

User Modeling Using User Preference and User Life Pattern Based on Personal Bio Data and SNS Data

Hyejin Song*, Kihoon Lee*, and Namme Moon*

Abstract

The purpose of this study was to collect and analyze personal bio data and social network services (SNS) data, derive user preference and user life pattern, and propose intuitive and precise user modeling. This study not only tried to conduct eye tracking experiments using various smart devices to be the ground of the recommendation system considering the attribute of smart devices, but also derived classification preference by analyzing eye tracking data of collected bio data and SNS data. In addition, this study intended to combine and analyze preference of the common classification of the two types of data, derive final preference by each smart device, and based on user life pattern extracted from final preference and collected bio data (amount of activity, sleep), draw the similarity between users using Pearson correlation coefficient. Through derivation of preference considering the attribute of smart devices, it could be found that users would be influenced by smart devices. With user modeling using user behavior pattern, eye tracking, and user preference, this study tried to contribute to the research on the recommendation system that should precisely reflect user tendency.

Keywords

Bio Data, Data Tracking, Life Pattern, Machine Learning, Social Behavior Analysis, User Modeling

1. Introduction

Owing to the popularization of smart devices and social media, we can access resources such as social network services (SNS), TV programs, retail websites among others without the constraints of space and time. Thus, the amount of unstructured data (generated through social activity, weblog, etc.) that has enabled the interpretation of user tendency has drastically increased. Several types of research are actively being conducted to develop a recommendation system that can analyze unstructured data, learn user tendencies and interests, and provide customized services [1,2].

Most preceding studies on the recommendation system used single data such as personal SNS data, shopping history, weblog, and so on. However, it is difficult to accurately reflect user tendency with the existing recommendation system that uses single data. In recent years, various kinds of research on the recommendation system have been conducted that combine heterogeneous data to better interpret user tendency.

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Manuscript received October 31, 2018; first revision December 14, 2018; accepted January 22, 2019.

Corresponding Author: Namme Moon (namme.moon@gmail.com)

* Dept. of Computer and Engineering, Hoseo University, Asan, Korea ({song120408, happy51738}@naver.com, namme.moon@gmail.com)

This study attempts to collect, combine, and analyze personal SNS data and eye-tracking data to derive intuitive and precise user preferences. In addition, this study aims to collect and analyze biometric data, extract user life pattern, and propose user modeling using Pearson correlation coefficient based on extracted life pattern and user preference, ultimately contributing to the development of the recommendation system.

As a result of eye-tracking experiments on smart TV and smartphone, the necessity of a recommendation system considering the attributes of smart devices was confirmed. In addition, to supplement the shortcomings of the existing recommendation system, this study attempts to combine and analyze biometric data and SNS data, derive preference, calculate the similarity between users using similarity detection algorithm based on extracted preference and life pattern, and propose user modeling using similar user data.

2. Related Research

2.1 Eye-Tracking

The eye-tracking technology has evolved to popularize eye-tracking equipment and smart devices. However, in recent years, a research was conducted on customized services using eye-tracking technology and equipment that analyzed the user's gaze and obtained the area of interest (AOI) [3,4].

In the previous research [5], the eye-tracker was used to analyze the flow of the user's gaze during web surfing. This was combined with the user's weblog to acquire more accurate and objective interests of the particular user. A research was also conducted to improve the quality of decision by taking feedback from the users who used complex decision-making systems in various fields of society based on eye-tracking [5].

In this paper, the evaluation score of the product is derived by combining the gaze tracking data such as the fixed time and the average fixed time of the product with the rating of the product. Based on this score, the problem of classification-based recommendation system is solved by improving the products with relatively lower interests. In this paper, we collected information about the user's gaze using the webcam and conducted an experiment by defining an appropriate AOI for the gaze tracking stimulus to interpret the user's interest. We also calculated the fixed time and the average fixed time for the AOI by comparing the fixed time of the user's gaze, the line of sight, and the coordinates of the AOI to derive the preferences for the gaze tracking stimulus.

The preferences for the gaze tracking stimulus was obtained by calculating the fixed time and the average fixed time for the AOI by comparing the gaze fixation time, the gaze flow, and the AOI.

2.2 Social Behavior Analysis

Social network behavior analysis is a core technology in providing information and services that reflect customer needs in all areas of society [6]. A research on the recommendation system that interprets and reflects the preference of the public through SNS analysis is actively underway.

