ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2016 December: 22(4): 1~18 http://dx.doi.org/10.13088/jiis.2016.22.4.001

Recommender Systems using SVD with Social Network Information*

Min-Gun Kim
Ucess Partners
(*mgkim@ucesspartners.com*)

Kyoung-jae Kim
Dongguk University_Seoul
(*kjkim@dongguk.edu*)

Collaborative Filtering (CF) predicts the focal user's preference for particular item based on user's preference rating data and recommends items for the similar users by using them. It is a popular technique for the personalization in e-commerce to reduce information overload. However, it has some limitations including sparsity and scalability problems. In this paper, we use a method to integrate social network information into collaborative filtering in order to mitigate the sparsity and scalability problems which are major limitations of typical collaborative filtering and reflect the user's qualitative and emotional information in recommendation process. In this paper, we use a novel recommendation algorithm which is integrated with collaborative filtering by using Social SVD++ algorithm which considers social network information in SVD++, an extension algorithm that can reflect implicit information in singular value decomposition (SVD). In particular, this study will evaluate the performance of the model by reflecting the real-world user's social network information in the recommendation process.

Key Words : Recommender systems, Social network information, Collaborative filtering, Singular value decomposition, Business analytics

Received : December 5, 2016 Revised : December 19, 2016 Accepted : December 21, 2016
Publication Type : Regular Paper Corresponding Author : Kyoung-jae Kim

1. Introduction

Due to the generalization of Internet commerce, the information overflow phenomenon in electronic transactions has deepened. The product recommender system plays an important role in supporting the purchase decision of the customer by mitigating the information overflow phenomenon of the Internet. In the meantime, various product recommendation techniques have been proposed to realize the recommender system, but collaborative filtering (CF) is widely used in industry and academia. (Funakoshi and Ohguri, 2000; Wang and Wu, 2012). Collaborative filtering is a method of predicting the preference of a new customer for another product by analyzing the purchase relationship between the existing customer and the customer who wants to receive

* This work was supported by the research program of Dongguk University, 2015.

the recommendation based on the evaluation score of the existing customer who performs a similar evaluation (Adomavicius and Tuzhilin, 2005; Sarwar et al., 2000).

Meanwhile, the previous research pointed out that the problem in recommendation process for new customers or products (a.k.a the Cold-start problem), sparsity problem, scalability problem may occur since collaborative filtering recommends using past purchasing or evaluation records (Ghazanfar and Prügel-Bennettb, 2014; Sarwar et al., 2000; Su and Khoshgoftaar, 2008).

Problems in the process of recommending for new customers or products are that new customers will not receive recommendations, or in the case of a new product, there is a problem that recommendation cannot be made because collaborative filtering will recommend based on past purchases or scores. This problem is commonly referred to as the Cold-start problem. On the other hand, the sparsity problem is a sparseness of the user-item matrix, which is the basis of collaborative filtering, so that it cannot properly select similar products or users in the process of collaborative filtering. In other words, in the case that there is no rating data for the same product among the users due to the scarcity of rating data, even users with similar preferences cannot be selected as similar users in the collaboration filtering and thus have no influence on the recommendation process it means. Therefore, the cold-start problem described above is a special form of the sparsity problem. Finally, the scalability problem means that the calculation process such as searching for similar users for collaborative filtering is delayed when the number of products to be traded is continuously increased or the number of users participating in the transaction is continuously increased. In the case of a rapid response needed such as an Internet shopping mall, the scalability problem is an important limitation to lose customers.

Recently, many researches have been carried out to reflect users' emotional tastes through the use of social network information according to the proliferation of social network service. Typical collaborative filtering uses only user's rating, which is external and explicit information, but there is a limit that users' emotional and implicit information cannot be reflected. In order to compensate for this, recent research attempts to incorporate implicit information such as user's purchase frequency, social network information, and Internet visit record into collaborative filtering. In addition, implicit information can be used even when there is no explicit evaluation score of the customer, which is explicit information, and can contribute to alleviate the sparsity problem or the cold-start problem of the collaboration filtering.

In this paper, we use a method to integrate social network information into collaborative filtering in order to mitigate the sparsity and scalability problems which are major limitations of typical collaborative filtering and reflect the user's qualitative and emotional information in recommendation process. In this paper, we use a novel recommendation algorithm which is integrated with collaborative filtering by using

Social SVD++ algorithm which considers social network information in SVD++, an extension algorithm that can reflect implicit information in singular value decomposition (SVD). In particular, this study will evaluate the performance of the model by reflecting the real-world user's social network information in the recommendation process.

SVD is known to improve the scalability in large-scale data using Eigen values that cannot be extracted from the existing preference score data, through decomposing the dimension of the preference score in the user-item matrix of collaborative filtering. There are advantages that SVD++ uses implicit feedback data such as user's number of visits and social network information in the recommendation process, while SVD only uses the preference score on user-item matrix, which is explicit feedback data. For this purpose, we will use modified algorithm to apply social network information to typical SVD++ and apply it to collaborative filtering.

