

Incorporating Social Relationship discovered from User's Behavior into Collaborative Filtering

Setha Thay

Department of Computer and
Information Engineering,
Inha University
(sethathay@eslab.inha.ac.kr)

Inay Ha

Department of Computer and
Information Engineering,
Inha University
(inay@eslab.inha.ac.kr)

Geun-Sik Jo

School of Computer and
Information Engineering,
Inha University
(gsjo@inha.ac.kr)

.....

Nowadays, social network is a huge communication platform for providing people to connect with one another and to bring users together to share common interests, experiences, and their daily activities. Users spend hours per day in maintaining personal information and interacting with other people via posting, commenting, messaging, games, social events, and applications. Due to the growth of user's distributed information in social network, there is a great potential to utilize the social data to enhance the quality of recommender system. There are some researches focusing on social network analysis that investigate how social network can be used in recommendation domain. Among these researches, we are interested in taking advantages of the interaction between a user and others in social network that can be determined and known as social relationship. Furthermore, mostly user's decisions before purchasing some products depend on suggestion of people who have either the same preferences or closer relationship. For this reason, we believe that user's relationship in social network can provide an effective way to increase the quality in prediction user's interests of recommender system. Therefore, social relationship between users encountered from social network is a common factor to improve the way of predicting user's preferences in the conventional approach.

Recommender system is dramatically increasing in popularity and currently being used by many e-commerce sites such as *Amazon.com*, *Last.fm*, *eBay.com*, etc. Collaborative filtering (CF) method is one of the essential and powerful techniques in recommender system for suggesting the appropriate items to user by learning user's preferences. CF method focuses on user data and generates automatic prediction about user's interests by gathering information from users who share similar background and preferences. Specifically, the intension of CF method is to find users who have similar preferences and to suggest target user items that were mostly preferred by those nearest neighbor users. There are two basic units that need to be considered by CF method, the user and the item. Each user needs to provide his rating value on items i.e. movies, products, books, etc to indicate their interests on those items. In addition, CF uses the user-rating matrix to find a group of users who have similar

* This research was supported by the MKE(The Ministry of Knowledge Economy), Korea and Microsoft Research, under IT/SW Creative research program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2012-H0503-12-1024).

rating with target user. Then, it predicts unknown rating value for items that target user has not rated. Currently, CF has been successfully implemented in both information filtering and e-commerce applications. However, it remains some important challenges such as cold start, data sparsity, and scalability reflected on quality and accuracy of prediction.

In order to overcome these challenges, many researchers have proposed various kinds of CF method such as hybrid CF, trust-based CF, social network-based CF, etc. In the purpose of improving the recommendation performance and prediction accuracy of standard CF, in this paper we propose a method which integrates traditional CF technique with social relationship between users discovered from user's behavior in social network i.e. Facebook. We identify user's relationship from behavior of user such as posts and comments interacted with friends in Facebook. We believe that social relationship implicitly inferred from user's behavior can be likely applied to compensate the limitation of conventional approach. Therefore, we extract posts and comments of each user by using Facebook Graph API and calculate feature score among each term to obtain feature vector for computing similarity of user. Then, we combine the result with similarity value computed using traditional CF technique. Finally, our system provides a list of recommended items according to neighbor users who have the biggest total similarity value to the target user.

In order to verify and evaluate our proposed method we have performed an experiment on data collected from our Movies Rating System. Prediction accuracy evaluation is conducted to demonstrate how much our algorithm gives the correctness of recommendation to user in terms of MAE. Then, the evaluation of performance is made to show the effectiveness of our method in terms of precision, recall, and F1-measure. Evaluation on coverage is also included in our experiment to see the ability of generating recommendation. The experimental results show that our proposed method outperform and more accurate in suggesting items to users with better performance. The effectiveness of user's behavior in social network particularly shows the significant improvement by up to 6% on recommendation accuracy. Moreover, experiment of recommendation performance shows that incorporating social relationship observed from user's behavior into CF is beneficial and useful to generate recommendation with 7% improvement of performance compared with benchmark methods. Finally, we confirm that interaction between users in social network is able to enhance the accuracy and give better recommendation in conventional approach.

Received : March 27, 2013 Revision : April 29, 2013 Accepted : May 04, 2013
 Type of Submission : English Fast-Track Corresponding author : Geun-Sik Jo

1. Introduction

Social network grows in tremendous popularity in recent years as an extremely large communication

service for enabling users to participate and connect with one another. In addition, users can share common interests, experiences, and their daily activities with family and friends. Users spend hours

per day in maintaining personal information and interacting with other people via posting, commenting, messaging, games, social events, and applications (Chen and Fong, 2010). Due to the growth of user's distributed information in social network, there is a great potential to utilize the social data to enhance the quality of existing recommender system. There are some researches focusing on social network analysis that investigate how social network can be used in recommendation domain (Kazienko and Musial, 2006; Groh and Ehlig, 2007). Among these researches, there are several topics related in taking advantages of the interaction between a user and others in social network that can be determined and known as social relationship (Kazienko and Musial, 2006; Yuan et al., 2009). Research study written in (Yuan et al., 2009) is one of the studies which exploit membership and friendship based social network as a part of the recommendation. The hybridization results by fusing users' social network embedded social relationship with CF method show the perfection of the recommendation quality. Furthermore, mostly user's decisions before purchasing some products rely on suggestion of people who have either the same preferences or closer relationship. For this reason, we believe that user's relationship in social network can provide an effective way to increase the quality in prediction user's interests of recommender system (De Meo et al., 2011). Therefore, social relationship between users encountered from social network is a common factor to improve the way of predicting user's preference in the conventional approach.

Recommender system is dramatically increasing in popularity and currently being used by many e-commerce sites such as *Amazon.com*, *Last.fm*, *eBay.com*, etc. Commonly there are two types of filtering model used in recommender system, content-based filtering and collaborative filtering. Collaborative filtering (CF) method is one of the essential and powerful techniques in recommender system to recommend the appropriate items to user by learning user's preferences. CF method focuses on user data and generates automatic prediction about user's interests by gathering information from users who share similar background and preferences (Chen and Fong, 2010). Specifically, the intension of CF method is to find users who have similar preferences and to suggest target user items that were mostly preferred by those nearest neighbor users. There are two basic units that need to be considered by CF method, the user and the item. Each user needs to provide his rating value on items i.e. movies, products, books, etc to indicate their interests on those items. In addition, CF uses the user-rating matrix to find a group of users who have similar rating with target user. Then, it predicts unknown rating for items that target user has not rated. Currently, CF has been successfully implemented in both information filtering and e-commerce applications, but it remains some important challenges. The first challenge is that it uses only explicit numerical rating therefore it is impossible to incorporate it into systems that contain no explicit rating (Sirawit and Kijirikul, 2011). Another weakness of current recommender system is cold start users, users who provided few

ratings or users who just started using the system. For these users, CF method tends to fail in providing recommendation since it cannot find the possible neighbors (Gao et al., 2011). Moreover, since both the number of items and consumers are large in popular e-commerce websites, user-rating matrix will also become really sparse. As a result, data sparsity can cause low accuracy and quality of recommender system.

