J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr

A Social Travel Recommendation System using Item-based collaborative filtering

Dae-ho Kim[1] Je-in Song[1] So-yeop Yoo[1] Ok-ran Jeong[1*]

ABSTRACT

As SNS(Social Network Service) becomes a part of our life, new information can be derived through various information provided by SNS. Through the public timeline analysis of SNS, we can extract the latest tour trends for the public and the intimacy through the social relationship analysis in the SNS. The extracted intimacy can also be used to make the personalized recommendation by adding the weights to friends with high intimacy. We apply SNS elements such as analyzed latest trends and intimacy to item-based collaborative filtering techniques to achieve better accuracy and satisfaction than existing travel recommendation services in a new way. In this paper, we propose a social travel recommendation system using item - based collaborative filtering.

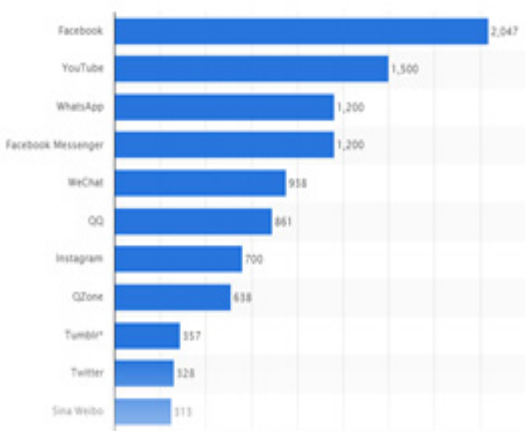
☞ keyword : Personalized Recommendation, Item-based Collaborative Filtering, Apache Mahout, Social Travel Trends, Intimacy

1. Introduction

In recent years, SNS (Social Network Service) such as Facebook, Twitter, Instagram is deeply embedded in our life and research is actively going on. In addition, as the wireless Internet becomes popular due to the recent spread of smartphones, it is possible to access the Internet from anywhere, and SNS can be utilized very much because it can access the SNS in real time through the GPS (Global Positioning System).

According to Statista's statistics [1], The number of users of social network services is enormous. Figure 1 shows statistical data on the actual number of users for each SNS. Facebook has the largest number of users, as of August 2017, the total number of users on Facebook is 16 billion and the number of active users is 2,047 million. Based on such a large number of users, the amount of data such as text and video accumulated in SNS per day is enormous. Based on the above information, you can get a lot of new information about the latest trends and people's candid opinions if you apply it to the existing research using the large amount of data that the SNS users and their large amount of data.

SNS can create more diverse and new social relations than offline social relations. By analyzing social relations, it is possible to extract places and items that the user likes, and the intimacy between users can be extracted through action data such as comments or like created by the user. Applying the intimacy on the SNS to the existing recommendation system can be improve the accuracy of recommendation.



(Figure 1) Active users of social network

1 Dept. of Software, Gachon University, Seongnam, 13120, Korea
*Corresponding author (orjeong@gachon.ac.kr)
[Received 31 October 2017, Reviewed 6 November 2017(R2 19 March 2018), Accepted 11 April 2018]

☆ This research was supported by Basic Science Research Program through the NRF(National Research Foundation of Korea), and the MSIT(Ministry of Science, ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute for Information & communications Technology Promotion) and Next-Generation Information Computing Development Program (Nos. NRF-2015R1C1A2A01051729, 2015-0-00932-003, 2017M3C4A7065980).

Existing personalized recommendation systems use a method of recommending items by analyzing similar users by using past history or ratings of users. In the case of Amazon, they provide a recommendation system using item-based collaborative filtering techniques, and the rate of purchasing recommended items accounts for a large percentage of total sales [2]. However, existing recommendation methods do not include basic information or social relations such as gender and age, and are not appropriately used even if included. When social network services are applied to existing personalization recommendations, various characteristics such as 'real-time', 'big data', 'social relations', and 'latest trend' can be applied.

A service that utilizes SNS as a personalized recommendation system is 'A Perfect storm for recommendation' created by Amazon and Facebook. People who purchase products from Amazon can get direct recommendation from users who have social connections on Facebook, and can also get recommendation through basic profile information on Facebook.

In this paper, we propose a personalized travel recommendation system to user by adding latest travel trend and intimacy in SNS to travel recommendation system and using Item-based collaborative filtering method.