In previous research we proposed an analytical method by collecting user input data, and Twitter data such as classification and product selection, writing reviews, and using collaborative filtering [7]. Based on the familiarity with Facebook news feeds and timelines, we extracted common items of interest and then proceeded with measuring the similarities. It is possible to effectively provide content-based information on a Facebook page by repeating the experiments continuously, filtering only the list of top-

level groups generated by the AOI. Moreover, research has been conducted to group Facebook users according to their interests and recommend content to these grouped users in order of importance [8].

We grouped Facebook users according to their interests as well and recommended content to the grouped users in order of importance.

In this paper, we used the web crawling technology to collect personal SNS data of users participating in the experiment. The analysis of the user's behavior determined their interests and tendency toward the classification. The users were grouped on the basis of their derived classifications and preferences. Recommendations were provided based on similar user preferences and information.

2.3 User Life Pattern Analysis

With the spread of smart devices and wearable devices, the amount of biometric data that can be extracted to users is increasing exponentially [8]. There is also a growing interest in technologies for collecting, filtering, and analyzing these growing bio-data.

Research on algorithms for analyzing the increase in bio data has been progressing steadily. In recent years, research has also been actively conducted on systems that recommend products or services through data analysis in the entire social market. Various studies are underway to improve the accuracy of the recommendation system [9,10]. The most important keyword among various studies is user's tendency. Analyzing your life pattern is one of the most important tasks when identifying these user needs [11].

In this paper, we collect and analyze various biometric data of users and extract user life patterns.

We also tried to increase the reliability of the experiment using eye-tracking by measuring the time of experiment by taking the extracted user life pattern into account.

3. System Overview

Fig. 1 shows an overview of the process of user modeling proposed in this study. Based on preceding research, personal SNS data was collected using web scraping technology targeting the users in their 20s who actively used SNS [9]. Eye-tracking data was collected through the webcam on smart TVs and smartphones. In the case of a smartwatch, the user's activity and sleep data were collected on smartphones. For the preference calculation in the preprocessing of data, AOI coordinates and social classification data set in this study were compared with the existing data and the data appropriate for deriving preference conditions was extracted and stored. In the data analysis process, user behavior was analyzed to derive preference, followed by combining and analyzing the preference data of the common classification of the two types of data. This study attempted to derive the final preference for the classification, measure the similarity between users based on their biometric data (amount of activity and sleep), and thereby become the ground for the recommendation system.

4. Preference Based User Modeling Experiment

4.1 Collecting Eye-Tracking Data

The flow of the user's gaze is unconscious body information and its result will be different depending on the device or the type of content being watched. To verify this, we conducted eye-tracking experiments

using smart TVs and smartphones and confirmed the necessity of providing recommended services considering the characteristics of smart devices.