In order to overcome these challenges, many researchers have proposed various kinds of CF method such as hybrid CF (Ahmed et al., 2011; Wang et al., 2010), trust-based CF (Bhuiyan et al., 2010), social network-based CF (Park and Cho, 2011; Liu et al., 2010), etc. In the aim of improving the recommendation performance and prediction accuracy of standard CF, in this paper we propose a method which incorporates social relationship between users discovered from user's behavior in social network i.e. Facebook¹⁾ with traditional CF technique. We identify user's relationship from behavior of user such as posts and comments text messages interacted with friends in Facebook. We believe that social relationship implicitly inferred from user's behavior can be likely applied to compensate the limitation of current approach. Therefore, we extract posts and comments in text form of each user by using Facebook Graph API²⁾ and calculate feature score among each term using Term Frequency-Inverse Document Frequency (TF-IDF)³⁾, a numerical statistic often used as a

weighting factor in information retrieval and text mining, to obtain feature vector for computing similarity of user. Then, we combine the result with the similarity value computed using traditional CF technique. Finally, our system provides a list of predicted items or top- N recommendation according to neighbor users who have the biggest total similarity value to the target user.

The rest of this paper is organized as follows. Related works are described in section 2. In section 3, we discuss in detail about how to incorporate social relationship discovered from user's behavior into CF. Section 4 shows about experimental evaluation and results of this research. Conclusion and future work are presented in the last section.

2. Related Works

2.1 User-based Collaborative Filtering

Collaborative filtering is one of the most popular techniques in recommender system used to recommend items based on preferences of similar users. CF method has two vital steps, collaborating phase and filtering phase. First, CF collects taste information from many users then using information gleaned from neighbor users for prediction and finally recommendation of user's interests were automatically generated. CF can be divided into two main categories, user-based and item-based CF (Mu et al., 2010). User-based CF is also called nearest-neighbor based CF that utilizes the entire user-rating data to generate prediction. It uses statistical technique to find user's nearest neighbors. Once the nearest neighbors of user are found,

1) <http://www.facebook.com>.

2) <http://developers.facebook.com/docs/reference/api>.

3) <http://en.wikipedia.org/wiki/Tf-idf>.

prediction of top- N recommendation for target user is computed using weighted average of neighbors' ratings. Unlike user-based CF algorithm, the item-based approach looks into the set of items the target user has rated before computing similarity value to the target item and then selects k -most similar items to give the prediction (Mu et al., 2010; Shi et al., 2008).

Since we apply our proposed recommendation method in movie domain, the rating information of users on movies plays a very important role for predicting user's preferences. Users are allowed to rate the movies in order to represent how they liked those movies. After users rate the movies, we collect rating information from each user and represent them into user-rating matrix. In user-rating matrix, there consists of a list of n movies denote as:

$M = \{Movie_1, Movie_2, Movie_3 \dots Movie_n\}$, a list of m users $U = \{User_1, User_2, User_3 \dots User_m\}$ and rating matrix $U \times M$, where R_{ui} defined as rating value of user u on item i .

Once the whole set of user-rating matrix is collected, we use the user-based CF technique to find the nearest neighbors of target user who share similar interests. In order to obtain the neighbors of target user, there are many different ways to calculate similarity between target user and other users such as Pearson Correlation, Cosine Similarity, Adjusted Cosine Similarity, etc to clarify who have the highest similarity value. Pearson Correlation is one of the most popular similarity measurements in CF, so we use Pearson Correlation to calculate the similarity of user in user-rating matrix as shown

in equation (1). The result shows the predicted likeliness of target user to other users and it also uses to select top nearest neighbor users.

$$Rating_Sim(u, v) = \frac{\sum_{j \in CRI} (r_{uj} - avg(r_u)) \times (r_{vj} - avg(r_v))}{\sqrt{\sum_{j \in CRI} (r_{uj} - avg(r_u))^2 \sum_{j \in CRI} (r_{vj} - avg(r_v))^2}} \quad (1)$$

where r_{uj} is the rating value of item j by user u , $avg(r_u)$ is the average rating of user u , and CRI is the co-rated item set rated by user u and v . In order to select neighbors of target user who will serve as recommenders, there are two different techniques that have been employed, threshold-based selection and top- n selection. Threshold-based selection chooses neighbor users whose similarity value exceeds a certain threshold. On the other hand, top- n technique chooses n -most similar neighbors. In this case, we choose top- n selection technique to select the n -most similar neighbors based on similarity value calculated using equation (1). Since we get the neighbors of target user, we can calculate the predicted rating value of user u on item j by performing a weighted average of neighbors' ratings in equation (2).

$$p_{u,j} = avg(r_u) + \frac{\sum_{v \in KNN(u)} sim(u, v) \times (r_{v,j} - avg(r_v))}{\sum_{v \in KNN(u)} |sim(u, v)|} \quad (2)$$

where $avg(r_u)$ is the average rating of the target user u , $r_{v,j}$ is the rating of the neighbor user v to the target item j , $avg(r_v)$ is the average rating of neighbor user v , $sim(u, v)$ is the similarity value of the target user u and neighbor v , and $KNN(u)$ is the set of the neighbors selected using top- n selection technique.

