ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2020 June: 26(2): 43~56 http://dx.doi.org/10.13088/jiis.2020.26.2.043

A Topic Modeling-based Recommender System Considering Changes in User Preferences*

So Young Kang
Department of Business Administration, Graduate School, Kyung Hee Univesity
(*ilios9@nate.com*)

Jae Kyeong Kim
School of Management
Kyung Hee University
(*jaek@khu.ac.kr*)

Il Young Choi
School of Business Administration
Kyung Hee Univesity
(*choice102@khu.ac.kr*)

Chang Dong Kang
Department of Business Administration, Graduate School, Kyung Hee Univesity
(*cdkang@dooco.kr*)

Recommender systems help users make the best choice among various options. Especially, recommender systems play important roles in internet sites as digital information is generated innumerable every second. Many studies on recommender systems have focused on an accurate recommendation. However, there are some problems to overcome in order for the recommendation system to be commercially successful. First, there is a lack of transparency in the recommender system. That is, users cannot know why products are recommended. Second, the recommender system cannot immediately reflect changes in user preferences. That is, although the preference of the user's product changes over time, the recommender system must rebuild the model to reflect the user's preference. Therefore, in this study, we proposed a recommendation methodology using topic modeling and sequential association rule mining to solve these problems from review data. Product reviews provide useful information for recommendations because product reviews include not only rating of the product but also various contents such as user experiences and emotional state. So, reviews imply user preference for the product. So, topic modeling is useful for explaining why items are recommended to users. In addition, sequential association rule mining is useful for identifying changes in user preferences. The proposed methodology is largely divided into two phases. The first phase is to create user profile based on topic modeling. After extracting topics from user reviews on products, user profile on topics is created. The second phase is to recommend products using sequential rules that appear in buying behaviors of users as time passes. The buying behaviors are derived from a change in the topic of each user. A collaborative filtering-based recommendation system was developed as a benchmark system, and we compared the performance of the proposed methodology with that of the collaborative filtering-based recommendation system using Amazon's review dataset. As evaluation metrics, accuracy, recall, precision, and F1 were used. For topic modeling, collapsed Gibbs sampling was conducted. And we extracted 15 topics. Looking at the main topics, topic 1, top 3, topic 4, topic 7, topic 9, topic 13, topic 14 are related to “comedy shows”, “high-teen drama series”, “crime investigation drama”, “horror theme”, “British drama”, “medical drama”, “science fiction drama”, respectively. As a result of comparative analysis, the proposed methodology outperformed the collaborative filtering-based recommendation system. From the results, we found that the time just prior to the recommendation was very important for inferring changes in user preference. Therefore, the proposed methodology not only can secure the transparency of the recommender system but also can reflect the user's preferences that change over time.

* This research was conducted under the financial support from the National Research Foundation of Korea (NRF-2017S1A5B8059804).

However, the proposed methodology has some limitations. The proposed methodology cannot recommend product elaborately if the number of products included in the topic is large. In addition, the number of sequential patterns is small because the number of topics is too small. Therefore, future research needs to consider these limitations.

Key Words : Recommender system, Topic modeling, Sequential association rule mining, Sequence analysis, Topic pattern, Explainability

Received : April 1, 2020 Revised : May 7, 2020 Accepted : May 14, 2020
Publication Type : Regular Paper Corresponding Author : Chang Dong Kang

1. Introduction

With the development of information technology, as digital information is generated innumerable every second, the information available to consumers is getting rich. However, due to the exponential increase of information, users face difficulties in selecting products suitable for their preferences. In order to solve this problem, the importance of recommender systems is rapidly increasing. Especially, recommender systems play important roles in internet sites such as Amazon, YouTube, Netflix so on (Koren et al., 2009) because recommender systems provide suggestions for users especially to individuals who lack experience or competence to evaluate the alternative products (Mahmood et al., 2009).

Many studies on recommender systems have focused on accurate recommendation (Ahn et al., 2006; Cho and Kim, 2004; Cho et al., 2002; Herlocker et al., 2000; Kim et al., 2009; Kim et al., 2010; Kim et al., 2018; Lee and Park, 2007; Shardanand and Maes, 1995; Sohn and Suh, 2006; Suh et al., 2014). However, recommender systems should be transparent (Dehuri, 2012). The transparency of recommender systems means their abilities to explain why the products were recommended. although recommender systems have become more accurate and more elaborate as technology evolves, users cannot understand how the systems work. Furthermore, recommender systems should consider changes in user preferences (Cho et al., 2005) because user preference change over time. Accordingly, some researches have been attempted to solve these problems. Abdollahi and Nasraoui (2016), Bellini et al. (2018), Cho et al. (2006), He et al. (2005) and Ren et al. (2017) have proposed the explainable recommender systems. And, Cho et al. (2005), Kim et al. (2018), Moon et al. (2010), and Moon et al. (2017) have attempted to develop the recommender system considering preference change. However, existing studies have limitations in solving only one of the transparency and preference change.

