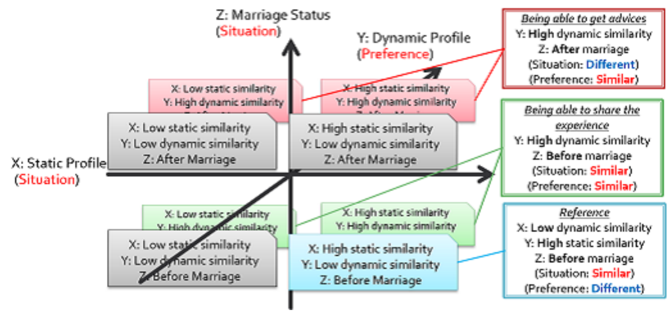

Figure 3. User detection when the original user is not yet married.

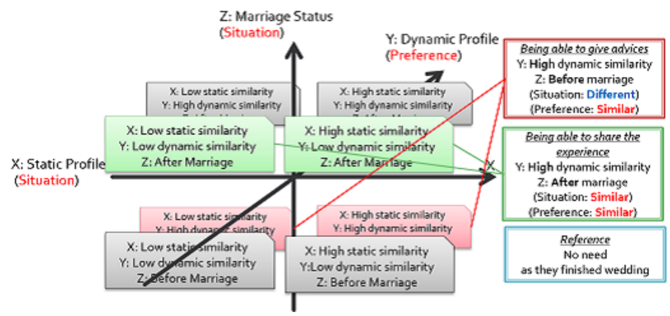

Figure 4. User detection when the original user is already married.

Other studies that have recommended analyzing user behavior on news websites [20][21] did not consider the relationships between users. In this paper, we first extract user posts to analyze user behavior, and then detect users to recommend by extracting the relationships between users.

Our proposed method involves activating user communications by recommending appropriate users and their comments. Xu et al. [22] proposed a novel user recommendation system based on users interests and networks of relationships. However, there have not been many studies regarding user recommendation systems that focus on specific community websites.

Although several automatic link generation methods for websites have been studied [23][24], these have primarily focused on web pages for knowledge support only; they did not consider communications among users. To address this deficiency, our proposed method recommends users in order to evoke communications.

III. Active Communication Mechanism for Wedding Websites

A. User Behavior Analysis on a Wedding Community Website

To evoke communication among users, our active communication mechanism recommends users and their comments by analyzing user behavior on a wedding community website. According to our previous work in [7], users can help other users when they search for the same web pages. Furthermore, in general users can communicate with each other easily when they share similar statuses or situations and have similar preferences. Therefore, in order to recommend users we analyze users' humans values by considering three kinds of profile information based on aspects of a wedding community website (see Figure 2). In particular, we consider the axes of "Static Profile Information," "Marriage Status," and "Dynamic Profile Information."

"Static Profile Information" and "'Marriage Status" indicate a user's situations. "Dynamic Profile Information" indicates a user's preferences. This system analyzes users' human values based on their situations and preferences. Static profile information is generated from user login information, such as age and location, which does not normally change. Marriage status implies a user's position regarding a wedding, such as before or after marriage. Dynamic profile information is generated from the extraction of user characteristics. This uses the data from posts from each user, and so this information often changes.

On wedding community websites, marriage status is the most important factor in analyzing users, because this implies a big difference in the purpose of use. The purpose of using this website for users who are preparing for a wedding is to obtain information regarding weddings. On the other hand, for users who have already finished their wedding, the purpose is to share information and their experiences of their wedding to provide advice and help others. Thus, "Marriage Status" is the most important factor in the user analysis for categorizing users.

1) User Login Information Extraction: We extract user login information by acquiring the information that users input during registration on a wedding community website. Users input information such as their age, area where they live, and marital status. We divide this user login information into user static profile information and marital status.

2) User Characteristic Extraction: We extract user characteristics by extracting all posts for 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 *TF* indicates term frequency, *IDF* indicates inverse document frequency, and $n_{i,j}$ denotes the term frequency of the word t_i in document d_j. In this work, d_j denotes the document that is integrated all posts of one user. Therefore,

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

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 documents that include the word t_i.