2.2 Friendship Fusing Collaborative Filtering

Yuan et al. (2009) learns the role of two types of social relationship, membership and friendship, in order to fuse with standard CF to more accurately predict user's interests and produce better recommendation. Moreover, they presented two different platforms to integrate explicit social relationship into standard CF method, the weighted-similarity fusion and the graph fusion via random walk, to identify the best performance platform. Weighted-similarity fusion is conducted by leveraging the two social relationships to strengthen user similarity calculation process by combining similarity from friendship and/or membership with similarity from user-rating matrix. There are some differences between fusing with friendship and membership. When fusing friendship with user similarity from rating matrix, first they need to get user similarity based on the friendship. They represent friendship of users in the format of user-user matrix. For instance, if two users u_i and u_j are friends, then the value given to cell u_{ij} is set to 1, otherwise 0. In addition, they use adopting Cosine Correlation in order to calculate the friendship similarity. Next, they calculate the final similarity value between two users by combining similarity calculated from user-rating matrix and similarity computed from user-user matrix in a weighted approach. On the other hand, first they need to obtain user-user similarity based on membership data when fusing the membership. Since membership focuses on the relationship between user and the communities or groups that user joined, they represent the membership in the form of user-group matrix where the

rows denote users and the columns denote the groups joined by users. Next, both user-rating matrix and user-group matrix are used to find the final similarity value for prediction. In addition, fusing these two relationships together on the rating matrix is also adopted in the experiment. However, novel graph fusion model is used to fuse the social relationship with rating matrix using random walk technique in order to generate neighbor users for recommendation. Before applying random walk and similarity measures to generate neighborhood similarities for recommendation, it is very important to construct graph for social community when fusing via graph. In the experimental analysis, it shows the significant improvement of recommendation quality by embedding social relationship in standard CF via graph model.

We choose the method proposed in (Yuan et al., 2009) using weighted-similarity fusion as one of the benchmark algorithms to compare with our proposed method. This method improves recommendation quality by considering two types of social relationship, membership and friendship. However, we take into account only friendship relationship in our experiment and named as Friendship Fusing Collaborative Filtering (FFCF) because friendship fusing outputs better results than membership and the combination of friendship and membership in their experiment.

3. Incorporating Social Relationship discovered from User's Behavior into CF

In this section, we will explain in detail about

our proposed method which is divided into two parts such as system architecture and our method named as social relationship-based collaborative filtering (SRBCF).

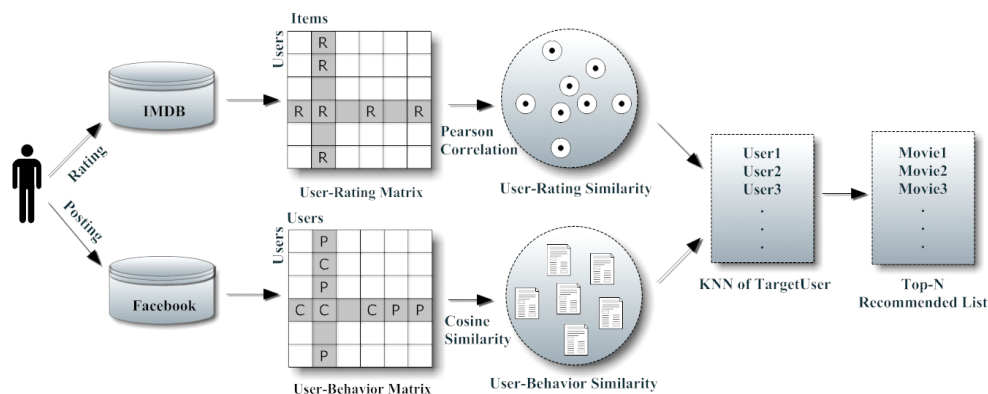
3.1 System Architecture

We introduce our system architecture as shown in <Figure 1> which merges the social relationship of each user discovered from user's behavior in Facebook with traditional CF method. Both posts text and comments text are considered as user's behavior in our approach. In our scenario, we consider the scope which a user is using Facebook to share their opinions or to give ideas to family and friends as suggestion or advices and user also has rated their favorite movies extracted from IMDB. First, from user's rating information we build a user-rating matrix that contains user, item, and rating value. Then, user-behavior matrix was conducted by crawling posts and comments text messages of user from Facebook using Facebook Graph API. Next, our system calculates

user-rating similarity value and user-behavior similarity value using Pearson Correlation and Cosine Similarity respectively (Hameed et al., 2012). In addition, we combine both similarity value from user-rating matrix and user-behavior matrix by using weighted-similarity combination approach and then select the nearest neighbor users who have the biggest total similarity value. The predicted rating value of target user is computed according to those neighbor users who share the similar preference to target user. Finally, our system provides a list of top- N recommended items to target user.

3.2 Social Relationship-based Collaborative Filtering (SRBCF) Method

In this section, we will discuss in detail about our method (SRBCF) in mathematical way for recommendation. Enrichment of social relationship from user's behavior in social network is presented in the first sub section and weighted-similarity combination approach is described in the next sub



<Figure 1> System Architecture

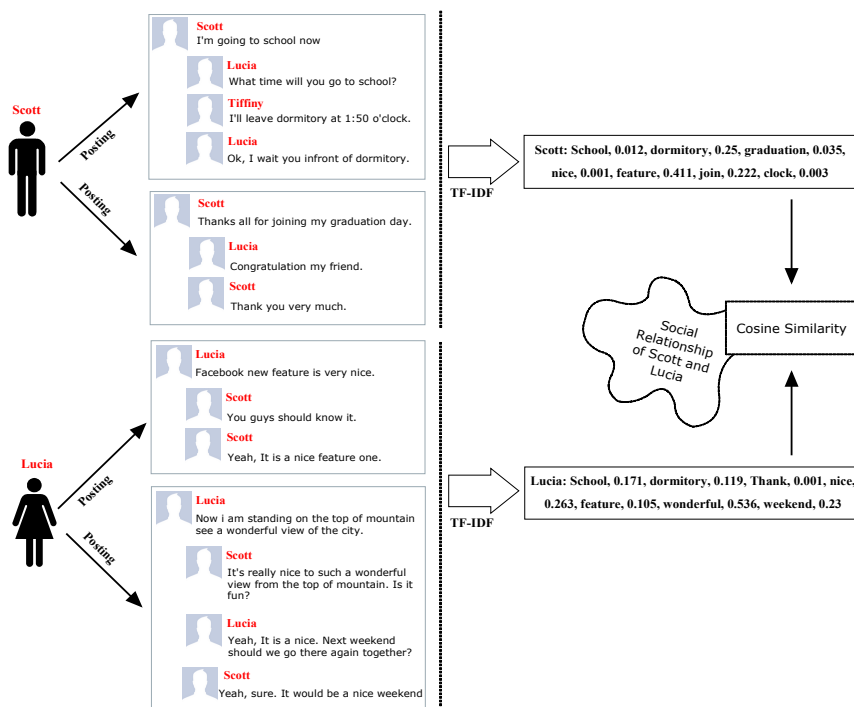
section.