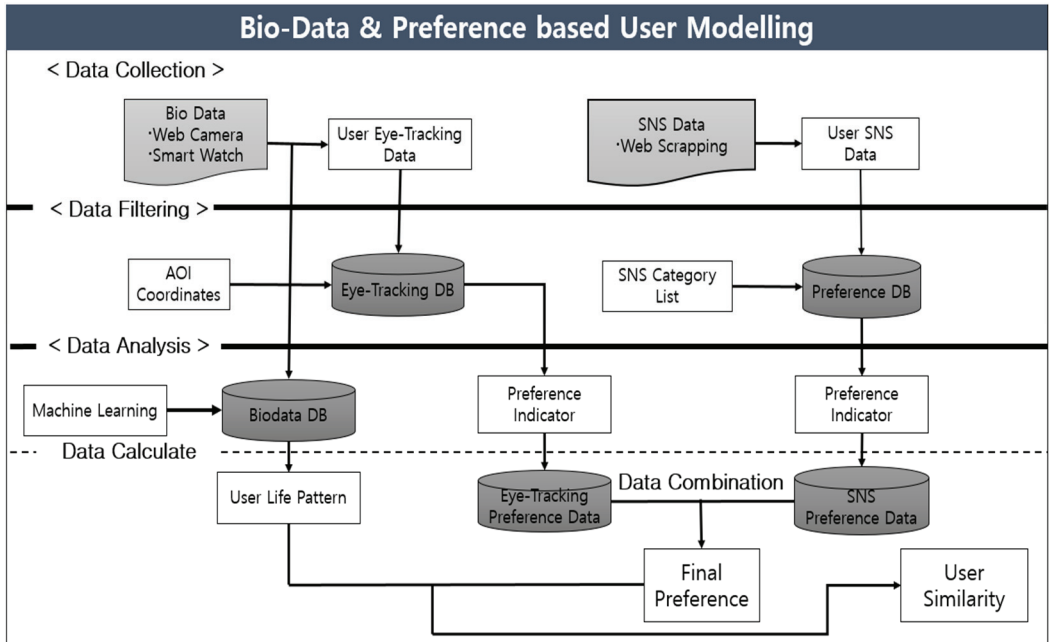


Fig. 1. System overview.



| | Contents | Time of appearance (sec) | Coordinate |
|---|--------------|--------------------------|---|
| 1 | Product Name | 0.89 | (593,633),(781,633),(593,697),(781,697) |
| 2 | Ad Copy | 1.33 | (113,249),(621,249),(113,397),(621,397) |
| 3 | Ad Character | 5.02 | (565,73),(829,73),(565,761),(829,761) |
| 4 | Product1 | 7.78 | (473,221),(1077,221),(473,521),(1077,521) |
| 5 | Product2 | 8.14 | (165,121),(573,121),(165,413),(573,413) |
| 6 | Ad Model | 13.24 | (255,48),(406,48),(255,525),(406,525) |
| 7 | Company Name | 13.24 | (627,232),(697,232),(627,329),(697,329) |

Fig. 2. AOI setting example.

Based on single webcam-based remote gaze tracking method, the gaze tracking data of smart TVs and smartphones was collected. In this paper, based on the previous research [3], the purchase records of high sales (houseware, household electronics, pharmaceutical) and smartphone sales (game, travel, fashion) were considered as experimental irritants. All six stimuli were shown in a 15-second video advertisement, at 30 fps, with 1280×720 magnification, and similar product exposure times. The AOI is composed of a rectangle model as shown in Fig. 2, starting from a point of interest mainly in products, product names, and advertisement models that are advertised through eye-tracking of users other than the experimental group.

The user’s line of sight has X, Y coordinates and measures the flow of gaze in the same way as shown in Fig. 3, in 0.1 second increments.

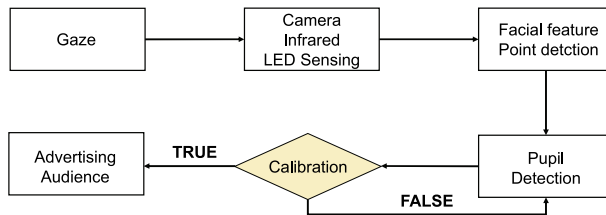


Fig. 3. Eye-tracking experiment course.

The coordinate values of the user and that of the AOI were compared with each other to determine whether the user has gazed at an AOI, thereby deriving the gaze tracking preference. The following Eq. (1) is used for measuring the user’s gaze. $Ugaze_i$ is a measure of gaze concentration the user looked in the advertisement. At_i refers to the total time of the advertisement. Au_i refers to the time taken by the user watched advertisement. Aot_i refers to the time taken for the AOI from Adu_i .

$$Ugaze_i = (Au_i / At_i \times 100) \times Aot_i / Au_i \tag{1}$$

From Eq. (2) below, the eye-tracking preference was derived, using $d Ugaze_i$.

$$UP_i = Ugaze_i / AUgaze_i \times 100 \tag{2}$$

UP_i denotes the user’s preference for classification, and $Ugaze_i$ is the intensity of the user’s gaze at the advertisement. $Ugaze_i$ denotes the average intensity of gaze at a particular classification and is calculated by the aggregation of users’ intensities of gaze at one classification divided by the number of users.

The derived $Ugaze_i$ verified different preferences according to smart devices and proved the need for providing recommended services.

4.2 Social Network Service

Several users utilize the search engine of SNS to follow specific pages and people or to receive news on the topics they are interested in. This behavior is passive therefore it can be implied that it represents the interests of the users. In this paper, we used web scraping technology to collect a list of ‘Liked’ pages on Facebook to derive preferences. The classifications provided by Facebook are composed of 12 upper classifications and 277 lower classifications, as shown in Table 1.