Based on the above, we use the obtained *TF-IDF* values and feature words for each user to determine a user's active profile information. This information changes every time a user posts on a thread on the wedding community site, and so users' dynamic profile information is normally changed by user posts.

B. User Detection and Recommendation

1) User Detection: "Marriage Status" takes an absolute value of either "preparing for wedding" or "finished wedding." Therefore, there are only two kinds of value. However, "Static Profile Information" and "Dynamic Profile Information" are represented by relative values. These will vary depending on each user.

To detect a recommended user, the system needs to calculate the similarity for each axis. First, the method of calculating the "Static Profile Information" depends on the dataset. We will explain this in Section . Second, the similarity for "Marriage Status" is very simple, as this consists of only two kinds of status, so that the relationship between users for "Marriage Status" must either be the same or different. Third, we calculate the similarities for "Dynamic Profile Information" between users using 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, and $|V|$ is the number of dimensions for the feature vector.

Thus, this system calculates the similarities between users by using three kinds of profile information. Based on these similarity values, the system detects recommended users for each original user.

2) User Recommendation: We recommend users to evoke communications with others by considering users who are in similar situations. Such users may easily relate to each other and share their experiences or advice. As mentioned in the previous section, we use three kinds of profile information: "Static Profile Information," "Marriage Status," and "Dynamic Profile Information." Considering the relationships between users, each type of profile information has a high or low similarity. However, we assume that "Marriage Status" has four possible combinations: "preparing for wedding" and "preparing for wedding," "preparing for wedding" and "finished wedding," "finished wedding" and "preparing for wedding," and "finished wedding" and "finished wedding." A total of 16 kinds of user can be extracted by the system. We separate these into two cases: where the original user is preparing for a wedding and where the original user has already finished a wedding.

Figure 3 illustrates the case in which the original user is preparing for a wedding. There are eight kinds of user that the system can detect in this case, based on three kinds of profile information. The idea of recommending users is to suggest an appropriate user who is able to provide advice or share related experiences, or could be a reference for a different style of wedding.

For a user to be recommended who can provide advice, they should have a high dynamic similarity, which means that their preferences should be similar. Because they have similar preferences, the recommended user can provide advice if they have already finished their wedding.

For a recommended user to be able to share their experiences, they should have high similarity for dynamic profile information, which means that their preferences are similar. In addition, the marital status should be the same for the users to be in the same situation. We assume that users who are in a similar situation with similar preferences can communicate easily to share their experiences.

To recommend a user who could be a reference, that user should be in a similar situation and have different preferences. To activate communications among users, recommending similar users is important, but users are also interested in others who are different. For instance, even though a user has an ideal style of wedding in mind, this might be changed by referring to others. A user recommended to another for reference has to have a low similarity for dynamic profile and high similarity for static profile information.

Figure 4 illustrates the case in which the original user has already finished their wedding. There are eight kinds of users in this case, based on three types of profile information. The idea of recommending users is the same as in 3; however, we assume that a user who has finished their wedding does not use this website for reference, because the motivation of these users is to provide advice or share experiences with other users.

Based on the three axes described in the previous subsection and the theory of recommending users, we classify five useful patterns of users on a wedding community website (see Table I).

For each user, we detect the other user that is most similar to them for Patterns 1, 2, 4, and 5. Moreover, we detect the user that is most different 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 in order to stimulate communications on a wedding community website. To recommend user comments, we calculate the most closely related comments from the recommended users that are derived using Eq. (3). The recommended comments suggest why the recommended user is relevant for the original user.

2) Recommendation Interface: Our active communication mechanism recommends other appropriate users and their comments in different scenarios corresponding 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 following login.

For the first method, the interface provides recommendation for Patterns 1, 3, and 5. The recommended users for these patterns have similar preferences. We expected that recommending other users using generated links in the comments would be an effective way to raise the interest of the original user, because the recommended users are detected by their preferences, which are extracted from their comments on threads.

This mechanism generates links in the comments. To generate links in the comments after users have posted, we attach the links to user information or comments to related words by extracting user characteristics (feature words).

In the second method, the interface provides recommendations for Patterns 2 and 4, and the mechanism presents users on the top page of the website following 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, because this method focuses on recommending users.