3.2.1 Social Relationship discovered from User's Behavior

There are many activities user can do such as posting an article, commenting on a topic, sharing photos or videos, etc via Facebook. Within a big amount of social data available in user's profile, enrichment of social relationship of user by using their activities or behavior between friends in social network is needed because we propose a method to combine this relationship strength with user's interests. In our scope, we consider posts and comments text as user's behavior because common functionality of social network is to share text

articles. In addition, when a user posts an article in Facebook, they may get a lot of responding comments from friends so we will take into account only comments of friends that we want to find relationship strength with target user.

First, we use Facebook Graph API to obtain posts and comments text between two users. Then, user-behavior matrix was built where the row and column are represented as users and the cell of the matrix is the text of posts and comments between two users. For instance, we have to find social relationship strength between Scott and his friend, Lucia, so we have to get posted articles from Scott and all comments of Lucia on each posted article of Scott and vice versa as shown in <Figure 2>.



<Figure 2> Example of Social Relationship Discovered from User's Behavior

Each posted article of Scott and Lucia is considered as a document so each document has both text messages of Scott and text messages of Lucia. Therefore, we separate both text messages of Scott and Lucia for all of the documents and then collect text messages of Scott and Lucia separately within all documents.

Next, our system will automatically generate the feature vector from the retrieved posts and comments text of Scott and Lucia. Feature score of each term is calculated using TF-IDF but in order to use TF-IDF properly and efficiently we have to run a few steps of preprocessing on retrieved text. We performed some preprocessing phases such as transforming from varied form to its root form i.e. plural noun to singular noun and adverb with postfix 'ly' to adjective, removing stopwords, special symbols (@, ?, &, *, etc), digits and multiple repetitions of characters in words (goood to good). In our system, we consider only text messages that are written in English.

$$Behavior_Sim(u, v) = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{(\sum_{i=1}^n A_i)^2} \times \sqrt{(\sum_{i=1}^n B_i)^2}} \quad (3)$$

where A_i is the feature score of each term of $User_1$ and B_i is the feature score of each term of $User_2$.

Finally, cosine similarity as shown in equation (3) is used for computing the similarity value between feature vectors of retrieved posts and comments text from two users. After calculating behavior similarity value of target user and his/her friends in Facebook, we get a list of

behavior similarity value that representing the social relationship.

3.2.2 Weighted-Similarity Combination Approach

Following system architecture shown in <Figure 1>, our approach is to combine similarity value calculating from user-rating matrix and user-behavior matrix using Pearson Correlation and Cosine Similarity respectively. In this case, we use weighted-similarity combination approach by control the weight of both similarity scores using α value changing from 0 to 1. We also did the experiment related to α value to see the effectiveness of α to the system. The weighted-similarity combination approach is defined as follows:

$$Total_Sim(u, v) = \alpha Behavior_Sim(u, v) + (1 - \alpha) Rating_Sim(u, v) \quad (4)$$

First, we need to get rating similarity from user-rating matrix by using equation (1) and behavior similarity from user-behavior matrix by using equation (3) in order to get total similarity of each user that computed by equation (4). Then, select nearest neighbors of target user who have the biggest total similarity value for predicting target user interests. We follow equation (2) to predict rating value of target user and finally recommend top-N movies to target user. Algorithm pseudocode of SRBCF is shown in <Figure 3>.

Algorithm

Input: Target user U_T , User-rating matrix URM , User-behavior matrix UBM , Neighborhood size of target user k , Similarity weighting α , Number of movie to recommend n

Output: Top- n movie recommendation

1. Algorithm SRBCF(U_T , URM , UBM , k , α , n)
2. Begin
3. For each user U_A in URM
4. $S_R = \text{Similarity}(U_T, U_A)$ //Computed using Pearson Correlation
5. End
6. $L_R = A$ list of S_R similarity value from user-rating matrix
7. For each user U_A in UBM
8. Get posts and comments text between U_T and U_A
9. $V_T = \text{TF-IDF}(\text{Text of } U_T)$ //Feature vector of user U_T
10. $V_A = \text{TF-IDF}(\text{Text of } U_A)$ //Feature vector of user U_A
11. $S_B = \text{Similarity}(V_T, V_A)$ //Computed using Cosine Similarity
12. End
13. $L_B = A$ list of S_B similarity value from user-behavior matrix
14. $L_T = \alpha L_B + (1-\alpha)L_R$ //Weighted-Similarity Combination
15. $NB = \text{Select top-}k \text{ users with the highest similarity value in } L_T$
16. Predict movie rating based on the nearest neighbors NB
17. Return top- n movies list to target user U_T
18. End

<Figure 3> SRBCF Algorithm

4. Experimental Environment and Results

In order to verify and evaluate our proposed method, we have performed experiment on dataset

collected from our Movies Rating System. Prediction accuracy evaluation is conducted to demonstrate how much our method gives the correctness of recommendation to user in terms of MAE. Then, the evaluation on performance is made to illustrate the effectiveness of SRBCF in terms of precision, recall, and F1-measure. Furthermore, we setup an experiment about prediction coverage to measure the domain of items over which the system can make recommendation. For confidently confirming that our proposed method is better in suggesting items to user, we also did the experiment compared with two benchmark algorithms such as Traditional Collaborative Filtering (TCF) and Friendship Fusing Collaborative Filtering (FFCF).

4.1 Dataset

We perform our experiment to observe the effectiveness of SRBCF by applying our proposed recommendation algorithm in movie domain which gets from IMDB⁴⁾ website using IMDB API⁵⁾ for downloading movies published in between year 2011 and 2012. After we retrieved movies from IMDB, we have implemented a Movies Rating System in order to allow users to express their favorite movies by rating the movies. The dataset contains 51 users and 2,507 movies. The rating value ranges from 1 to 5 for user to state their favorite movies and only Facebook users are allowed to use the system. Furthermore, we use Facebook Graph API to get approximate 1,750

4) <http://www.imdb.com>.

5) <http://www.imdbapi.com>.

latest posts of each user and all comments text for each post after getting permission by authenticating with our system. In this process, we only consider posts from user's news feed. We run experiment by randomly split dataset into 80% training set to apply our proposed method and 20% test set to assess the performance, effectiveness, and prediction accuracy of recommendation.