In this study, we propose a recommendation methodology for addressing these problems. The proposed methodology uses the review data. Online reviews such as blogs and websites influence users' decision-making (Filieri and

McLeay, 2013; Gupta and Harris, 2009). Especially, product reviews provide useful information for recommendations because product reviews include not only rating of the product but also various contents such as user experiences and emotional state. So, reviews imply user preference for the product. The proposed methodology employs topic modeling and sequential association rule mining. Topic modeling in product reviews is used for explainable recommendation because topic modeling estimates the distribution of topics per review (Pyo and Kim, 2014). This means that the topics are clustered with the same reviews. and, sequential association rule mining is used for reflect changes in user preferences. Sequential association rule mining in recommender systems is useful for identifying patterns of preference changes (Cho et al., 2005; Moon et al., 2010; Moon et al., 2013).

2. Related works

Recommendation systems help users find products that suit their preferences. In general, recommender systems are widely classified into content-based filtering and collaborative filtering. The former creates a profile for each user or product to characterize its natures (Pazzani and Billsus, 2007). In other words, the latter analyzes the relationship between the user. Both methods have pros and cons. The former is better in cold start problem, where a recommender system is not able to recommend existing products to new users or new products to existing users, than the latter because the former is based on their products and user description. But the latter is more diverse than the former because the latter recommends the purchased products of neighborhoods who have similar preferences to a target user.

To overcome various problems such as sparsity, cold-start problem, lack of explainability, and use of static preference, text mining and data mining are used in the recommendation system. A typical technique among text mining is topic modeling. Topic modeling is “latent variable models of documents that exploit the correlations among the words and latent semantic themes” (Blei and Lafferty, 2007). That is, the main purpose of topic modeling is to find hidden pattern as topics in documents. The most popular topic-modeling technique is Latent Dirichlet Allocation (LDA) which is a probabilistic model of a corpus. User preferences can be inferred through LDA (Pyo and Kim, 2014). Therefore, topic modeling is used in the recommendation. Choi et al. (2015) developed a model to predict user’s repurchase using topic modeling, and Jin et al. (2018) proposed a recommendation methodology combining deep learning and topic modeling.

A representative data mining technique that can infer preference changes is sequential association rule mining. Sequential association rule mining is one of the unsupervised learning techniques that find frequent patterns and correlations considering time. Sequential association rule is presented as $A \Rightarrow B$. It means that users who bought A will buy B after a certain period of time. However, it is important to find useful rules because many rules are generated through sequential association rule mining. In general, minimum support and minimum confidence are used to find useful rules. Support and confidence are computed as follows. Let sequential association rule be $A \Rightarrow B$.

$$\text{Support}\ (A \Rightarrow B) = P(A \cup B)\ and\ \text{Confidence}\ (A \Rightarrow B) = \text{P}(B \mid A),$$

here, P is probability.

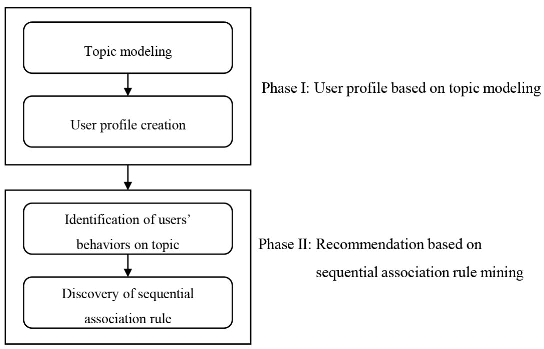
Cho et al. (2005) used the self-organizing map to cluster user preference and then discovered sequential patterns between clusters. Moon et al. (2010) used sequential association rule mining to discover sequential patterns of preferences in an exhibition environment. And Kim et al. (2018) combined sequential association rule mining and association rule mining to infer user's preference in grocery shopping. So, in this study, we propose a recommendation methodology using topic modeling and sequential association rule mining.