Especially, in this study, we will verify the performance of user's point-of-interest recommender system by using external evaluation score and social network information about user's point-of-interest. For this purpose, we will gather social network information and preference score data of real-world users.

The composition of this study is as follows. Section 1 describes the background and necessity of the research. Section 2 introduces the related research and Section 3 explains the Social SVD++ used in this study. Section 4 explains the research data, design, and experimental results. Section 5 describes the conclusions and limitations of the study and future research directions.

2. Prior Research

2.1 Recommender systems and dimension reduction

When a user evaluates a product, his / her preference is assigned a score in collaborative filtering. The preference of a person is determined by various factors. That is, the individual's preference for a particular product is determined by a combination of various factors such as price, quality, fun, taste, and so on. Therefore, analytical techniques that can analyze such high-dimensional data are needed.

The user-item matrix of collaborative filtering is generally expressed as a two-dimensional matrix with users as rows and products as columns, which is a matrix. The purpose of collaborative filtering is to estimate the empty evaluation scores that should be predicted based on the established evaluation scores. To do this, we analyze the association between users and products in the evaluation score matrix. As described above, since the preference score is high-dimensional data reflecting the user's preference for the product, it is very difficult to predict the preference score of another person by analyzing it as it is. Therefore, decomposing into a low-dimensional data space that is easy to understand makes prediction easier.

To do this, we can use a method such as LSI (Latent Semantic Indexing), which has done much research in the field of information retrieval.

Originally, LSI is a method to improve the accuracy of search through analyzing the patterns of simultaneous appearance between documents and terms appearing in the document, and clarifying the association (Deerwester et al., 1990). In order to extract potential semantic information through the association analysis between documents and terms, LSI generates a high-dimensional initial term-document matrix and decomposes it into a low-dimensional matrix through singular value decomposition (SVD). Through such a decomposition process, high-dimensional data characteristics expressed in a raw matrix can be expressed as a low-dimensional disassembled matrix. Likewise, the user-item matrix of collaborative filtering can be decomposed into low-dimensional matrices using SVD. Through this process, it is possible to extract the characteristics and patterns of various factors that cannot be found in the original matrix by decomposing the complex relationship between the various factors that cannot be extracted from the high-dimensional data form.

In particular, it is known that SVD can mitigate the sparseness of data by generating low-dimensional element values by decomposing the data with no element values in the process of decomposing into a low-dimensional matrix space. It has been known that SVD has the advantage of reducing the problem of scalability because it has a function to reduce the factors having the pattern in the case of large data. Due to these reasons, it has recently been used in many recommender system studies.

SVD is used for collaborative filtering in the Netflix contest, which complements the performance of recommendation algorithms (Funk, 2006; Gorrell and Funk, 2006), Sarwar et al. (2000) proposed a method of estimating the score after reducing the dimensions of movie and commerce data using SVD. Lee and Kim (2007) applied SVD to EachMovie data to extract user's taste space and use it as a recommendation. Kim et al. (2002) used SVD and collaborative filtering to create a hybrid recommendation model with knowledge base recommender system using MovieLens data. On the other hand, Billsus and Pazzani (1998) used artificial neural networks and SVD-based recommendation algorithms to reduce the dimension of feature space. Several previous studies have used dimension reduction techniques such as SVD for the recommender system, and the following briefly explains the principle of SVD.

2.2 SVD

SVD is a form of matrix factorization. When used in a recommendation system, SVD is a method of internally projecting the user's and the product's rating score matrix to a specific dimension of user and product combining latent factor space (Ricci et al., 2011). For example, suppose that there is a user-item matrix consisting of user u and product i, and we can think of matrix S with each user's evaluation score as an

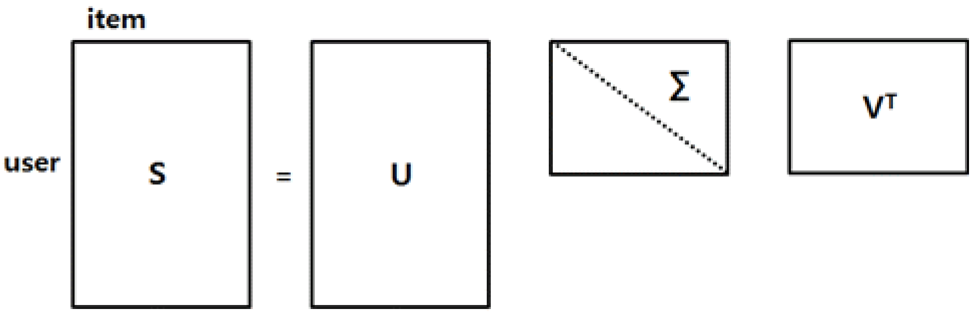

Fig. 1 Decomposition of user-item matrix suing SVD (Edited after excerpt from Deerwester et al., 1990)

element for each product. Through SVD, the matrix S is decomposed into several small matrices. That is, the matrix S can be decomposed as shown in Fig. 1 through SVD. If the size of the matrix S is u×k, U consists of u×u, Σ consists of u×k, and V^T consists of k×k sized matrix.