IV. Active Communication Mechanism for e-Learning

In previous section, we discussed how to activate communication among users on wedding community websites. However, we expect that this method for recommending other users based on user characteristics in order to evoke communication will also be effective on other community websites, such as e-learning community sites. To enhance the user community on e-learning websites, we discuss an active communication mechanism for e-learning in this section.

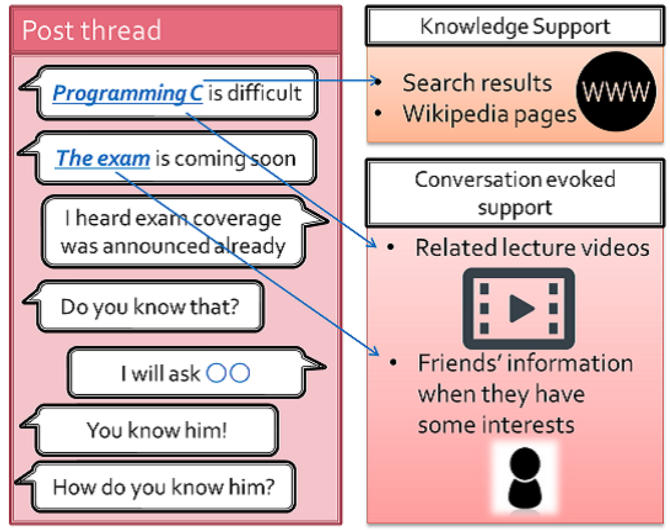

Figure 5. Conceptual diagram of automatic link generation.

A. System Overview for e-Learning

In recent years, massive open online courses (MOOC) have emerged as a new form of education for students who wish to attend courses at any level or cannot access traditional education because of constraints on time, location, or other factors. However, it is difficult to maintain students' motivation for self-learning. Currently, many students can collaborate during online courses through SNS, such as Facebook and Twitter. Liao et al. [25] and Chen et al. [26] investigated the use of SNS in online education. Students can communicate with each other when they use post threads on SNS. Because of the different levels of knowledge between students, communications cannot proceed smoothly. This is a problem that we aim to solve in this research. Therefore, it is necessary to extract user characteristics from SNS user behavior, which indicate users' knowledge and interests, and supplement users' posts with related information (e.g., search results, Wikipedia pages, lecture videos, or information on other friends) in the post thread.

In this work, the goal is to develop a novel automatic link generation method by analyzing SNS user behavior from user posts in e-learning. The proposed method generates two kinds of links: 1) knowledge support for receivers who receive posts from senders, and 2) conversation evoked support for receivers to give information to senders (see Figure 5).

Although several automatic link generation methods for websites have been studied [27][28], these have focused on web pages for knowledge support only, and they do not solve the abovementioned issues regarding user communication in e-learning. Other studies regarding user behavior on news websites [29] have not considered the relationships between users.

In this paper, we first extract user characteristics by analyzing SNS user behavior in e-learning from their posts. Then, we can detect users' relationships based on user characteristics. Thus, 1) links for knowledge support are attached to posts by using search results or Wikipedia pages for unknown vocabulary words for receivers, and 2) links for conversation evoked support for receivers to offer information regarding

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

senders are attached to posts by using related lecture videos for topics of posts or other information on other friends with the same interests. Thus, the proposed novel method encourages users to communicate with each other during conversations on e-learning platforms. This method activates communication not only by recommending other users, but also by generating links to relevant knowledge for users' studies.

In this section, we describe the proposed link generation method for active communications in e-learning. The proposed method extracts user characteristics by analyzing user behavior from their posts, selects vocabulary words as links based on user characteristics from their posts, and selects vocabulary words as links based on user characteristics. We first extract feature words for each user from their posts as user characteristics. Then, low or high weight vocabulary words are selected as link candidates based on users' relationships. The link information is classified into two categories: 1) knowledge support and 2) conversation evoked support.