2. Related Research

With the development of wireless Internet services and smart phones, it is now possible to access the Internet anywhere. In this environment, the usability of SNS is continuously increasing, and the size of the generated data has also increased. Therefore, studies [3-10] that analyze the data generated in the SNS are active. That kind of research focused on the personalized features such as user profile or preference. We can implement the personalized recommendation system by using these SNS features.

One of the various researches on the ongoing SNS is the study of extracting the interest or the intimacy through the social relation analysis in the user's SNS. The interest and intimacy extracted from these studies are used to improve the quality of the recommendation results, and the recommendation accuracy of the existing recommendation system is further improved by utilizing the SNS data.

[11] uses the social network to calculate the intimacy through the user's relationship analysis and applies it to the personalized recommendation system. However, if we calculate the intimacy through only the user's relationship analysis in the intimacy algorithm part, it is impossible to overlook the important interaction element in the social network and deeply analyze the relationship. When analyzing the social relations on the social network, it is possible to derive a deeper and more reasonable intimacy by analyzing the interaction between users such as likes, tag, and shares, as well as relationship analysis.

[12] proposes a travel recommendation system using data mining based collaborative filtering. In this paper, user-based collaborative filtering technique is used for recommendation of travel, and Pearson coefficient is used for similarity measurement. However, in the experimental results, the recommendation accuracy is 0.5 or less on the average, and the recommendation accuracy is low.

[13] proposes a personalized system that improves recommendation accuracy by using user's personal information in order to solve the sparseness problem which is a disadvantage of collaborative filtering. The user 's personal information uses personal profile information on the social network to suggest personalized recommendations based on the user' s tendency and recent activity information. However, if the user's profile has not changed recently, or if there is no activity information, there is a limitation that it is not much different from the existing system that analyzes and recommends the static profile.

[14] proposes a travel recommendation system. In this paper, they use CBR(Case-based Reasoning), which is one of the data mining techniques, by using weather and user information for travel recommendation. However, this method has the problem that it is very difficult to find the outlier when there is an exceptional case, and it shows that the recommendation performance can be remarkably lowered when the exception cases accumulate.

In this paper, we propose a recommendation system based on the latest travel trends and the intimacy of the user through social relationship analysis. Especially, in the step of extracting intimacy, we calculate how much interaction has occurred with the user in the social network. And we give different weights according to each type of interaction, and

calculate the intimacy by summing the interaction results. If we use the analyzed intimacy in the personalization recommendation system, we recommend more reasonable and satisfactory recommendation than the recommendation system using the existing relationship analysis This method is distinguished from the existing research in that it can be used not only in the travel recommendation system but also in other recommendation systems.

3. A Social Travel Recommendation System

The travel recommendation system proposed in this paper enables travel recommendation customized to the user.

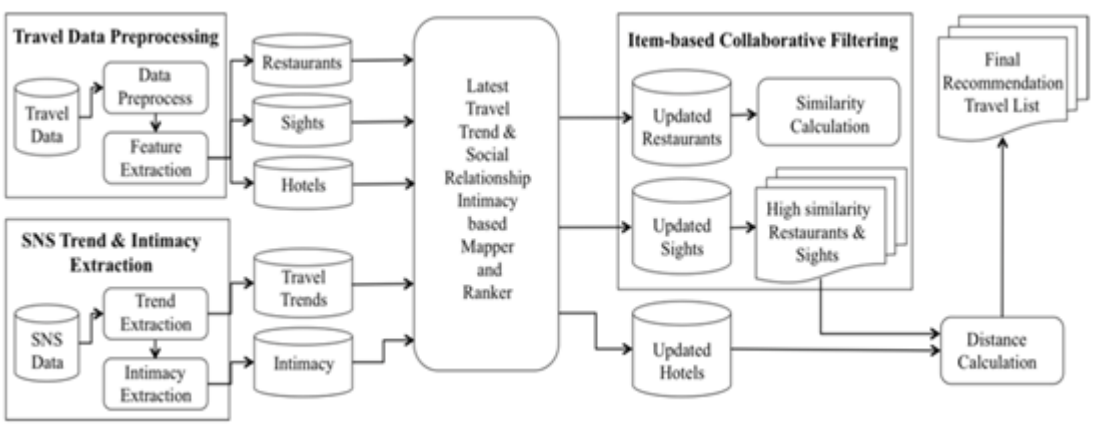
In the case of the existing travel recommendation system, it generally recommends the place most visited by people. However, these recommendation results do not reflect information such as age, gender and interests of the people who are going to travel, so they show the limit of recommendation results with low satisfaction. Because each person's taste is different, the recommendation system must be able to provide an optimized travel information that reflects their interest and preference to user.