In Fig. 1, the matrix U is a left singular vector with orthogonal matrices, and is a matrix expressing the characteristics of the rows of the matrix S. The matrix V^T is a right singular vector with orthogonal matrices and is composed of the singular values of the matrix S. The matrix Σ is a diagonal matrix composed of singular values for the matrix S. A diagonal matrix means a matrix with only diagonal values and the remainder being zero. U is a matrix of eigenvectors corresponding to columns of a source matrix, and V is a matrix of eigenvectors corresponding to rows of a source matrix. That is, the matrix U is a product characteristic in a user-item matrix, and the matrix V is a matrix that reflects user characteristics. The matrix Σ is a diagonal matrix having eigenvalues of the corresponding elements. The eigenvalues are arranged in descending order, which means the importance of the features. By reducing the dimension of the matrix Σ, the dimension of the matrix U and the matrix V are reduced as well, and the dimensionally reduced matrix is constructed by extracting only the important features of the original matrix. It becomes possible to disassemble into low-dimensional matrices with easy and non-critical factors removed. In other words, in the recommendation process, some of the noise of the evaluation data can be removed, and the pure characteristic factor values can be extracted, thereby reducing the complexity of the data.

In general, SVD has the advantage of removing the noise and extracting the singular value to express the factor characteristic which is difficult to represent in the original matrix, but there is a disadvantage that some information may be lost through the noise removal process. In addition, there is a disadvantage in that it does not utilize other implicit data by using only the rating data of the user's product, which is explicitly existing explicit data. To overcome the disadvantage of not using the implicit data, the proposed method is SVD++ which will be explained next.

2.3 SVD++

SVD++ was proposed by Koren (2008) and can be said to be an extended SVD technique that allows the existing SVD to take into account further implicit feedback data. Koren (2008) proposed SVD++ as a hybrid algorithm that extended the NSVD model proposed by Paterek (2007) and the existing SVD. SVD++ takes into account not only the explicit data, but also the implicit data in the user-item matrix. Implicit data refers to the evaluation data of various users' products that are not explicitly expressed, for example, even if a user evaluates a specific product regardless of the evaluation score of the product, there is a possibility that there will be an interest in the product. For example, in the case of movie DVD, regardless of evaluation, it can be judged that there is a preference for the movie with only one record of rental. This is not considered in SVD, so SVD is expanded so that such data can be considered in SVD++.

In general, SVD++ is known to increase the accuracy of prediction by taking additional data into consideration while SVD is somewhat less predictive due to information loss in decomposition process. The NSVD mentioned above is suggested by Paterek (2007), it proposes an asymmetric factor model during matrix factorization to improve the performance of SVD. The most important advantage of NSVD is that it uses the product characteristic matrix instead of the user's characteristic matrix. It is proved through experiments that this method can reduce the size of the input data and raise the performance of the recommendation system since the number of users is usually larger than the number of products. In order to explain SVD++, it is necessary to start from baseline predictors and go through SVD, but the explanations are described in detail by Koren and Bell (2011), so a detailed description will be omitted.

Each product i is associated with a factor vector $q_i \in \mathbb{R}^f$, and when a user u is associated with a factor vector $p_u \in \mathbb{R}^f$, the factors q_i for a particular product i measure how much the product holds the factors positively or negatively. Also, the factors p_u for a particular user u measure how much interest the user has on relevant factors. The inner product $q_i^T p_u$ based on this result expresses the degree of the user's overall interest in the relationship between the user and the product, that is, the characteristics of the product. Thus, the rating of the user's product is predicted through Eq. (1).

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T p_u \qquad \text{Eq. (1)}$$

Here, μ mean the overall average rating, and b_u and b_i mean the observed deviations from the mean value for user u and product i, respectively. To estimate the coefficients of the model used in Eq. (1), we need to optimize Eq. (2) to minimize the regularized squared error.

Constants λ is generally determined through cross-validation and control the degree of constraint. Optimization is usually done by stochastic gradient descent or least squares.

On the other hand, SVD++ adds a second set of product factors to the above SVD to associate a factor vector $y_i \in \mathbb{R}^f$ with each product i. These new product factors are used to characterize users based on a set of products that users have evaluated. The modified model is shown in Eq. (3).