B. Analysis of SNS User Behavior

To analyze SNS user behavior, we first extract high-frequency words from each post of each user using the Yahoo! Web API[1]. Next, we calculate the average weight of each extracted high-frequency word. Then, we extract user characteristics by extracting feature words for each user. We calculate the weight of each word i that appears in each user's posts using the following formula:

$$\frac{\text{weight of } i \text{ by Yahoo!Web API}}{\#\text{posts with } i} \times \frac{\text{total } \#\text{posts}}{\#\text{posts with } i} \tag{4}$$

The left part of Eq. (4) calculates the average weight of i that appears in each post. The right part of Eq. (4) gives an *IDF* value of i in all posts of each user. In addition, “Like” and “Share” options are available on SNS, to respectively mark interests or spread a post. Therefore, we can improve the calculation method by adding the numbers of “Like” and “Share” actions to the weight of each word.

C. Generation of Links in Chats

To generate links in chats between users, we attach link information to vocabulary words in posts based on users' relationships by using user characteristics (feature words). The link information for knowledge support is intended to supplement unknown vocabulary words for the receiver. We select the low weight words as unknown vocabulary words to be as link candidates in the post from the extracted feature words from the receiver's characteristics. Then, we attach search results and Wikipedia pages for the low weight words to the post. In Figure 5, it is determined that the receiver has a lack of knowledge regarding “C Programming,” and “C Programming” is generated as a link in the post to the receiver.

The link information for conversation evoked support is intended to promote user communication for receivers by offering information regarding the senders. We select the high weight words from extracted feature words of the receiver's characteristics, and detect information on other friends (linked with the receiver through the network) related to the post as link candidates using cosine similarity the high weight words through the formula given below. In Figure 5, it is determined that a friend is similar to the sender and mentions “the exam”, and “the exam” is generated as a link in the post to the receiver.

$$Sim(\overrightarrow{x}, \overrightarrow{y}) = \frac{\sum_{i=1}^{|V|} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{|V|}(x_i)^2} \cdot \sqrt{\sum_{i=1}^{|V|}(y_i)^2}}, \tag{5}$$

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

TABLE IV. NUMBERS OF COMMENTS.

#posts	#users
1 - 10	440 users
11- 20	64 users
21 - 30	37 users
31 - 40	14 users
41 - 50	1 user
51 - 60	1 user

V. EVALUATION

In this section, we first extract the actual data from a wedding community website, in order to verify the user characteristic extraction method by extracting feature words for all posts of each user. Next, we detect similar users by comparing the cosine similarity with collaborative filtering. The dataset consists of 6,361 posts by 588 users during six months. Table IV shows the distribution of the numbers of comments for each user.

A. Experiment 1: Verification of User Characteristic Extraction on a Wedding Community Website

To evaluate our user characteristic extraction method, we extracted feature words from all posts of each user. We

[1]http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.html

TABLE V. 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

compared the following three methods:

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 indicate 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 as feature words. We determined that calculating using *IDF* is a more effective way of extracting 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, then 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 those after marriage. Thus, we considered different definitions of document groups, which are not limited to the marital status.

Our results suggest that in the future we need to remove common words, because 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 on a Wedding Community Website

In our active communication mechanism, the similarities between users constitute the key aspect for recommending users. In the previous section, we described our classification scheme that classifies users based on similarities of three axes. In this manner, we 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, specifically the content-based recommendation method using the cosine similarity with active profile information. The second method is an item-based recommendation method that uses collaborative filtering with static profile information and marriage status. As mentioned previously, we calculated the cosine similarity based on user characteristics, which consist of feature words for 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 for each user. This is mainly used to recommend other items to users according to the following formula:

TABLE VI. 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

$$Sim(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}. \tag{6}$$

This equation calculates the similarity between users X and Y. On a wedding community website, users create individual accounts by answering questions regarding 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, 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 answer 1 and 5, the average value would be 3.

The users evaluated for our proposed user characteristic extraction method 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 V shows the distribution of results for 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 relating to their wedding planning. However, some user combinations induce a high cosine similarity.

Table VI 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

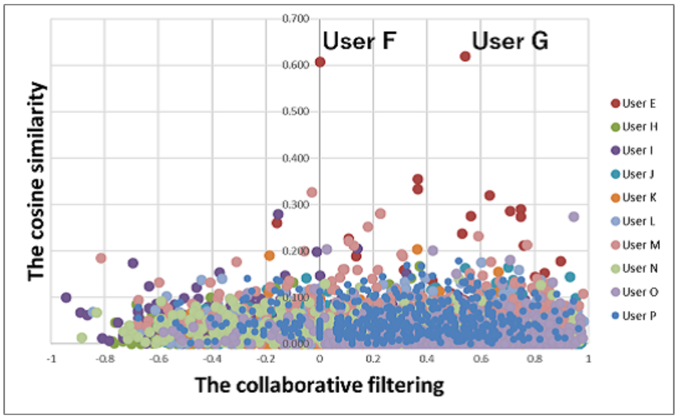

Figure 6. Distribution of the cosine similarity and collaborative filtering 1.