Figure 2 shows the system architecture of the proposed system. In the Travel Data Processing process, in order to collect data about travel sights, data of restaurants, hotels, and sights are crawled. And we did the morphological analysis to extract the information such as restaurant name or address from crawled documents. And in the Trend & User relationship Extraction process, we use SNS API to extract the latest travel trends and intimacy and apply them to crawled travel sights data. Each travel sights data comes with new data with the latest travel trends and familiarity. In the final recommendation step, we use Apache Mahout to calculate the similarity between restaurants and sights, and calculate the distance between hotel and <restaurant,sight> set with high similarities and extract the hotel with the closest distance. Finally the user is recommended to <restaurant,sight,hotel> set.

In this paper, we propose a personalized travel recommendation system through analysis of the latest travel trends and social relations in SNS.

3.1 Data Crawling and Preprocessing

In order to recommend personalized travel sights to users, it is necessary to analyze recent trends and individual users as well as travel sights information such as restaurants, hotels and sights. In order to obtain such data, the restaurant data is crawled by 'MenuPan.com' [15], the hotel data by 'hotels.com' [16], and the sights data by the Korea Tourism Organization's TourAPI [17]. To analyze the latest travel trends, we crawl the latest social media data and extract



(Figure 2) System Architecture

nouns using Korean morphological analyzer KOMORAN [18]. The extracted nouns are used as weights for previously crawled travel sights data reflected in the travel ranking, and stored in the database by category. Finally, social interaction data is crawled through API [19] provided by SNS to extract intimacy.

3.2 Extract Latest Travel Trends

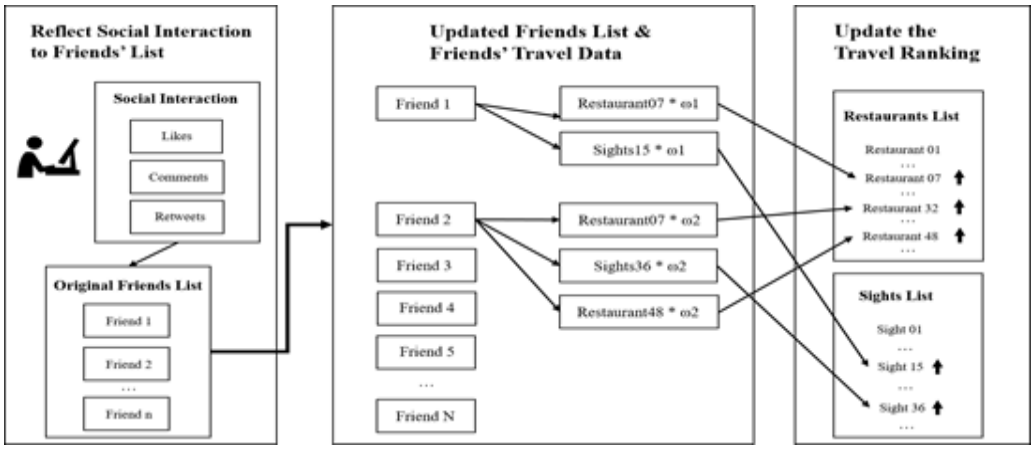
In the case of the latest travel trend information, only articles related to travel are extracted from the SNS timeline article. In the first step, crawls SNS public timeline data through the streaming API provided by the SNS. To extract travel-related articles from the crawled timeline data, check-in data included in the article is used. Finally, the extracted latest travel trend data is used as a weight in the previously stored travel list and reflected in the travel ranking.