$\mathrm{R}(u)$ contains the products evaluated by user u. Here, user u is modeled as $p_u + |\mathrm{R}(u)|^{-\frac{1}{2}} \sum_{j \in \mathrm{R}(u)} y_j$. In Eq. (3), the user factor vector p_u is derived by the given explicit evaluation data. This vector is supplemented by the sum of $|\mathrm{R}(u)|^{-\frac{1}{2}} \sum_{j \in \mathrm{R}(u)} y_j$, which implies the implicit data. Since the y_j in equation (3) are controlled close to zero, their sum is normalized by $|\mathrm{R}(u)|^{-\frac{1}{2}}$ to stabilize the deviation over the range of the observed values of $|\mathrm{R}(u)|$. The coefficients of the model are determined by minimizing the associated constraint squared error function through a stochastic gradient descent algorithm.

In addition, various types of implicit data can be considered simultaneously by using a set of additional product factors. For example, if user u has an implicit preference for products in $\mathrm{N}^1(u)$ (for example, if the user rented the products) and different type of implicit data to the product $\mathrm{N}^2(u)$ (for example, if the user searched for the products), we can be use Eq. (4).

The relative importance of each source of implicit data is automatically learned according to the algorithms.

$$\min_{b_*, q_*, p_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \mu - b_i - b_u - q_i^T p_u)^2 + \lambda\ (b_i^2 + b_u^2 + \|q_i\|^2 + \|p_u\|^2) \qquad \text{Eq. (2)}$$

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T \left(p_u + |\mathrm{R}(u)|^{-\frac{1}{2}} \sum_{j \in \mathrm{R}(u)} y_j \right) \qquad \text{Eq. (3)}$$

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T \left(p_u + |\mathrm{N}^1(u)|^{-\frac{1}{2}} \sum_{j \in \mathrm{N}^1(u)} y_j^{(1)} + |\mathrm{N}^2(u)|^{-\frac{1}{2}} \sum_{j \in \mathrm{N}^2(u)} y_j^{(2)} \right) \qquad \text{Eq. (4)}$$

3. Social SVD++ CF

As mentioned in the earlier, Kumar et al. (2014) proposed the Social SVD++, which is a special form of the SVD++ model proposed by Koren (2008), which considers the social network data as one of the implicit evaluation data. In other words, it is assumed that the preference of a certain user is related to the preference of the other user having socially connected relation. In previous research, it has been suggested that preferences for products tend to show a similar tendency when there is a relationship of friendship in the social network.

In the recent recommendation system research, researches have been carried out to utilize the social network information and the result of social network analysis in recommendation process. In the early studies, the information of the social network was mainly reflected in the recommendation process, and the social network analysis was not used in earnest. Golbeck (2006) proposed a recommender system called Film Trust that incorporates Web-based social network information into a movie recommender system. Her research has shown that more accurate recommendations can be made when various ratings are presented compared to the average. Jyun-Cheng & Chui-Chen (2008) proposed a trading relationship recommender system based on seller's trust relationship in online auction. They proposed a collaborative based recommender system using two social network indicators, and claimed that 76% of the bad user accounts could be found through that system. Yang & Dia (2008) used social network information to generate targeted advertisements. They used the interaction data of customers to find out the detailed group with strong bonds, and estimate the likelihood of each customer by the product category. Debnath et al. (2008) proposed a recommender system that combines a content-based recommender system and collaborative filtering, and estimates the weight of the attributes used in the content-based recommendation through the linear regression analysis obtained from the social network graph. In addition, Liu & Lee (2010) proposed a method that reflects or includes weighted preference information of users in friendship in collaborative filtering using friend relationship information in social network service. They suggested that the preference of the users in the friendship relationship is more reflected in the recommendation process, and that the recommendation performance by the proposed method was improved compared with the existing method. Although the above studies attempted to incorporate the information of the social network into the recommendation process, it was not an actual use of the social network analysis, but an exploratory study without the empirical analysis or only the information of the social network was reflected in the recommendation process.

Relatively recent researches have tried to incorporate social network analysis rather than social network information into the recommendation process. Park et al. (2009) proposed a system that recommends products by combining the centrality concept of social network analysis with the product purchase network. They proposed a method of recommending products to other users by using the purchasing list of the upper level users with high centrality value, after establishing a customer network connecting users who purchase common products. In addition, Cho & Bang (2009) constructs a network of products purchased by the same customers, identifies the purchase relationship among the products using the concept of centrality of the social network analysis and they suggested a method to recommend new products. Meanwhile, Kang (2010) proposed a method of recommending new products to customers after extracting customers who are in a structural hole position of the social network analysis after establishing a customer network. In addition, Kang (2011) applied structural hole analysis to product network and applied it to recommend new products.