Table 1. Facebook classification

| Classification name | Number of sub-classification |
|------------------------|------------------------------|
| Web site & blog | 19 |
| Sports | 12 |
| Event source | 69 |
| Company & organization | 37 |
| People | 33 |
| Others | 9 |
| Business | 48 |
| Book & magazine | 9 |
| TV | 10 |
| Music | 13 |
| Movie | 8 |
| Brand & product | 36 |

The basic classifications provided by Facebook are commercially available and lack preference. In this This paper, we removed the classifications that are not relevant to the users in their 20s to fit the experimental group and compared it with the classification data provided via other social services. Thus, we derived the classifications, as shown in Table 2.

Table 2. Reclassified Facebook classification (Scategory)

| Classification name | Number of sub-classification |
|------------------------|------------------------------|
| Food | 8 |
| Fashion | 10 |
| Travel & culture | 8 |
| Company & organization | 17 |
| Game | 6 |
| Life & health | 10 |
| TV program | 8 |
| Sports | 21 |
| Distal appliances | 13 |
| Life & health | 10 |
| Music | 11 |

This is called *Scategory*. Its frequency is calculated based on the classification name. It is used to derive SNS preference. The data that is not included in *Scategory* is deleted.

Before measuring the user's SNS preference, the frequency of the user's 'Liked' pages is derived using the classification process shown in Fig. 4. This is based on the *Scategory*. The frequency UF_i is calculated by counting the number of words based on the classification name. It is called SNS preference measurement.

Social preference (SPr_i) refers to a user's preference for social behavior for a specific classification. AUF_i is the average frequency of users for a particular classification. The derived SPr_i confirms that the interests of each user are different. Eq. (3) is used in this study to measure the preferences of the user's social behavior.

$$SPr_i = \frac{UF_i}{AUF_i} \times 100 \quad (3)$$

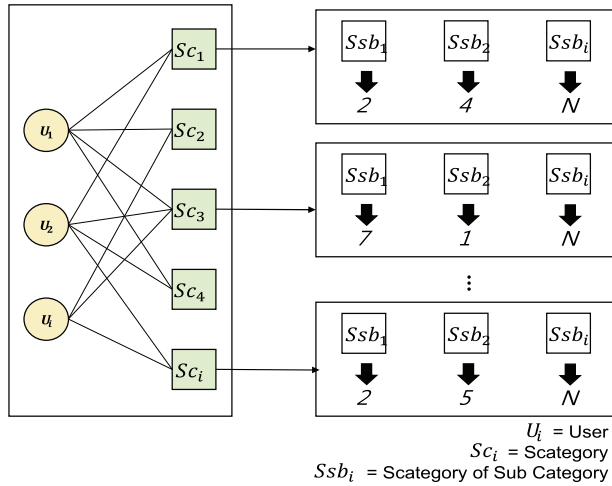


Fig. 4. Social behavior data classification process.

4.3 User Life Pattern

The sleep data (total amount of sleep and amount of non-REM sleep) of the users and their activity data (total travel) for a day were collected using the 3-axis acceleration sensor and heart rate sensor of smartwatch by connecting to smartphones. Twenty students of the same class of H. University, who had similar jobs, were selected for the experiment. As they were in the same class, they had similar attributes such as school timing and weather. These attributes were suitable for the experiment because the activity time for each participant was the same.

In this paper, we tried to increase the reliability by performing the two experiments in the same environment at the time of relatively higher concentration by dividing the life patterns of individuals into two groups, morning and evening, because the number of subjects was not large.

Table 3. Collected bio data samples

| Date | User ID | Sleep time (hr) | Deep sleep time (hr) | Activity (km) |
|------------|---------|-----------------|----------------------|---------------|
| 2018/01/12 | User 1 | 7.12 | 1.49 | 7.47 |
| 2018/01/12 | User 2 | 5.13 | 1.30 | 8.13 |
| 2018/01/12 | User 3 | 9.10 | 0.51 | 5.42 |
| 2018/01/12 | User 4 | 6.52 | 2.10 | 9.12 |
| ... | ... | ... | ... | ... |

Collected data is shown in Table 3, which was stored by time and user. User life pattern was derived using clustering based on sleep data and activity data.