3. Proposed methodology

3.1 Overall view

We propose a recommendation methodology using topic-modeling to explain why products are recommended and sequential association rule mining to reflect changes in user preferences. In other words, the proposed methodology employs topic-modeling and sequential association rule mining. The former is useful to find users' preferences (Ovsjanikov and Chen, 2010) and the latter is useful to discover the changes of preferences (Cho et al., 2005). In this study, we assume that if the user has written a review of a product, that product has been purchased. In other words, we assume that users write reviews on purchased products.

The proposed methodology consists of two phases as shown in <Figure 1>. The first phase is to create user profile based on topic modeling. After extracting topics from user reviews on products, a user profile on topics is created. The second phase is to recommend products using sequential rules that appear in buying behaviors of users as time passes. The buying behaviors are derived from a change in the topic of each user.



〈Figure 1〉 Overall procedure

3.2 User profile based on topic modeling

To cluster users' preferences, topic modeling is conducted because topics represent users' preferences (Ovsjanikov and Chen, 2010). In this study, Latent Dirichlet Allocation (LDA) is used for topic modeling. To extract topics from user reviews, collapsed Gibbs sampling for LDA is used. Collapsed Gibbs sampling is a technique to remove some unnecessary variables from sampling. When performing collapsed Gibbs sampling, the number of topics should be determined. We use the harmonic mean of the log likelihood to determine the optimal number of topics.

LDA estimates the distribution of words per topic and the distribution of topics per review. Therefore, users' reviews have more than one topic. Suppose the purchase history is shown in <Table 1>. When topic modeling is performed on product review, each product can be expressed as a probability to be included in each topic as shown in <Table 2>.

〈Table 1〉 Example of purchase history

User	Product	Time	Review
$User_1$	P_1	20200225	My understanding is that most of the non-USA versions can handle 2 SIM cards and come with a tray that can hold 2 SIM cards.
$User_1$	P_2	20200301	For all those waiting on Samsung's promised update, hoping it'll fix the issue
$User_1$	P_3	20200305	This phone is awesome it took my a little bit to get use to some of the stuff on here since I was use to the a50 but I'm almost a pro now.
$User_2$	P_1	20200122	Didn't come with the offered Buds and duo charger. What's up!!
$User_2$	P_2	20200205	It did not come with complete items
$User_2$	P_3	20200307	Very good product, go for the plus s20 is almost same looking and size as the s9 & s10

〈Table 2〉 Example of purchase history rearranged by topic

User	Product	Time	Probability				
			Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
$User_1$	P_1	20200225	0.69	0	0	0	0
$User_1$	P_2	20200301	0	0.34	0	0	0
$User_1$	P_3	20200305	0	0	0.30	0	0
$User_2$	P_1	20200122	0	0	0	0.36	0
$User_2$	P_2	20200205	0	0.17	0	0	0.53
$User_2$	P_3	20200307	0.2	0	0	0	0

〈Table 3〉 Example of user profile on topic

User	Product	Time	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
$User_1$	P_1	t-2	1	0	0	0	0
$User_1$	P_2	t-1	0	1	0	0	0
$User_1$	P_3	t	0	0	1	0	0
$User_2$	P_1	t-2	0	0	0	1	0
$User_2$	P_2	t-1	0	1	0	0	1
$User_2$	P_3	t	1	0	0	0	0

Therefore, the user profile is represented as a collection of m user' preference on n topics at t time. Let $r_{i,j}^{t}$ be user i' profile on topic j at time t. $r_{i,j}^{t}$ is represented as follows.

$$r_{i,j}^{t} = \begin{cases} 1 & \textit{if user i's review is included in topic j at time t} \\ 0, & \textit{otherwise} \end{cases}$$

where i = 1 to m, and j = 1 to n

For example, <Table 2> is converted to a user profile as shown in <Table 3>.

3.3 Recommendation using Sequential Association Rule Mining

User preferences for items change from time to time. If the user's preference can be tracked through pattern analysis, a product that fits the user's preference can be recommended. This phase is divided into two steps as follows.

The first step is to identify users' behaviors on topic. Suppose there are q topics. Then $T = \{T_1, T_2, \cdots, T_q\}$, where T means topic. The behavior of user i at time t represents as a topic. If L_i is the behavior locus of user i for k periods at time t, L_i represent as follows.

$$L_i = <T_{i,t-k}, \cdots, T_{i,t-1}, T_{i,t}>$$

where i = 1 to m, $T_{i,t-k} \in$ T, k = 0 to t-1, and $t \geq 1$

For example, <Table 3> is converted to a user profile as shown in <Table 4>.