Recent studies have attempted to integrate the results of social network analysis into typical collaborative filtering using typical users' preference scores. Kim & Ahn (2010) proposed a recommender system using cluster-indexing collaborative filtering, which extracts the top 5 users and calculates them as an initial cluster center after calculating the centroid index of social network analysis. In addition, Kim and Kim (2014) proposed a recommender system using cluster-indexing collaborative filtering, which extracts the top 5 users after extracting the structural hole of social network analysis and uses them as an initial cluster center point. The proposed systems of the above two studies suggested a recommender system that integrates the social network analysis and the collaboration filtering by performing the cluster-indexing collaboration filtering using the results of the centrality analysis and the structural hole analysis, which are known as very important indicators in the social network analysis.

The contribution of the previous studies is that the results of the social network analysis or the social network data were reflected in the recommendation process and the recommendation process was improved by reflecting the qualitative factors such as the friendship relation between the users, which could not be extracted from the user-product matrix. However, there is another limitation in these researches that collaborative filtering does not take into account the limitations of sparsity and scalability that are inherent. However, the recommender system used in this research not only reflects the social network data including the user's qualitative information in the recommendation process but also alleviates the sparsity and scalability problems of collaborative filtering by using SVD++.

As described above, SVD has an advantage of alleviating the sparsity and scalability problems of collaborative filtering. SVD++ is an extended SVD algorithm, and the existing SVD is limited to merely use the preference score of the user-item matrix. We can improve the quality of the recommender system by considering the implicit data. The method used in this study reflects the social network information of the user which can be considered to contain the user's qualitative and emotional information among the implicit data that can be considered in SVD++. In this study, this is a recommender system that integrates social network information and SVD++ and is simply referred to as Social SVD++ CF.

The basic principle of Social SVD++ CF reflects social network information in the implicit data of SVD++. That is, the number of friends on the social network of users and the number of users who have evaluated about product are considered as implicit data. Social SVD++ CF has been developed by Kumar et al. (2014), but it was not verified using the real-world social network data. In this study, we try to verify the usefulness the model by using real-world users' product rating and social network information.

In order to verify the performance of Social SVD++ CF, we first collect data of preference evaluation in real-world and social network data of users. To do this, we build a web-based data collection system and collect data. After the user-item matrix is created, it is decomposed into low-dimensional data space by the SVD and the social network information included in the social network data is reflected by the modified SVD++. Finally, it is driven by collaborative filtering to generate recommendation results and verify the recommendation performance. For this purpose, the MAE (Mean Absolute Error), which means the absolute value of the difference between the predicted value of the product and the actual value, is calculated for each model, and the results of the SVD based collaborative filtering, the SVD++ based collaborative filtering, and the social SVD++ based collaborative filtering are compared. And we check the generalizability of the differences among models using paired sample t-test.

4. Research Data and Experiments

4.1 Research data

Prior research on recommender systems using SVD evaluated the usefulness of the research model through recommendation performance evaluation of large amount of data such as MovieLens data or EachMovie data. On the other hand, the data used in this study are product rating and social network data of real-world users. We will use two difference data sets. The first data (the Research Data 1) is data on overall preferences for customers visiting restaurants and cafés and their social networking data. The second data is the data used in Kim and Ahn (2010), which is rating information related to movie preference and social network information between the users. In order to optimize and generalize the parameters of SVD, the research data were divided into two parts: training set and test set, and then SVD, SVD++, and Social SVD++ were sequentially applied to the test set.

To collect first research data, a web-based customer rating data collection system was constructed. A total of 45 data were collected from domestic college students, but finally 39 research data were collected excluding 6 data with inaccurate answers. This research data collected various information such as satisfaction, reason of visit, and visit time etc. for restaurants, cafes, and bars located in or near the university. We also collected data on relationships among friends in social network services such as Facebook, Naver Band, and others.

The research data used in this study were collected through mobile and internet online surveys by building a customer rating data system. As shown in Fig. 2, there are many kinds of data was collected such as frequent visiting days, frequent visiting time, visiting purpose, taste of goods, price, atmosphere, convenience, satisfaction for restaurants, cafes in the university or nearby. Finally, total of 288 survey items were collected by defining 32 items with 9 preference evaluations per each item. In this study, only the overall satisfaction score was utilized.

The screen of the customer rating system on the mobile device is shown in Fig. 3. This screen provides images of demographic characteristics, 32 restaurants and cafes. The users evaluate the point-of-interest using score from 0 to 7.

Meanwhile, research data 2 is data used by Kim and Ahn (2010), and includes ratings data and social network information of real-world users about movie preference. Research data 2 includes 100 movies rating data for 90 users and includes friendship data from social network services such as Twitter, Facebook, MySpace, Cyworld, and others. See Kim and Ahn (2010) for a detailed description of Research Data 2.