4.2 Evaluation Metrics

Mean Absolute Error: There are several methods that have been widely used in evaluating the accuracy of recommender system such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Normalized Mean Absolute Error (NMAE). In our approach evaluation, we use Mean Absolute Error metric, a statistic accuracy measurement adopted to evaluate recommendation accuracy, to determine the correctness of our proposed method. MAE evaluates the prediction accuracy through computing the value distinction between predicted rating value generated by system and actual rating value provided by user.

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_{u,i} - r_{u,i}| \quad (5)$$

where $r_{u,i}$ denotes the actual rating that user u give to an item i , $p_{u,i}$ is the predicted rating of user u on item i generated by our approach, and n is the number of movies in test set.

Precision, recall, and F1-measure: The "Precision-Recall" method is used to evaluate the recommendation performance. This evaluation

method has been primarily suggested by (Cleverdon and Keen, 1966) as evaluation metric in the field of information retrieval system and has been adapted to evaluate the performance of a set of recommended items (Herlocker et al., 2004). Due to the simplicity and popularity of these two metrics, they have been adopted for recommender system evaluation in the top- N recommendation approach. Equation (6) and (7) are used to compute precision and recall of top- N recommendation (Bhuiyan et al., 2010).

$$Precision = \frac{|\{\text{relevant movies}\} \cap \{\text{retrieved movies}\}|}{|\{\text{retrieved movies}\}|} \quad (6)$$

$$Recall = \frac{|\{\text{relevant movies}\} \cap \{\text{retrieved movies}\}|}{|\{\text{relevant movies}\}|} \quad (7)$$

Precision and recall are often conflict properties because increasing the recommendation set size is likely to improve recall, but reduce precision. To solve this conflict, we use F1-measure that is one of the most popular techniques for combining precision and recall together for completely evaluating performance of recommender system. It is defined as the harmonic mean of precision and recall as shown in equation (8).

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (8)$$

Coverage: According to (Herlocker et al., 2004), the coverage of recommender system is a measurement of the domain of items in the system over which the system can form predictions or make recommendations. System with lower cover-

age maybe less valuable to user since they will be limited in the decisions they are able to help with. The most common coverage computation is the measurement of the number of items for which predictions can be formed as a percentage of the total number of items (Ge et al., 2010). This kind of coverage is often called prediction coverage and defined as follow:

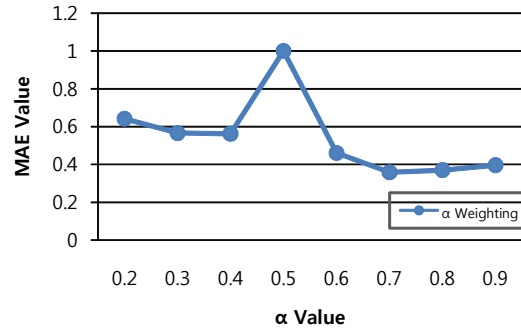
$$\text{Prediction Coverage} = \frac{|I_p|}{|I|} \quad (9)$$

where I denotes the set of available items and I_p defined as the set of items for which a prediction can be made. In addition, techniques that use to obtain I_p are different depending on the recommender technique. For instance, in this case I_p can be considered as the set of items that the rating value exceeds c .

4.3 Experimental results and Discussion

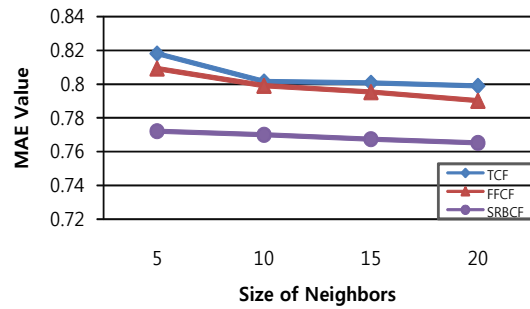
First of all, in the equation (4) of weighted-similarity combination approach we calculate the total similarity value of target user by weighting α . Therefore, in our experimental result as shown in <Figure 4> we can achieve a small MAE value with $\alpha = 0.7$ which means that our method provides small error rate in prediction by giving $\alpha = 0.7$. Furthermore, it also shows that user's rating matrix contributes 30% of the weight while user's behavior contributes 70% in calculating user-user similarity. In addition, we will take this α value in our further experiment.

Second, we did the experiment about MAE



<Figure 4> α Weighting

value to evaluate the prediction accuracy of our method. Since the size of neighborhood has a significant effect on the prediction accuracy, we calculate the MAE with number of neighbors ranging from 5 to 20. <Figure 5> shows the comparison of prediction accuracy with benchmark algorithms.



<Figure 5> Prediction Accuracy Comparison

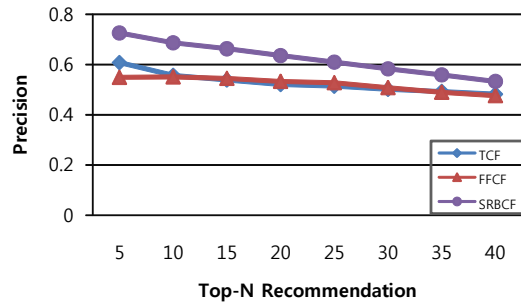
In <Figure 5>, it shows that our proposed method carry out lower MAE value compared with benchmark methods and it means that SRBCF can produce recommendation with higher prediction accuracy than TCF and FFCF in term of MAE value within our dataset. SRBCF also shows the

decreasing of MAE value when size of neighbors grows up. It implies that our system will offer high accuracy in recommending process with large size of neighbor users. According to the graph, FFCF can generate recommendation with smaller error rate than TCF but it gets higher MAE value compared with SRBCF. Therefore, social relationship implicitly inferred from user's behavior in social network of SRBCF algorithm is more powerful than social relationship explicitly encountered from friendship.