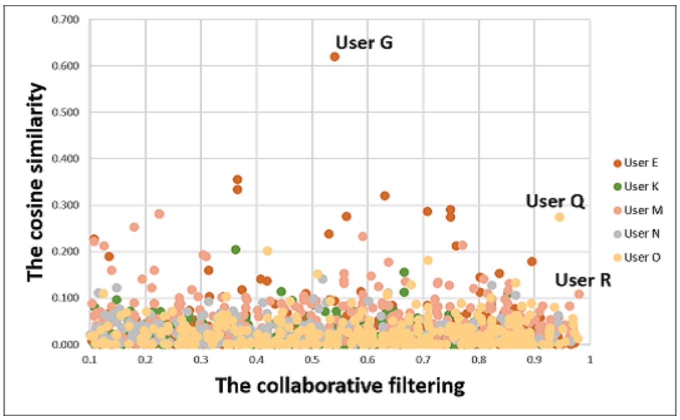

Figure 7. Distribution of the cosine similarity and collaborative filtering 2.

when users are creating accounts on the wedding community website. A high value implies that 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.

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 the wedding community site as a main user. We calculated all combinations with user E. Therefore, there were a total of 588 values for the cosine similarity and 588 values for collaborative filtering.

Figure 6 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 for user E, specifically F and G. Both of these users have high cosine similarity values of over 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 cousin's impressive wedding and the groom's tears. On the other hand, a post of user F describes how their cousin's 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 one as mentioned above. A post from user G describes their cousin's wedding, with tears resulting from a letter about a grandmother who has passed away. These posts both mention the same type of wedding and their cousin's 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 for detecting similar comments. However, we found that calculating both the cosine similarity and collaborative filtering is effective. Therefore, these two methods can help to detect similar user comments in order to evoke communications among users. However, we must still evaluate different situations for a user by considering other users' axes and marriage statuses.

Figure 7 shows the distribution of the cosine similarity and collaborative filtering for users E, K, M, N, and O. We found several users that are particularly similar to these users, such as users Q and R. In the future, we plan to propose methods for clustering using the cosine similarity and collaborative filtering.

C. Experiment 3: Verification of Recommendation Effect

We interviewed five users to evaluate effect of our proposed recommendation system. On experiments 1 and 2, we verified that our proposed methods work effectively to detect recommended users. For this experiment, we evaluated the similarities for comments and the communication desire through interviews.

We interviewed five users of the wedding community website. On this community website, majority of users want to communicate with others to share their experiments, and want to obtain information from other users. The purpose of using this community website for the five users is same as for the majority. Therefore, these five users' opinions should reflect the opinions of the majority on this website.

The interviewer evaluated the two questions below regarding the users ' own comments and those of recommended users.

1) Similarity of comments
2) Communication desire

As mentioned in Section III, this system is able to recommend five patterns of different users (see Table I). For this evaluation, we considered two kinds of comments: 1) comments of recommended users whose preferences are similar according to one of the patterns 1, 2, 4, and 5; and 2) recommended users' comments whose situation is similar, but whose preferences are different, corresponding to pattern 3.

1) Other users' comments
whose preferences are similar,
(corresponding to one of the patterns 1, 2, 4, 5)
2) Other users' comments
whose situation is similar,
but whose preferences are different,
(corresponding to pattern 3)

TABLE VII. VERIFICATION OF RECOMMENDATION EFFECTS.

Recommended user	Comments similarity (adv)	Communication desire (adv)
1. High posting similarity (users with similar preferences)	3.00	3.33
2. Low posting similarity, but high static similarity (different preferences and similar situation)	2.89	3.56

TABLE VIII. EFFECTIVE RECOMMENDED USER FOR ACTIVATING CONVERSATION.