〈Table 4〉 Example of behavior locus of user by time

User \ Time	t-2	t-1	t
$User_1$	T1	T2	T3
$User_2$	T4	T2	T1
		T6	T1

The second step is to find a topic of a target user at time t through sequential association rule mining. In this study, we use sequential association rule mining proposed by Cho et al. (2005). That is, we find the sequential patterns ($A \Rightarrow B$) that satisfy minimum support and minimum confidence. Here, A and B are topic locus at time t-1 and t, respectively. So, if the target user exhibits the behavior of A at time t-1, it is predictable to be in the topic locus of B at time t. Therefore, we recommend products include in each topic according to the meaningful sequential association rules.

4. experiments and evaluations

4.1 Dataset

Amazon review data is used to evaluate the performance of the proposed methodology. The datasets are composed of the real-world reviews of 37,127 users who bought 1,685 amazon instant video from January 2007 to December 2013. We deleted the data before 2010 because it was too old. Also, we targeted only users who left reviews of more than 5 products. We used “reviewer ID”, “asin”, “review text”, and “review time” from the sample data in <Figure 2>. Here, “asin” means video id.

4.2 Evaluation measures

To evaluate recommender systems, various metrics such as mean absolute error (*MAE*), root mean square error (*RMSE*), accuracy, error rate, sensitivity, recall, precision, or $F1$ have been used (Good et al., 1999; Herlocker et al., 2000; Herlocker et al., 2004; Kim et al., 2009; Sarwar et al., 2000). In this study, we used accuracy, recall, precision, and F1.

Suppose that the confusion table is as shown in <Table 5>. Accuracy, recall, precision, and F1 are computed as follows.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F1} = \frac{2 \text{ X } Recall \text{ X } Precision}{Recall + \ Precision}$$

“reviewerID": "A2SUAM1J3GNN3B",
"asin": "0000013714",
"reviewerName": "J. McDonald",
"helpful": [2, 3],
"reviewText": "I bought this for my husband who plays the piano. He is having a wonderful time playing these old hymns. The music is at times hard to read because we think the book was published for singing from more than playing from. Great purchase though!",
"overall": 5.0,
"summary": "Heavenly Highway Hymns",
"unixReviewTime": 1252800000,
"reviewTime": "09 13, 2009"

〈Figure 2〉 Data sample

〈Table 5〉 Confusion matrix

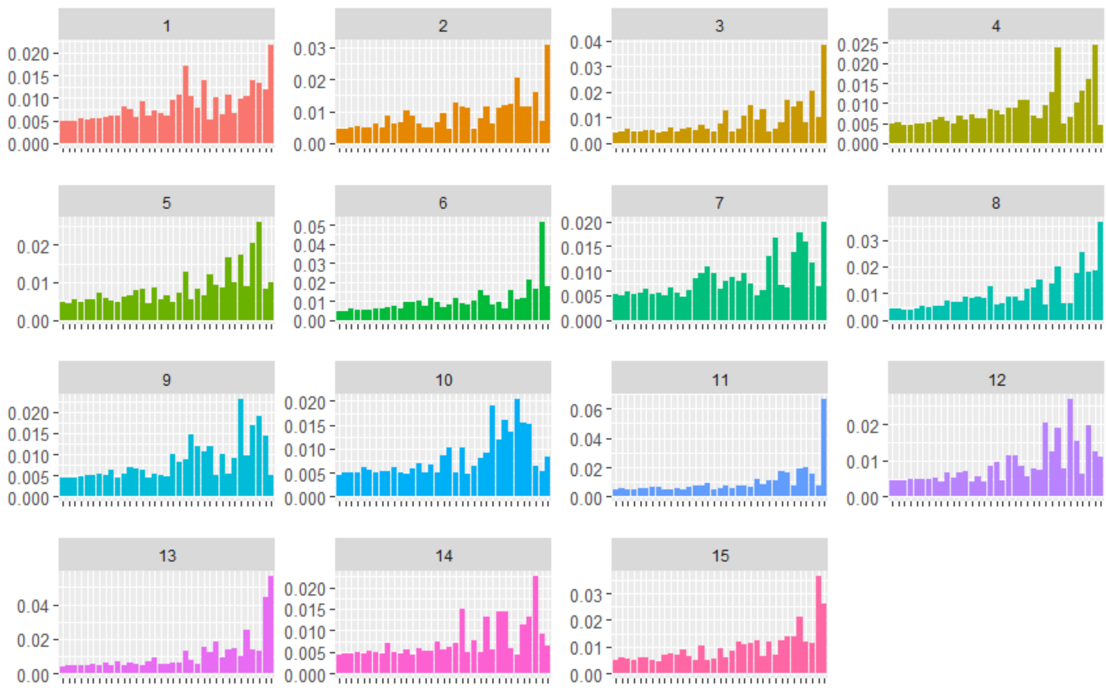
		Predicted class	
		True	False
Actual class	True	True positive (TP)	False negative (FN)
	False	False positive (FP)	True negative (TN)