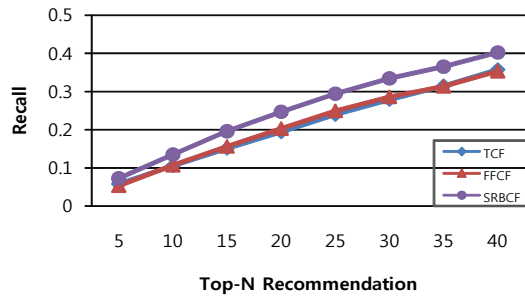
The result also demonstrates that the effectiveness of user's behavior in social network particularly showing the significant improvement by up to 6% on recommendation accuracy. Furthermore, it shows that the activities or behavior can illustrate the relationship strength between two users. User's behavior also can determine that which users share similar interests with him/her that help the recommender system in the neighborhood selection process. Sometimes, neighbor users selected by calculating similarity from user-rating matrix cannot be the good recommender for target user because they have no relation with target user. Taking into account of the interaction of users by their posts and comments in Facebook can help system filters out neighbor users that have no relation with target user. In addition, each user prefers suggestion from friends who keep in touch with them rather than the users who share only the same rating information. By analyzing posts and comments in Facebook to get user's social relationship, recommender system can increase accuracy of predicting user's interests. Therefore,

we can confirm that interaction between users in social network is able to enhance the prediction accuracy and give better recommendation.

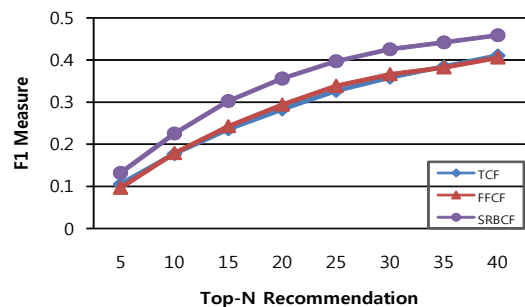
After the experiment about prediction accuracy, in order to evaluate performance of our method we did another experiment of precision, recall, and F1-measure that are well-known for evaluating top- N recommendation. Therefore, we apply precision, recall, and F1-measure metrics with top- N ranging from 5 to 40 to observe how well SRBCF can give a good recommendation result. <Figure 6>, <Figure 7>, and <Figure 8> are depicted the result of precision, recall, and F1-measure respectively. According to the results, we claim that SRBCF can produce recommendation with better performance than TCF and FFCF while these two benchmark methods can produce recommendation nearly the same performance. Furthermore, precision and recall can be linked to probabilities that directly affect the user. Our method can provide precision up to 70% which means that user can expect that on average 7 out of 10 recommended items will be relevant. According to F1-measure, SRBCF is able to achieve up to 7% improvement of performance and it means that relationship explored from interaction between users stated using posts and comments is better than relationship stated using friendship in integrating with CF. Integrating friendship with CF cannot enhance performance of conventional approach as much as SRBCF based on precision, recall, and F1-measure experiment. As a result, incorporating social relationship into CF algorithm improve not only the prediction accuracy but also



<Figure 6> Performance Evaluation with Precision



<Figure 7> Performance Evaluation with Recall

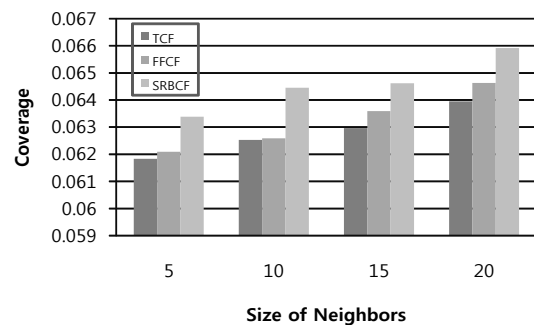


<Figure 8> Performance Evaluation with F1-measure

the performance of recommendation.

Finally, <Figure 9> shows the prediction coverage experiment of the three methods and it shows that our proposed method can form prediction or make better recommendation in terms

of prediction coverage. According to the graph, we observe that prediction coverage can be compared between FFCF and TCF when size of neighbors equal to 10. Prediction coverage constructed in the experiment based on our dataset shows SRBCF can form better prediction with big size of neighbors. In addition, according to the increasing trend of coverage and decreasing trend of accuracy in <Figure 5> generated by SRBCF, it is very efficient to form recommendation prediction with large size of neighbors. Recommender system with low prediction coverage cannot guarantee to generate a good recommendation which means that not only prediction accuracy is significant for recommendation method but also does the prediction coverage. Therefore, coverage must be measured in combination with accuracy so that recommenders are not tempted to raise coverage by making false prediction. From the experiment results of both prediction accuracy and prediction coverage, it shows that social relationship discovered from user's behavior also takes part to contribute the prediction coverage of the recommendation with perfect accuracy.



<Figure 9> Prediction coverage

5. Conclusion and Future Work

In this paper, we propose an effective collaborative filtering method that integrates social relationship of each user discovered from user's behavior in Facebook with traditional CF in order to improve performance and prediction accuracy of recommender system. It is a method which absorbs social data from user's profile in social network and rating information by gathering it using Movies Rating System and Facebook Graph API. From the Movies Rating System, we get rating information on movies of each user and also get permission from user to obtain social data from their personal profile by allowing each user to authenticate with our website before they can rate movies. We use Facebook Graph API to obtain posts and comments text from Facebook and apply TF-IDF technique to obtain feature score in order to calculate similarity of each user. Finally, in order to generate top- N recommendation we combine similarity calculated from user's behavior and rating information by using weighted-similarity combination approach. The experiment and evaluation have been conducted in order to illustrate the prediction accuracy and performance of the system. According to the high prediction accuracy of recommendation generated by SRBCF compared with TCF and FFCF within our dataset and high performance in recommending user's interests, we can claim that our method is more suitable than benchmark methods in incorporating social relationship into standard CF. Social relationship discovered from user's behavior, posts and

comments, can provide user with more accurate in prediction user's interests with better performance. Our method seeks to establish the effectiveness of user's behavior in social network that particularly shows the significant expansion by up to 6% on the accuracy in the conventional approach. SRBCF is also able to achieve up to 7% improvement of performance in terms of F1- measure. Thus, it reveals that relationship explored from interaction between users stated using posts and comments is capable to integrate with CF than relationship stated using friendship. Likewise, our system provides a good quality of recommendation with the big size of neighbors according to experiment on prediction accuracy and coverage. It means that when incorporating social relationship encountered from user's behavior into CF, it is common to perform recommendation with large size of neighbors. Moreover, our approach is easy to integrate with existing recommender systems in some environments such as smart phone or smart pad because currently those systems already integrated Facebook connect components i.e. Facebook login button. Lastly, applying our method can help system exploit social data available in user's profile in social network to improve accuracy, performance, and quality of conventional recommender system.