4.4 Derivation of Preference and Calculation of User Similarity

To derive precise preference through a combination between biometric data and SNS data, eye-tracking classification and SNS classification were combined and analyzed to derive final preference. In this study, to determine preference combination, similar types of classification such as trip/culture (game, fashion, travel), life/health (medicine manufacture), houseware (kitchenware, decorative items), and household

electronic equipment (digital/home appliance) were used to derive final preference based on Eq. (4).

$$FP_i = (UP_i \times 0.6) + (SP_i \times 0.4) \tag{4}$$

Final preference (FP_i) for each classification was derived by adding eye-tracking preference (UP_i) to social activity preference (SP_i). The weight for each preference was $x=0.6$ and $y=0.4$, respectively. The performance was the highest when the result of recommendation was predicted based on root mean square error (RMSE) and mean absolute error (MAE). In the experiment to derive the weights for Eq. (4), the optimal weight values were calculated using Table 4: $x=0.7$ and $y=0.5$, for different weights of x and y . As a result of the experiment, the prediction rate was higher when the weight of eye-tracking preference was more than that of SNS preference. These results indicate that the unconscious data collected through users could be helpful in improving the precision of the recommendation.

Table 4. Predictions based on preference rate

| | RMSE | MAE |
|---------------------|----------|----------|
| $x=0.5$ and $y=0.7$ | 4.193639 | 3.900385 |
| $x=0.6$ and $y=0.6$ | 4.264093 | 3.944725 |
| $x=0.7$ and $y=0.5$ | 4.153472 | 3.834875 |

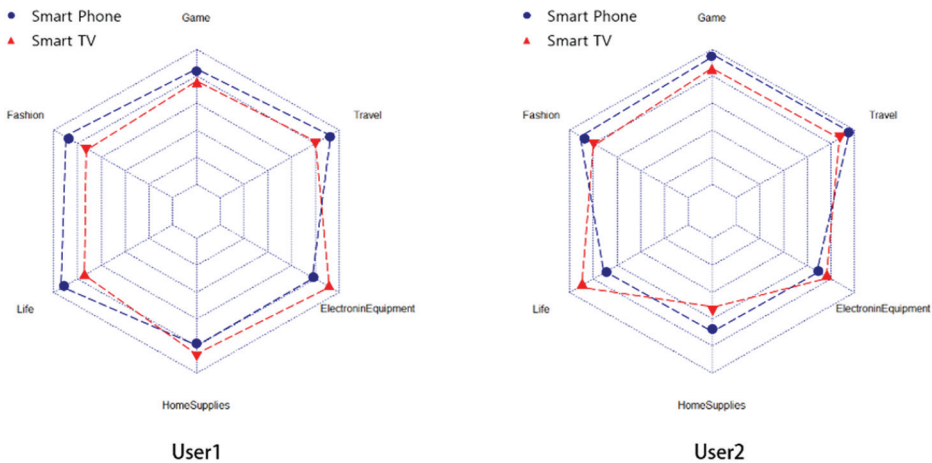


Fig. 5. User preference example.

Fig. 5 visualizes FP_i of users with similar types of life pattern, which shows the similarity between User1 and User2. User similarities were measured using Pearson correlation coefficients, which is appropriate for this study, with preference values for classifications that multiple users commonly rated. The Pearson correlation coefficient equation is demonstrated in Eq. (5).

$$\frac{n(\sum ab) - (\sum a)(\sum b)}{[n \sum a^2 - (\sum a)^2][n \sum b^2 - (\sum b)^2]} \tag{5}$$

Additionally, it was found that preference varies with smart TV and smartphone, which verifies the validity of the experiments conducted on the two devices used in this study. Based on FP_i and user life

pattern, the similarity between users on each smart device was examined using Pearson correlation coefficient. As a result, the recommendation based on smart TV showed 98.5% precision, and that based on smartphone showed 96.5% precision.

5. Experimental Environment and Method

An eye-tracking and social behavior preference-based recommendation system was implemented in an environment equipped with Intel Core i5-750 CPU @2.67 GHz and 16 GB RAM. For eye-tracking, the webcam was installed at a distance of 10–13 cm from the user's gaze. The analysis was done in Micro Visual Studio 2017 using Python 3.6 and R3.3.3. We collected data from 20 participants for a week. SNS data collection target was the same as the gaze tracking data collection target. In addition, data from 20 participants was collected during the first semester using smart watch to collect bio-data.