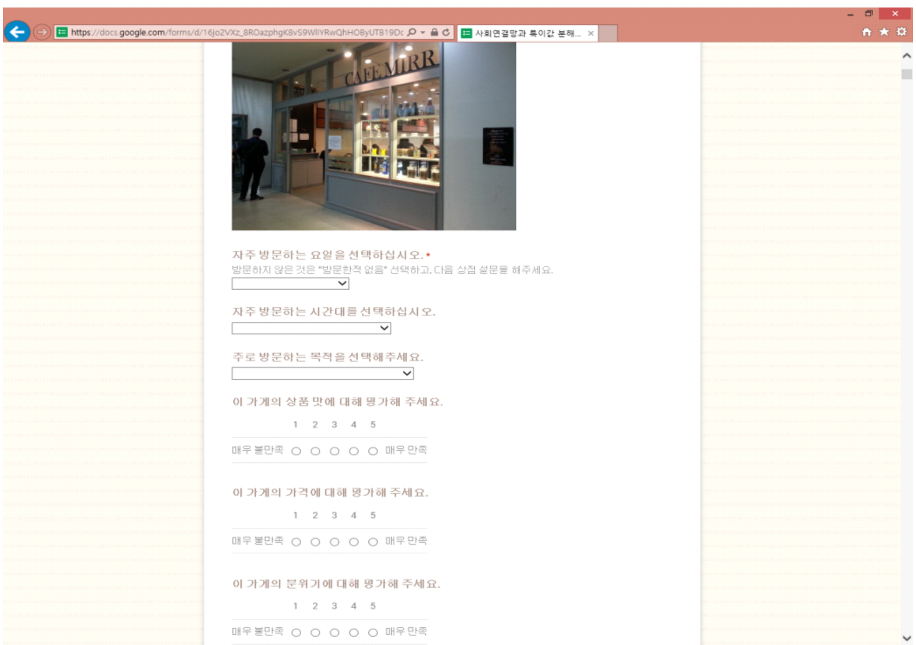

Fig. 2 Example screen of the data collection system

Fig. 3 Example screen of mobile survey system

4.2 Experiments and results

Experiments were carried out to predict the preference of the Social SVD++ CF and to confirm its usefulness. Experiment 1 sequentially conducted experiments for evaluating SVD, SVD++, and Social SVD++ prediction performance on the Research Data 1. Experiment 1 consisted of 482 training sets and 120 test sets with the ratio of 6: 4 data of 32 restaurants and cafes evaluated by 39 users in Research Data 1. Experiment 2 consisted of 2816 training sets and 1877 test sets with the ratio of 6: 4 data of 90 movies evaluated by 100 users in Research Data 2.

In this study, to compare SVD, SVD++, and Social SVD++, the difference between the predicted preference score and actual preference score is evaluated as MAE (Mean Absolute Error), and the difference is statistically significant (Paired Sample T-Test) was performed. The MAE is a statistical measure for evaluating the prediction accuracy of the recommended performance by comparing the score predicted from the analysis with the score evaluated by the actual user. It is frequently used for evaluating the prediction accuracy of CF-based recommender systems.

The results of the comparative analysis between the models are shown in Table 1. First, in case of Research Data 1, the MAE is the same when the performance of SVD and SVD++ is compared. When the SVD++ and Social SVD++ are compared, the MAE is lowered by 0.011. Although the performance evaluation does not show any significant difference, it has been shown that Social SVD++ outperforms the other two models.

In the result of the Research Data 2 based on the movie rating, the MAE value is lowered by 0.002 when the performance of SVD and SVD++ is compared, and the MAE is lowered by 0.001 when compared with Social SVD++. In comparison between SVD++ and SVD, MAE of SVD++ was improved only in Experiment 2.

Table 1. Results of comparative analysis (MAE)

Model	Research Data 1	Research Data 2
SVD	0.808	1.027
SVD++	0.808	1.025
Social SVD++	0.797	1.024

Table 2 and Table 3 show the results of paired sample t-test between SVD and SVD++, SVD and Social SVD++, and SVD++ and Social SVD++ to see whether the difference in MAE between the three models is statistically significant. As shown in Table 2, the difference in MAE between models was not statistically significant in Research Data 1.

Table 2. Results of statistical significance test for the Research Data 1 (p-value)

	SVD++	Social SVD++
SVD	0.481	0.535
SVD++		0.505

As shown in Table 3, SVD++ is different from SVD with statistical significance in Research Data 2 and significance is 1%. However, no statistical significance was found in comparison with Social SVD++.

Table 3. Results of statistical significance test for the Research Data 2 (p-value)

	SVD++	Social SVD++
SVD	0.000	0.438
SVD++		0.784

5. Conclusions

In this study, we tried to confirm the usefulness Social SVD++ CF model, which considers social network information of real-world users as implicit data by integrating typical collaborative filtering and extended SVD++ model. Social SVD++ CF has the advantage of reflecting the user's qualitative and emotional information in the recommendation process because it can reflect the user's social network information in the recommendation process. Also, considering the social network information and SVD, it is possible to mitigate the sparsity problem, and it will be helpful to improve the scalability which is another classical problem through the process of dimension reduction of SVD. Experiments on two real-world data have also been found to improve the performance of recommendation. The improvement of the recommendation performance is judged to be the result that the user's emotional and qualitative information is additionally considered in the recommendation process, while typical CF-based recommender systems utilize only the explicit evaluation information.