3.3 Extract Social Relationship and Intimacy

On a social network, users are generally more likely to have a social relationship with people who are close friends or tend to have the same tendency or belong to the same group. The social relationship in the social network can be used as a good analytical indicator for the user's interest or preference. Therefore, by analyzing social relations in social networks and recommending a higher weight to the travel destinations visited by highly-interested users, it is possible to recommend more reliable personalized travel destinations than general travel recommendation services. Figure 3 shows the method of extraction and application of intimacy in this paper. The process of extracting and applying intimacy is largely divided into three steps. First, we extract elements that can represent intimacy in SNS. In the SNS, the elements of intimacy are defined as how much interaction with the user is in the SNS, and we have assumed that the more interaction, the higher the intimacy. We used intimacy elements like 'Likes', 'Retweet', and 'Comments' to get interaction. In the second step, the existing friend list is ranked and updated in order of extracted high intimacy. Finally, the ranking of the existing travel list is updated by assigning a higher weight to the travel sight visited by friends having high intimacy.

$$R'_k = T_k I_k R_k \quad (1)$$

The following equation 1 is an example of the formula for how we apply the weights obtained from analysis results of latest trends and intimacy to the original ranking value. Rk is the original ranking value when crawling on the restaurant site, and Tk is the weight of how many the people have uploaded that Rk to the public timeline. Ik is the weight of how many people who have familiarity with user visited Rk. Because the purpose of the proposed system in this paper is a personalized



(Figure 3) Social relationship and Intimacy extraction in SocialNetwork Service

travel recommendation system, weights of intimacy were set to a little higher than the latest trend weights.

3.4. Travel Recommendation Using Item-based Collaborative filtering

Travel data that reflects the latest trends and intimacy extracted through the above process is recommended to the user through item-based collaboration filtering [20].

In this paper, we propose an item-based recommendation system using Mahout which is Apache machine running open source [21]. Existing collaborative filtering methods use user's rating data and similarity measures are also used, such as Pearson correlation coefficient [22] or Cosine similarity [23]. On the other hand, because the system does not have rating data in the travel data of the proposed system, we use log-likelihood similarity [24].

In this paper, we calculate the similarity between restaurants and sights by using Mahout, then sort them in descending order of similarity such as <restaurant,sight>. After that, we calculate the distance between the above results and the hotel data. As a result, users can get recommended travel sights with the latest travel trends and intimacy. In addition, personalized and travel recommendations can be made compared to existing recommendation services.

4. Experiment

4.1 Experiment Setup

The experiment in this paper is based on the social media data of Twitter crawled from January to June 2016. We used Twitter's public timeline data to extract the latest travel trends, and we used the social interaction data for each user on Twitter to extract the intimacy. I also crawled friend's twitter timeline data related to the travel that 100 friends had visited which includes about 600 travel data. The extracted latest travel trend, intimacy, and friend's travel sights data are reflected in the previously crawled travel sights ranking. Only the top 50 data of each travel sights such as restaurant, hotel, sights, was extracted and stored in the database.

4.2 Experimental Comparison with similar theme

Before evaluating the recommendation accuracy of this paper, we would like to compare it with other papers with similar themes. The reason for comparing with other papers is that first, the recommendation result proposed in this paper has high accuracy and it is evaluated whether the recommendation result is reasonable, The second is to compare how much superior it is to existing papers with similar themes. The first comparative paper [17], proposes a personalization recommendation algorithm that takes advantage of the features of social network services and considers user intimacy. It is similar to this paper in that it considers social network services and social relations. The experiment measures are MAP (Mean Average Precision), MRR (Mean Reciprocal Rank), and Normalized Discounted Cumulative Gain (NDCG). The results are 0.9067, 0.9195, and 0.9451, respectively.

The second comparative paper is [25].

In this paper, traveler 's stay time is considered and travel schedule is recommended using semantic movement trajectory and semantic category set. In this paper, the Recall was measured as an experiment measure and the results were compared with those without consideration of the stay time. The results of the recommendation results considering the stay time were higher than those not considered, but there was no significant difference.

4.3 Experiment Method

In this paper, we used MAE (Mean Absolute Error) which is widely used as the recommendation system evaluation. We compared the test results with the training results. The MAE can be calculated using the following formula.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (2)$$

MAE is a method of comparing the training value and the test value. It compares the difference between the result of the training set and test set by calculating the error between the two result values.

In this paper, we use the recommended results from Top 1 to Top 5 of training value and test value. The error of the

two results is obtained through the MAE, and we can see how good the recommendation accuracy of the proposed system is.