In our work, we mainly consider two types of social attributes, posts and comments represented as user's behavior in social network. However there are many attributes such as personal information, privacy, user liked page, user's interests, etc. In future work, we can take more attributes into account to get better prediction accuracy and

performance of recommender systems and we will do the experiment with cold-start or sparsity problem. In addition, we will generalize the results by conducting a large-scale empirical studying and testing our approach on large movie dataset such as MovieLens. Moreover, we will try to observe which social attributes are the most valuable and useful for social relationship in recommender systems.

References

- Ahmed, S., J. W. Kim, and S. G. Kang, "Enhanced Recommendation Algorithm using Semantic Collaborative Filtering : E-commerce Portal," *Journal of Intelligence and Information Systems*, Vol.17, No.3(2011), 79~98.
- Bhuiyan, T., Y. Xu, and A. Josang, SimTrust: A New Method of Trust Network Generation, in Proceeding of 2010 8th IEEE/IFIP International Conference on Embedded and Ubiquitous Computing (EUC), Hong Kong, China, 2010.
- Chen, W. and S. Fong, Social Network Collaborative Filtering Framework and Online Trust Factors : A Case Study on Facebook, in Proceeding of Fifth International Conference on Digital Information Management (ICDIM), Thunder Bay, ON, Canada, 2010.
- Chen, W., R. Khoury, and S. Fong, "Web 2.0 Recommendation service by multi-collaborative filtering trust network algorithm," *Information Systems Frontiers*, September, 2012.
- Choi, K. H., G. W. Kim, D. H. Yoo, and Y. M. Suh, "New Collaborative Filtering Based on Similarity Integration and Temporal Information," *Journal of Intelligence and Information Systems*, Vol.17, No.3(2011), 147~168.
- Cleverdon, C. and M. Keen, "Factors Determining the Performance of Indexing Systems," *ASLIB Cranfield Research Project*, Cranfield, 1966.
- De Meo, P., E. Ferrara, and G. Fiumara, A. Provetti, Improving Recommendation Quality by Merging Collaborative Filtering and Social Relationships, in Proceeding of International Conference on Intelligent Systems Design and Applications (ISDA), Cordoba, Spain, 2011.
- Gao, Y. L., B. Xu, and H. M. Cai, Information Recommendation Method Research Based on Trust Network and Collaborative Filtering, in Proceeding of IEEE 8th International Conference on e-Business Engineering (ICEBE), Beijing, China, 2011.
- Ge, M., C. Delgado-Battenfeld, D. Jannach, Beyond Accuracy : Evaluating Recommender Systems by Coverage and Serendipity, in RecSys 2010 Proceeding of the fourth ACM conference on Recommender Systems, Barcelona, Spain, 2010.
- Groh, G. and C. Ehmig, Recommendation in Taste Related Domains : Collaborative Filtering vs. Social Filtering, in Proceeding of the 2007 international ACM conference on supporting group work, Sanibel Island, Florida, USA, 2007.
- Hameed, M. A., O. A. Jadaan, and S. Ramachandram, "Collaborative Filtering Based Recommendation System : A survey," *International Journal on Computer Science and Engineering (IJCSSE)*, Vol.4(2012).
- Herlocker, J. L., J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *Journal of ACM Transactions on Information Systems*

- (TOIS), Vol.22(2004), 5~53.
- Kazienko, P. and K. Musial, "Recommendation Framework for Online Social Networks," *Advance in Web Intelligence and Data Mining Studies in Computational Intelligence*, Vol.23 (2006), 111~120.
- Liu, F. and H. J. Lee, "Use of social network information to enhance collaborative filtering performance," *Expert Systems with Applications*, Vol.37(2010), 4772~4778.
- Mu, X. W., Y. Chen, and T. Y. Li, User-Based Collaborative Filtering Based on Improved Similarity Algorithm, in Proceeding of 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Chengdu, China, 2010.
- Park, J. H. and Y. H. Cho, "Social Network Analysis for the Effective Adoption of Recommender Systems," *Journal of Intelligence and Information Systems*, Vol.17, No.4(2011), 305~316.
- Shi, X. Y., H. W. Ye, and S. J. Gong, A Personalized Recommender Integrating Item-based and User-based Collaborative Filtering, in Proceeding of ISBIM 08 International Seminar on Business and Information Management, Wuhan, China, 2008.
- Sirawit, S. and B. Kijisirikul, A Step Towards High Quality One-class Collaborative Filtering using Online Social Relationships, in Proceeding of International conference on Advanced Computer Science and Information System (ICACSIS), Jakarta, Indonesia, 2011.
- Wang, Q., X. H. Yuan, and M. Sun, Collaborative Filtering Recommendation Algorithm based on Hybrid User Model, in Proceeding of 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Yantai, Shandong, China, 2010.
- Yuan, Q., Sh. Zhao, L. Chen, Y. Liu, Sh. Ding, X. Zhang, and W. Zheng, Augmenting Collaborative Recommender by Fusing Explicit Social Relationships, in Proceeding of ACM RecSys '09 Workshop on Recommender Systems and The Social Web, New York, USA, 2009.

Abstract

사용자 행동 기반의 사회적 관계를 결합한 사용자 협업적 여과 방법

타이쎈타* · 하인애* · 조근식**

소셜 네트워크는 사용자들의 공통된 관심사, 경험, 그리고 일상 생활들을 함께 공유하기 위해 소셜 네트워크 상 사람들을 서로 연결시켜주는 거대한 커뮤니케이션 플랫폼이다. 소셜 네트워크상의 사용자들은 포스팅, 댓글, 인스턴스 메시지, 게임, 소셜 이벤트 외에도 다양한 애플리케이션을 통해 다른 사용자들과 소통하고 개인 정보 관리하는데 많은 시간을 소비한다. 소셜 네트워크 상의 풍부한 사용자 정보는 추천시스템이 추천 성능을 향상시키기 위해 필요한 큰 잠재력이 되었다. 대부분의 사용자들은 어떤 상품을 구매하기 전 가까운 관계이거나 같은 성향을 가진 사람들의 의견을 반영하여 의사 결정을 하게 된다. 그러므로 소셜 네트워크에서의 사용자 관계는 추천시스템을 위한 사용자 선호도 예측을 효율적으로 높이는데 중요한 요소라 할 수 있다. 일부 연구자들은 소셜 네트워크에서의 사용자와 다른 사용자들 사이의 상호작용 즉, 소셜 관계(social relationship)와 같은 소셜 데이터가 추천시스템에서 추천의 질에 어떠한 영향을 미치는가를 연구하고 있다.