However, this study has various limitations.

First, we confirmed the improvement effect of the recommendation performance, but the improvement difference is not statistically significant and it is difficult to generalize the result. This is because the number of data used in the experiment is small and the current performance difference does not generalize the study results. Therefore, in future research, it is necessary to gather enough research data to generalize the research results.

Second, although the data used in this study is real-world data, it may be a problem in the reliability of the data because it is collected in a controlled environment. In order to solve this problem, it is necessary to acquire data from the service providers that links the actual recommendation system and the social network service, which should be supplemented in future studies.

Finally, the hypothesis that Social SVD++ CF will improve the classical problems of collaborative filtering, sparsity and scalability was not verified. In case of sparsity, it is possible to verify some degree through the result of recommendation performance, but it is difficult to confirm directly. In case of scalability, verification in real-world environment is needed. This point should be improved in future research.

References

Adomavicius, G. and A. Tuzhilin, “Toward the next generation of recommender systems : A survey of the state-of-the-art and possible extensions,” *IEEE Transactions on Knowledge and Data Engineering*, Vol.17, No.6(2005), 734~749.

Billsus, D. and M. J. Pazzani, “Learning Collaborative Information Filters,” *In Proceedings of Recommender Systems Workshop. Tech. Report WS-98-08, AAAI Press*, 1998, 46~54,

Cho, Y. and J. Bang, “Social network analysis for new product recommendation,” *Journal of Intelligence and Information Systems*, Vol.15, No.4(2009), 183~200.

Debnath, S., M. Granguly, and P. Mitra, “Feature weighting based recommendation system using social network analysis”, *Proceeding of WWW 2008*, 2008, 1041~1042.

Deerwester, S., S. RT. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, “Indexing by Latent Semantic Analysis,” *Journal of the American Society for Information Science*, Vol.41, 1990, 391~407.

Funakoshi, K. and T. Ohguri, “A Content-based collaborative recommender system with detailed use of evaluations,” *Proceedings of the 4th International conference on knowledge-based Intelligent Engineering systems and Allied Technologies*, 2000, 253~256.

Funk, S., “Netflix Update: Try This At home,” http://sifter.org/~simon/journal/20061211.html, 2006

Ghazanfar, M. A. and A. Prügel-Bennett, “Leveraging Clustering Approaches to Solve the Gray-Sheep Users Problem in Recommender Systems," *Expert Systems with Applications*, Vol.41, No.7(2014), 3261~3275.

Golbeck, J. and J. Hendler, "Filmtrust: Movie recommendations using trust in web-based social networks," *Proceedings of the IEEE Consumer Communications and Networking Conference*, 2006, 282~286.

Gorrell, G. and S. Funk, "Generalized hebbian algorithm for incremental latent semantic analysis," *Proceedings of Interspeech*, 2006, 1325~1328.

Jyun-Cheng, W., and C. Chui-Chen, "Recommending trusted online auction sellers using social network analysis", *Expert Systems with Applications*, Vol. 34, 2008, 1666~1679.

Kang, B. -S., "A novel web recommendation method for new customers using structural holes in social networks," *Journal of Industrial Economics and Business*, Vol.23, No.5(2010), 2371~2385.

Kang, B. -S., "Structural hole of social network analysis for recommending new items," *Journal of the Korean Data Analysis Society*, Vol. 13, No. 3(B), 1365~1377.

Kim, K. and H. C. Ahn, "Hybrid recommender systems using cluster-indexing collaborative filtering and social network analysis", *Proceedings of the 2010 Fall Conference of Korean Management Information Systems Society*, 2010, 604~609.

Kim, M. G. and K. Kim, "Recommender systems using structural hole and collaborative filtering," Journal of Intelligence and Intelligent Systems, Vol.20, No.4(2014), 107~120.

Kim, Y. S., H. C. Lee, and H. J. Chun, "Development of hybrid recommender systems in e-commerce using SVD", *Journal of Korea Society of Supply Chain Management*, Vol.2, No.1(2002), 21~30.

Koren, Y., "Factorization meets the neighborhood: a multifaceted Collaborative filtering model," *Proceedings of the 14th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*, 2008, 426~434.

Koren, Y. and R. Bell, "Advances in collaborative filtering," Recommender Systems Handbook, Edited by Ricci, F., L. Rokach, B. Shapira, and P. B. Kantor, Springer US, 2011 145~186.