6. Conclusion

This study extracted intuitive and precise user preferences by combining eye-tracking data, SNS data, and heterogeneous data. It also determined the user life pattern through clustering. User modeling for the recommendation system was developed by utilizing user preferences and life pattern. The proposed user model was 98.5% accurate for smart TV and 96.5% accurate for smartphone. By deriving the preferences for the smart TV and smartphone, the influence of smart devices on their users could be verified. The collection of user biometric data and life pattern through clustering could lead to more precise user similarity. A follow-up study needs to focus on improving the precision of the recommendation system using the biometric data of more diverse users.

Acknowledgement

This research was supported by the Academic Research fund of Hoseo University in 2016 (20160286).

References

- [1] G. Bello-Orgaz, J. J. Jung, and D. Camacho, "Social big data: recent achievements and new challenges," *Information Fusion*, vol. 28, pp. 45-59, 2016.
- [2] S. Aghababaei and M. Makrehchi, "Activity-based Twitter sampling for content-based and user-centric prediction models," *Human-centric Computing and Information Sciences*, vol. 7, article no. 3, 2017.
- [3] J. R. Nurse and O. Buckley, "Behind the scenes: a cross-country study into third-party website referencing and the online advertising ecosystem," *Human-centric Computing and Information Sciences*, vol. 7, article no. 40, 2017.
- [4] W. Song, G. Sun, S. Fong, and K. Cho, "A real-time infrared LED detection method for input signal positioning of interactive media," *Journal of Convergence*, vol. 7, article no. 16071002, 2016.
- [5] A. Souri, S. Hosseinpour, and A. M. Rahmani, "Personality classification based on profiles of social networks' users and the five-factor model of personality," *Human-centric Computing and Information Sciences*, vol. 8, article no. 24, 2018.

- [6] L. Chen, F. Wang, and P. Pu, "Investigating users' eye movement behavior in critiquing-based recommender systems," *Ai Communications*, vol. 30, no. 3-4, pp. 207-222, 2017.
- [7] S. H. Park and J. Kim, "A method to utilize inner and outer SNS method for analyzing preferences," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 19, no. 12, pp. 2871-2877, 2015.
- [8] L. Chen and F. Wang, "An eye-tracking study: implication to implicit critiquing feedback elicitation in recommender systems," in *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, Halifax, Canada, 2016, pp. 163-167.
- [9] B. A. Galitsky, "Providing personalized recommendation for attending events based on individual interest profiles," *Artificial Intelligence Research*, vol. 5, no. 1, pp. 1-13, 2016.
- [10] S. I. Chiu and K. W. Hsu, "Efficiently processing skyline query on multi-instance data," *Journal of Information Processing Systems*, vol. 13, no. 5, pp. 1277-1298, 2017.
- [11] Korea Creative Content Agency, *2014 Korean Content Industry Statistics*. Naju: Korea Creative Content Agency, 2015.



Hyejin Song <https://orcid.org/0000-0003-0182-3669>

She received B.S. degrees in School of Division of Computer Information Engineering from Hoseo University in 2016 and her M.S. degrees in Department of Computer Engineering from Hoseo University in 2018. She is currently working at Chungnam Culture Technology Industry Agency since 2018. Her current research interests include Machine Learning, Big Data Processing and Analysis.



Kihoon Lee <https://orcid.org/0000-0001-7445-832X>

He received B.S. degree in School of Division of Computer Information Engineering from Hoseo University in 2018 and he is studying towards a M.S. degree in the Department of Computer Engineering from Hoseo University. His current research interests include machine learning, big data processing and analysis.



Nammee Moon <https://orcid.org/0000-0003-2229-4217>

She received B.S., M.S., and Ph.D. degrees in School of Computer Science and Engineering from Ewha Womans University in 1985, 1987, and 1998, respectively. She served as an assistant professor at Ewha Womans University from 1999 to 2003. From 2003 to 2008, she is a professor of Department of Digital Media, Graduate School of Seoul Venture Information. Since 2008, she is currently a professor of Division of Computer Information Engineering at Hoseo University. Her current research interests include social learning, HCI and user centric data, big data processing, and analysis.