	Similar Situation	Different Situation
Similar Preference	Good	Good
Different Preference	Very Good	—

Figure 8. An example of recommender system.

Table VII presents the results of the interview experiments. The five users answered using the numbers between 1 and 5. The number 1 indicates strong disagreement, and 5 means strong agreement. According to the results, we can see that recommending other users is effective for activating communication, as the users want to communicate with others.

However, we could determine whether recommending a user who has a similar situation and different preferences is more effective than recommending similar users. We assume that recommending a similar user would activate communication, but this turned out not to be true at all (see Table IV).

Through this experience, we created one example of our proposed recommendation system (see Figure 8). In this example, based on the original user, the recommended user's situation is similar, but their interests, which also correspond to preferences, are different. We could determine that their conversations will be activated through this experiment by interviews.

D. Experiment 4: Verification of User Characteristics on SNS

The purpose of this evaluation is to verify whether our proposed method is useful for extracting user characteristics based on SNS user behavior. We acquired posts for public online course pages using the Facebook API[2] as follows:

- A: the latest 50 posts of "Exciting Programming Starting from Elementary School"
- B: the latest 50 posts of "Online Programming Learning Service on APP Development"

A is an online course for programming beginners, and B is an online course for advanced learners in programming. In this evaluation, we extracted feature words as user characteristics of A and B by using the following four methods:

1) The weight of word i given by Eq. (1)
2) 1) $\times$ #likes of posts with i
3) 2) $\times$ #shares of posts with i
4) 3) $+$ 1) for each reply $\times$ #likes for each reply

[2]https://developers.facebook.com/

Here, #likes and #shares for A or B were normalized to fit within the range of 0 to 1. As described above, we proposed the following methods to test: 1) *TF-IDF*, which was calculated using only the text that the user has posted; 2) integrating the value of 1) with the number of likes, which was normalized as the weight; and 3) integrating the value of 2) with the number of shares, which was normalized as the weight. In addition, 4) we verified user characteristics (feature words) by adding the value of 3) to the value obtained by integrating 1) *TF-IDF* for each comment with the number of likes for each comment. Table IX shows the top 15 feature words for A and B obtained using each method. Bold words denote feature words that are related to A or B.

We found that many feature words are proper nouns, such as "Graduate School of Information Science and Technology, The University of Tokyo" and "jQuery UI." In all methods, the ranking orders of the same feature words are different, and several feature words are different. For example, the feature word "Kenichirou Mogi" of A ranked highest for method 1, and does not occur in the top 15 for methods 2-4. In this work, high weight feature words representing user characteristics are used to extract the relationships between users. Therefore, these feature words for generating feature vectors for each user are useful for both receivers and senders in conversations. According to the correlations between the method 1 and other methods found by comparing the rankings of the top 15 feature words based on Spearman's rank correlation coefficient, the correlation coefficient for methods 1 and 2 is 0.77, that for methods 1 and 3 is 0.76, and the value for 1 and 4 is 0.72. Although the correlations between method 1 and the other methods are similar, method 4, which considers #likes, #shares, and replies, is different from method 1. In addition, the correlations between method 1 and the other methods are not high, and we could confirm that the feature words obtained using these four methods are different.

As discussed above, many proper nouns occur that do not require knowledge assistance. Conversely, common words are widely used, and they are not useful for our proposed system. In the future, we need to improve the calculation method in order to remove common words when extracting user characteristics.

E. Future Work

We verified the validation of our proposed system by evaluation, however, we still need to consider several points of our proposed methods as future work. 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

TABLE IX. TOP 10 FEATURE WORDS OF A AND B.