Kumar, R., B. K. Verma, and S. S. Rastogi, "Social popularity based SVD++ recommender system," *International Journal of Computer Applications*, Vol.87, No.14(2014), 33~37.

Lee, H. D. and H. J. Kim, "Improving Collaborative Filtering with Rating Prediction Based on Taste Space", *Journal of KIISE: Database*, Vol.34, No.5(2007), 389~395.

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.

Park, J. H., Y. H. Cho, and J. K. Kim, "Social network: a novel approach to new customer recommendations", *Journal of Intelligence and Intelligent Systems*, Vol.15, No.1(2009), 123~140.

Paterek, A., "Improving regularized singular value decomposition for collaborative Filtering," *KDD-Cup and Workshop*, ACM press, 2007, 39~42,

Ricci, F., L. Rokach, B. Shapira, and P.B. Kantor, "Recommender Systems Handbook", Springer, 2011

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl, "Application of dimensionality reduction in recommender systems-a case study," In ACM WebKDD Workshop, 2000

Su, X. and T. M. Khoshgoftaar, "Collaborative Filtering for Multi-Class Data Using Bayesian Networks," International Journal on Artificial Intelligence Tools, Vol.17, No.1(2008), 71~85.

Wang, H. F. and C. T. Wu, "A Strategy-oriented Operation Module for Recommender Systems in E-commerce," Computers and Operations Research, Vol.39, No.10, 2012, 1837~1849.

Yang, W. and J. Dia, "Discovering cohesive subgroups from social networks for targeted advertising," Expert Systems with Applications, Vol. 34, 2008, 2029~2038.

국문요약

사회연결망정보를 고려하는 SVD 기반 추천시스템

김민건* · 김경재**

협업필터링은 사용자의 선호도 평가자료를 이용하여 특정 사용자의 특정 상품에 대한 선호도를 예측하고 이를 이용하여 유사한 사용자에게 상품을 추천한다. 협업필터링은 전자상거래에서의 정보 과잉현상을 줄여 주기에 가장 인기 있는 개인화 기법이다. 그러나 협업필터링은 희소성과 확장성 문제 등을 가지고 있다. 본 연구에서는 희소성과 확장성 문제와 같은 협업필터링의 주요 한계점을 보완하고 추천과정에 사용자의 정성적이고 감성적인 정보를 반영하도록 하기 위하여 사회연결망 정보와 협업필터링을 접목하는 방안을 이용한다. 본 논문에서는 특이값 분해에 내재적인 정보를 반영할 수 있도록 확장한 SVD++에 사회연결망 정보를 고려할 수 있도록 한 Social SVD++ 알고리듬을 협업필터링에 접목한 새로운 추천 알고리듬을 이용한다. 특히, 본 연구는 추천과정에 실제 사용자의 사회연결망 정보를 반영하여 모형의 성과를 평가할 것이다.

주제어 : 추천시스템, 사회연결망정보, 협업필터링, 특이값 분해, 비즈니스 애널리틱스

논문접수일 : 2016년 12월 5일 논문수정일 : 2016년 12월 19일 게재확정일 : 2016년 12월 21일
원고유형 : 일반논문 교신저자 : 김경재

* 유세스파트너스㈜
** 교신저자: 김경재
동국대학교_서울 경영대학 경영정보학과
30 Pildong-ro 1gil, Chung-gu, Seoul 100-715, Korea
Tel: 02-2260-3324, Fax: 02-2260-3684, E-mail: kjkim@dongguk.edu

저 자 소 개

김 민 건
현재 유세스파트너스(주)에서 전임연구원으로 재직 중이다. 동국대학교 컴퓨터멀티미디어학부에서 공학사를 받았으며, 동 대학원 경영정보학과에서 비즈니스인텔리전스 전공으로 석사학위를 취득하였다. 주 연구 관심분야는 고객관계관리, 사회연결망분석, 데이터마이닝, 소셜네트워크 서비스(SNS), 텍스트분석 등이며 지능정보연구에 논문을 게재하였다.

김 경 재
현재 동국대학교 경영대학 경영정보학과 교수로 재직 중이다. KAIST에서 경영정보시스템을 전공으로 박사학위를 취득하였으며, 경영과학, 경영학연구, 지능정보연구, 지식경영연구, Annals of Operations Research, Applied Intelligence, Applied Soft Computing, Asia Pacific Journal of Information Systems, Computers & Operations Research, Computers in Human Behavior, Expert Systems, Expert Systems with Applications, Information, Intelligent Data Analysis, International Journal of Electronic Commerce, Intelligent Systems in Accounting, Finance and Management, Journal of Information Technology Applications & Management, Neural Computing & Applications, Neurocomputing 등의 학술지에 논문을 게재하였다. 연구 관심분야는 고객관계관리, 데이터마이닝, 비즈니스 애널리틱스 등이다.