추천시스템은 아마존, 이베이, Last.fm과 같은 큰 규모의 전자상거래 사이트 또한 채택하여 사용되는 시스템으로, 추천시스템을 위한 방법으로는 협업적 여과 방법과 내용 기반 여과 방법이 있다. 협업적 여과 방법은 사용자들의 선호도 학습에 의해 사용자가 아직 평가하지 않은 아이템 중 선호할 수 있는 아이템을 정확하게 제안하기 위한 추천시스템 방법 중 하나이다. 협업적 여과는 사용자들의 데이터에 초점을 맞춘 방법으로 유사한 배경과 선호도를 가지는 사용자들로부터 정보를 수집하여 사용자들의 선호도 예측을 자동적으로 발생시킨다. 특히 협업적 여과는 근접한 이웃 사용자들에 의해서 목적 사용자가 선호할 수 있는 아이템을 제시하는 것으로 유사한 이웃 사용자를 찾는 것이 중요하다. 좋은 이웃 사용자 발견은 사용자와 아이템을 고려하는 방법이 일반적이다. 각 사용자는 아이템 즉, 영화, 상품, 책 등에 자신의 선호도를 나타내기 위하여 평가 값을 입력하고, 시스템은 이를 바탕으로 사용자-평가 행렬을 구축한다. 이 사용자-평가 행렬은 목적 사용자와 유사하게 아이템을 평가한 사용자 그룹을 찾기 위한 것으로, 목적 사용자가 아직 평가하지 않은 아이템에 대하여 사용자-평가 매트릭스를 통해 그 평가 값을 예측한다. 현재 이 협업적 여과 방법은 전자상거래와 정보 검색에서 적용되어 개인화 시스템에 효율적으로 사용되고 있다. 하지만 초기 사용자 문제, 데이터 희박성 문제와 확장성 그리고 예측 정확도 향상 등 해결해야 할 과제가 여전히 남아 있다.

이러한 문제들을 해소하기 위해 많은 연구자들은 하이브리드, 신뢰기반, 소셜 네트워크 기반 협업적 여과와 같은 다양한 방법을 제안하였다. 본 논문에서는 전통적인 협업적 여과 방식의 예측 정확도와 추천 성능을 향상시키기 위해 소셜 네트워크에 존재하는 소셜 관계를 이용한 협업적 여과 시스템을 제안한다. 소셜 관계는 소셜 네트워크 서비스 중 하나인 페이스북 사용자들이 남긴 포스팅과 사용자의 소셜 네트워크 친구와 의견 교류 중 남긴 코멘트와 같은 사용자 행동을 기반으로 정의된다. 소셜 관계를 구축하기 위해 소셜 네트워크 사용자의 포스팅과 댓글을 추출하고, 추출된 텍스트에 불용어 및 특수 기호 제거와 스테밍 등 전처리를 수행하였다. 특징 벡터는 TF-IDF를 이용하여 전처리된 텍스트에 나타난 각 단어에 대한 특징 점수를 계산함으로써 구축된다. 본 논문에서 이웃 사용자를 결정하기 위해

* 인하대학교 컴퓨터정보공학과

** 인하대학교 컴퓨터정보공학부, 교신저자

사용되는 사용자 간 유사도는 특징 벡터를 이용한 사용자 행동 유사도와 사용자의 영화 평가를 기반으로 한 전통적 방법의 유사도를 결합하여 계산된다. 제안하는 시스템은 목표 사용자와 제안한 방법을 통해 결정된 이웃 사용자 집단을 기반으로 목표 사용자가 평가하지 않은 아이템에 대한 선호도를 예측하고 Top-N 아이템을 선별하여 사용자에게 아이템을 추천하게 된다.

본 논문에서 제안하는 방법을 확인하고 평가하기 위하여 IMDB에서 제공하는 영화 정보 기반으로 영화 평가 시스템을 구축하였다. 예측 정확도를 평가하기 위해 MAE 값을 이용하여 제안하는 알고리즘이 얼마나 정확한 추천을 수행하는지에 대한 예측 정확도를 측정하였다. 그리고 정확도, 재현율 및 F1값 등을 활용하여 시스템의 성능을 평가하였으며, 시스템의 추천 품질은 커버리지를 이용하여 평가되었다. 실험 결과로부터 본 논문에서 제안한 시스템이 보다 더 정확하고 좋은 성능으로 사용자에게 아이템을 추천하는 것을 볼 수 있었다. 특히 소셜 네트워크에서 사용자 행동을 기반으로 한 소셜 관계를 이용함으로써 추천 정확도를 6% 향상시킴을 보였다. 또한 벤치마크 알고리즘과의 성능비교 실험을 통해 7% 향상된 추천 성능의 결과를 보여준다. 그러므로 사용자의 행동으로부터 관찰된 소셜 관계를 CF방법과 결합한 제안한 방법이 정확한 추천시스템을 위해 유용하며, 추천시스템의 성능과 품질을 향상시킬 수 있음을 알 수 있다.

Keywords : 추천시스템, 소셜 관계, 소셜 네트워크 분석, 협업적 여과 방법, 사용자 행동

저 자 소 개



Setha Thay

Setha Thay received a B.S. degree in Computer Science from the Royal University of Phnom Penh, Cambodia, in 2010. He is a M.S. Candidate majoring in Computer and Information Engineering at Inha University, Korea. His research interests include Social Network, Recommender System, Data Mining, and Semantic Web.



Inay Ha

Inay Ha received a B.S. degree Computer Science from Suwon University, Korea, in 2005, and a M.S. degree in Information Engineering from Inha University, Korea in 2007. She is currently a Ph.D. Candidate in Information Engineering of Inha University, Korea. Her research interests include Information Personalization, Semantic Web, Social Network and Recommender System.



Geun-Sik Jo

Geun-Sik Jo is a Professor in Computer and Information Engineering, Inha University, Korea. He received the B.S. degree in Computer Science from Inha University in 1982. He received the M.S. and the Ph.D. degrees in Computer Science from City University of New York in 1985 and 1991, respectively. He has been the General Chair and/or Technical Program Chair of more than 20 international conferences and workshops on artificial intelligence, knowledge management, and semantic applications. His research interests include knowledge-based scheduling, ontology, semantic Web, intelligent E-Commerce, constraint-directed scheduling, knowledge-based systems, decision support systems, and intelligent agents. He has authored and coauthored five books and more than 200 publications.