4.4 Experiment Results

In order to evaluate the performance of the system proposed in this paper, we used the travel data from January to June as a training set to derive recommended travel sights set. And this data are compared the results of the recommended travel sights set as a test set, collected in July with those actually visited by people to evaluate the recommendation accuracy. Table 1 shows the results of sample data from Top1 to Top5 in order of similarity from recommended travel sights using training sets and test sets, respectively. Restaurants are marked R1~R50, and sights are marked S1~S50. We can confirm the accuracy of 80% by calculating the difference between the recommendation results of the training data and test data respectively. In other words, the recommendation accuracy of the proposed system is very high and reliable. Figure 4 shows the same result as Table 1.

(Table 1) Experiment result

Training Data	Test Data	Error(E_i)
1(R21-S39)	2(R25-S40)	1-2=1
2(R25-S40)	1(R21-S39)	2-1=1
3(R13-S33)	3(R13-S33)	3-3=0
4(R45-S24)	6(R31-S18)	4-6=2
5(R45-S26)	5(R45-S26)	5-5=0

5. Conclusions

We propose a personalized travel recommendation system using the travel trends and the intimacy on the SNS. It is possible to analyze the current travel trends by considering the popular travel sights. Computing the intimacy enables for the user to obtain the recommended travel sights where other users in the same group or with the similar interests have visited. However, the limitation of this study is that as the permission for the SNS API has been strengthened recently, the data that can be analyzed from the SNS is greatly

(Figure 4) Graph of experiment result

decreasing. Therefore, we could recommend more reliable and satisfying travel than the existing recommendation system.

References

[1] https://www.statista.com

[2] Linden, Greg, Brent Smith, and Jeremy York. "Amazon. com recommendations: Item-to-item collaborative filtering." *IEEE Internet computing* 7.1, 76-80, 2003. https://doi.org/10.1109/MIC.2003.1167344

[3] Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M., "Recommendations in location-based social networks: a survey," GeoInformatica, 19(3), pp. 525-565, 2015. https://doi.org/10.1007/s10707-014-0220-8

[4] Filtering in SNS", Database Society, p53, KIISE, 2012.

[5] Bao, J., Zheng, Y., Mokbel, Mohamed F., "Location-based and Preference-Aware Recommendation using Sparse Geo-Social Networking Data," In Proc. of the 20th International Conference on Advances in Geographic Information Systems, ACM, pp. 199-208, 2012.

[6] Hu, B., Jamali, M., & Ester, M., "Learning the strength of the factors influencing user behavior in online social networks," In Proceedings of the 2012 *International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012),* IEEE Computer Society, pp. 368-375, August 2012.

[7] Durr, M., Protschky, V., & Linnhoff-Popien, C., "Modeling social network interaction graphs," In

Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), IEEE Computer Society, pp. 660-667, August 2012.

[8] Hu, B., Song, Z., & Ester, M., "User features and social networks for topic modeling in online social media," In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012),* IEEE Computer Society, pp. 202-209, August 2012. https://doi.org/10.1109/ASONAM.2012.43

[9] Naseri, S., Bahrehmand, A., Ding, C., & Chi, C. H., "Enhancing tag-based collaborative filtering via integrated social networking information," In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on* IEEE, pp. 761-765, August 2013. https://doi.org/10.1145/2492517.2492658

[10] WEI, Jian, et al. "Collaborative filtering and deep learning based recommendation system for cold start items," *Expert Systems with Applications*, 69: 29-39, 2017.

[11] Seo, YoungDuk, JeongDong Kim, and DooKwon Baik. "PReAmacy: A Personalized Recommendation Algorithm considering Contents and Intimacy between Users in Social Network Services," Journal of Computing Science and Engineering 179-181, 2013.

[12] Lee TaeRin, and Jeong YoonWon, "A scheme for recommending locations to tourist using collaborative filtering based on data mining," *The Korean Institute of Communications and Information Science,* 463-464, 2017.

[13] Kim BooSeong, et al. "Movie recommendation system using collaborative filtering based on Deomographic information" *Korea Intelligent Information System Society* 23.2, 63-64, 2013.