Method	A	B
1)	**Kenichiro Mogi**, **Nikkei software**, **debate**, **Graduate School of Information Science and Technology**, **The University of Tokyo, Kuramoto Daishi**, innovation, self-expression, **industrial competitiveness conference**, **robot programming teaching materials**, **programming compulsory subject**, **Newsweek Japanese version**, **trilingual**, **study Roh**, **Mitsuru Sugaya**, account every single	**CSS3**, **EdTech JAPAN Pitch Festival vol.4**, **go to jappan**, **Higher or Lower**, **IE KMD Venture Day Tokyo**, **jQuery UI**, **Tech academy**, **u-note**, **parallax**, Engineering, good, **SF JAPAN NIGHT semi Finals team**, learning, **Now we're hiring a great web designer**, **SF JapanNight**
2)	**Hour of Code Japan**, **Graduate School of Information Science and Technology, The University of Tokyo**, **programming compulsory**, **scratch Di**, **Prof. Yoshiaki Hashimoto**, **PC away**, a few lines, **study Roh**, **Show&Tell**, **Touch& Try**, **Code.org**, **World Business satellite**, self-expression, **Minecraft EDU**, **robot programming**	**CSS3**, **jQuery UI**, **Thanks for Five Thousand Fans**, learning, u-note, **Higher of Lower**, feedback, **SF JAPAN NIGHT semifinalists decision**, **intern**, **we'll launch a radical web wervice which**, **Trello**, **Pyhonista**, **Now we're hiring a great web designer**, **SF Japan Night**, This new service has already decided
3)	**scratch Di**, **Graduate School of Information Science and Technology, The University of Tokyo**, **Show &Tell**, **Touch &Try**, **Prof. Yoshiaki Hashimoto**, **Hour of Code Japan**, **PC away**, **study Roh**, **robot programming**, **Code.org**, **World Business Satellite**, **programming a compulsory subject**, **Nikkei style**, the former, co-workers	**Yukihiro Matsumoto**, learning, **object-oriented scripting language**, **jQuery UI**, **server-side scripting language**, tab, **SF JAPAN NIGHT semi fainalists decision**, **Higher or Lower**, **the three-column layout**, already learned, inquiry, voice, learning situation, **Mats**, **CSS3**
4)	nowadays, **education**, **scratch Di**, **faculty side**, **Graduate School of Information Science and Technology, The University of Tokyo**, **Show &Tell**, **Touch &Try**, **compulsory**, **high school**, **Nikkei BP booth**, challenge, **Prof. Yoshiaki Hashimoto**, maximum, case, **Hour or Code Japan**	very, sue, **Yukihiro Matsumoto**, learning, **Koushou Kawasoe**, **object-oriented scripting language**, **jQuery UI**, **server-side language ban**, tab, **SF JAPAN NIGHT semifinalists decision**, Higher or Lower, **three-stage assenmbly layout**, already learned, inquiry, voice

by constructing a matrix based on user behavior, as in our previous work [30].

Our proposed method has only been applied to users of an existing wedding community website in Japan. It should also be considered for intercultural communications. Furthermore, we are planning to verify this mechanism on other community websites.

Moreover, we are planning to propose a more effective user interface that considers the security implications of sharing and combining the information. Some users may feel annoyed to be suggested to other users, and so this system requires additional options to protect privacy. This problem should be discussed more in future research.

VI. CONCLUSION

In this paper, we proposed an active communication mechanism for a wedding community website. This mechanism recommended users who may potentially evoke communications, as well as their comments. To detect users, this mechanism classified all users according to three axes, specifically "Static Profile Information," "Marriage Status," and "Dynamic Profile Information." We then calculated the similarities between users using the cosine similarity. To extract comments that were posted on a wedding community website by recommended users, our mechanism detected the most closely related comments. We also proposed an activation mechanism for e-learning. Finally, we evaluated the method for extracting user characteristics from posts by comparing *TF-IDF* methods, and evaluated the similarity calculation methods using the cosine similarity and collaborative filtering. Moreover, we verified the extraction of user characteristics using data from Facebook.

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grant Numbers 15K00162, 16H01722, and 17K12686.

REFERENCES

[1] T. Hayashi, Y. Wang, Y. Kawai, and K. Sumiya, "A Proposal of Activation Mechanism for User Communication based on User Behavior Analysis on Wedding Community Sites," in Proceeding of The Ninth International Conference on Advances in Multimedia (MMEDIA 2017). IARIA, APR 2017, pp. 67–72, ISBN: 978-1-61208-548-7.

[2] M. Rathakrishnan, R. Ahmad, and C. L. Suan, "Online discussion: Enhancing students' critical thinking skills," AIP Conference Proceedings, vol. 1891, no. 1, 2017, p. 020120.