[14] Sim BeomSoo, Lee HanJoon, and Yoo DongHea. "Development of Customized Trip Navigation System Using Open Government Data" Korea Society for Internet Information, 17.1, 15-21, 2016. https://doi.org/10.7472/jksii.2016.17.1.15

[15] Menupan.com, http://www.menupan.com.

[16] Hotels.com, http://kr.hotels.com.

[17] Tour API 3.0, Korea Tourism Organization, http://kto.visitkorea.or.kr/kor/gov30/tourapi.kto.

[18] Choi, M., Shin, J., & Kim, H., "Robust Feature Extraction Method for Automatic Sentiment Classification of Erroneous Online Customer Reviews," *International Information Institute (Tokyo). Information*, *16*(10), 7637, 2013.

[19] TwitterAPI, https://dev.twitter.com/overview/api.

[20] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J., "Item-based collaborative filtering recommendation algorithms," In *Proceedings of the 10th international conference on World Wide Web,* ACM, pp. 285-295, April 2001.

[21] WALUNJ, Sachin Gulabrao; SADAFALE, Kishor, "An online recommendation system for e-commerce based on apache mahout framework," In *Proceedings of the 2013 annual conference on Computers and people research.* ACM, pp. 153-158, 2013. https://doi.org/10.1145/2487294.2487328

[22] Benesty, J., Chen, J., Huang, Y., & Cohen, I., "Pearson correlation coefficient," In *Noise reduction in speech processing,* Springer Berlin Heidelberg, pp. 1-4, 2009.

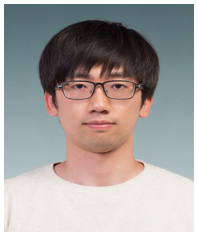
[23] Dehak, N., Dehak, R., Glass, J. R., Reynolds, D. A., & Kenny, P., "Cosine similarity scoring without score normalization techniques," In *Odyssey,* pp. 15, June 2010.http://groups.csail.mit.edu/sls/publications/2010/Dehak_Odyssey.pdf

[24] Pinheiro, J. C., & Bates, D. M. "Approximations to the log-likelihood function in the nonlinear mixed-effects model." *Journal of computational and Graphical Statistics, 4*(1), 12-35, 1995.

[25] Park SaeHwa, Park Seok, "Personalized Itinerary Recommendation System based on Stay Time," Journal of Computing Science and Engineering, 22(1), 38-43, 2016. https://doi.org/10.5626/KTCP.2016.22.1.38

◐ 저 자 소 개 ◑

김 대 호 (Dae-Ho Kim)
2017년 2월 가천대학교 소프트웨어학과 (공학사)
2017년~현재 가천대학교 일반대학원 IT융합공학과 소프트웨어전공 (석사과정)
관심분야 : 빅데이터, 소셜 미디어, 추천 시스템
E-mail : ikimdh91@gmail.com



송 제 인 (Je-In Song)
2016년 2월 가천대학교 소프트웨어학과 (공학사)
2018년 2월 가천대학교 일반대학원 IT융합공학과 소프트웨어전공 (공학석사)
2018년~현재 줌인터넷
관심분야 : 데이터마이닝, 머신러닝, 소셜 컴퓨팅
E-mail : jeinsong200@gmail.com

유 소 엽 (So-Yeop Yoo)
2014년 2월 가천대학교 소프트웨어설계·경영학과 (공학사)
2016년 2월 가천대학교 일반대학원 소프트웨어설계·경영학과 (공학석사)
2016년~2018년 가천대학교 소프트웨어중심대학사업단 (연구원)
2018년~현재 가천대학교 일반대학원 IT융합공학과 소프트웨어전공 (박사과정)
관심분야 : 소셜 컴퓨팅, 데이터마이닝, 머신러닝
E-mail : bbusso@gachon.ac.kr

정 옥 란 (Ok-Ran Jeong)
2005년 이화여자대학교 컴퓨터공학과 (공학박사)
2005년~2006년 서울대학교 컴퓨터공학부 (박사후 연구원)
2007년 Univ. of Illinois of Urbana Champaign (visiting scholar)
2008년~2009년 성균관대학교 정보통신공학부 (연구교수)
2009년~현재 가천대학교 소프트웨어학과 (부교수)
관심분야 : 웹 마이닝, 정보검색, 추천 시스템, 소셜 컴퓨팅
E-mail : orjeong@gachon.ac.kr