[3] F. Abel, I. I. Bittencourt, E. Costa, N. Henze, D. Krause, and J. Vassileva, "Recommendations in online discussion forums for e-learning systems," IEEE transactions on learning technologies, vol. 3, no. 2, 2010, pp. 165–176.

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

[5] 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.

[6] L. H. Shaw and L. M. Gant, "In defense of the internet: The relationship between internet communication and depression, loneliness, self-esteem, and perceived social support," Cyberpsychology & behavior, vol. 5, no. 2, 2002, pp. 157–171.

[7] 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.

[8] J. Preece and D. Maloney-Krichmar, "Online communities: Design, theory, and practice," Journal of Computer-Mediated Communication, vol. 10, no. 4, 2005, pp. 00–00.

[9] E. Park, R. Rishika, R. Janakiraman, M. B. Houston, and B. Yoo, "Social dollars in online communities: The effect of product, user and network characteristics," Journal of Marketing, 2017.

[10] K. Randhawa, E. Josserand, J. Schweitzer, and D. Logue, "Knowledge collaboration between organizations and online communities: the role of open innovation intermediaries," Journal of Knowledge Management, no. just-accepted, 2017, pp. 00–00.

[11] P. Grace-Farfaglia, A. Dekkers, B. Sundararajan, L. Peters, and S.-H. Park, "Multinational web uses and gratifications: Measuring the social impact of online community participation across national boundaries," Electronic Commerce Research, vol. 6, no. 1, 2006, pp. 75–101.

[12] J. Preece, "Sociability and usability in online communities: Determining and measuring success," Behaviour & Information Technology, vol. 20, no. 5, 2001, pp. 347–356.

[13] P. J. Batterham and A. L. Calear, "Preferences for internet-based mental health interventions in an adult online sample: Findings from an online community survey," JMIR mental health, vol. 4, no. 2, 2017.

[14] C. Liang, C.-C. Chang, W. Rothwell, and K.-M. Shu, "Influences of organizational culture on knowledge sharing in an online virtual community: Interactive effects of trust, communication and leadership," in Organizational Culture and Behavior: Concepts, Methodologies, Tools, and Applications. IGI Global, 2017, pp. 972–992.

[15] 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.

[16] D. Godes and D. Mayzlin, "Using online conversations to study word-of-mouth communication," Marketing science, vol. 23, no. 4, 2004, pp. 545–560.

[17] J. M. Kayany and P. Yelsma, "Displacement effects of online media in the socio-technical contexts of households," Journal of Broadcasting & Electronic Media, vol. 44, no. 2, 2000, pp. 215–229.

[18] P. K. Chintagunta, S. Gopinath, and S. Venkataraman, "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets," Marketing Science, vol. 29, no. 5, 2010, pp. 944–957.

[19] K. Hemmings-Jarrett, J. Jarrett, and M. B. Blake, "Evaluation of user engagement on social media to leverage active and passive communication," in Cognitive Computing (ICCC), 2017 IEEE International Conference on. IEEE, 2017, pp. 132–135.

[20] 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.

[21] L. Li and T. Li, "News recommendation via hypergraph learning: encapsulation of user behavior and news content," in Proceedings of the sixth ACM international conference on Web search and data mining. ACM, 2013, pp. 305–314.

[22] K. Xu, Y. Cai, H. Min, X. Zheng, T. Wong et al., "Uis-lda: a user recommendation based on social connections and interests of users in uni-directional social networks," in Proceedings of the International Conference on Web Intelligence. ACM, 2017, pp. 260–265.

[23] 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

[24] 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.

[25] H.-C. C. Yi-Wen Liao, Yueh-Min Huang and S.-H. Huang, "Exploring the antecedents of collaborative learning performance over social networking sites in a ubiquitous learning context," Computers in Human Behavior, vol. 43, 2015, pp. 313–323.

[26] R. Cheng and S. K. Sharma, "Learning and self-disclosure behavior on social networking sites: the case of facebook users," European Journal of Information Systems, vol. 24, no. 1, 2015, pp. 93–106.

[27] 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, 2015, pp. 401–403.

[28] Y. Aoki, R. Koshijima, and M. Toyama, "Automatic determination of hyperlink destination in web index," Proc. of IDEAS 2015, 2015, pp. 206–207.

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

[30] 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.