To evaluate the proposed methodology, we developed model 1 and model 2 considering the behavior locus over time as shown in <Table 4>. Model 1 is a sequence model to recommend topic at time t considering the behavior locus from time t-2 to time t-1, and Model 2 is the sequence model to recommend topic at time t considering the behavior of only time t-1. Also, we developed collaborative filtering-based recommender system (CF), which does not consider the behavior locus but only the user preference between time t-1 and time t, as a benchmark system because CF is known as commercially successful recommender systems that used in Amazon (Linden et al., 2003), Google (Das et al., 2007), Netflix (Bennett and Lanning, 2007), and so on.

4.3 Experimental results and discussion

4.3.1 Topic modeling

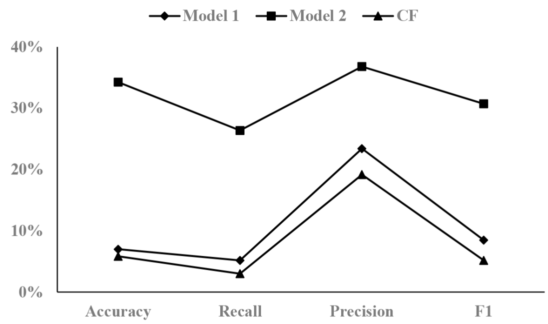
To cluster user preference, topic modeling through collapsed Gibbs sampling was conducted. The number of topics was selected from the maximum value of the harmonic mean of the



〈Figure 3〉 Distribution of the most frequently used words within each topic log-likelihood value (Ponweiser, 2012). We extracted 15 topics. The distribution of the most frequently used words in each topic is as shown in <Figure 3>. The main topics are as follows. First, topic 1 is related to "comedy shows" because topic 1 has a high frequency of words such as comedy, humor, and laugh. Second, topic 3 is related to "high-teen drama series" because topics 3 mainly includes words such as glee and student. Third, topic 4 is related to "crime investigation drama" because we can see the word such as police, agent, and killer. Fourth, topic 7 is related to "horror theme" because the frequency of words such as horror, paranormal, and scary is high. Fifth, topic 9 is related to "British drama" because we can see words such as detective, holmes, and British. Sixth, topic 13 is related to "medical drama" because we can see the word such as doctor and season. Finally, Topic 14 is related to "science fiction drama" because words such as aliens, star, and trek appear.

4.3.2 Recommendation evaluation

Our experiments were performed to compare the accuracy, recall, precision, and F1 of the proposed methodology with those of the benchmark system. The experimental result is presented in <Figure 3>. First, the values of accuracy, recall, precision, and F1 on model 1 are 6.98%, 5.17%, 23.40%, and 8.47%, respectively. Second, the values of accuracy, recall, precision, and F1 on model 2 are 34.23%, 26.35%, 36.78%, and 30.71%, respectively. Finally, the values of accuracy, recall, precision, and F1 on the benchmark system are 5.83%, 2.98%, 19.16%, 5.15%, respectively. From these results, we found that the proposed model 2 is better than the model 1 and the benchmark system. So, we can judge that the time just prior to the recommendation is very important for inferring changes in user preference.



〈Figure 4〉 Comparative analysis of recommended performances

5. Conclusion

One of the functions required in the recommendation system is a function that can explain to users why these products are recommended. Also, the recommender systems should be able to reflect user preferences that change over time. Therefore, we proposed a recommendation methodology using topic-modeling and sequential association rule mining. The experimental results show that the performance of the proposed methodology is better than that of the CF-based recommender system.

The contributions of the proposed methodology are below. First, the proposed methodology can explain why the products are recommended to the user because topics derived through topic modeling represent user preferences. Second, the proposed methodology can catch the preference changes of the users because this methodology can find the locus of changes in user preferences through sequential association rule mining. Therefore, corporates will be able to advertise products and sell more according to the user's preference change, and users can get more satisfaction if corporates employ the proposed methodology. That is, the proposed methodology will be beneficial to both corporates and users.

However, the proposed methodology has some limitations. First, the proposed methodology does not recommend products directly but recommends products related to the topic. However, if there are too many products related to the topic, the sophistication of the recommendation may be deteriorated. Second, the number of sequential patterns is small because the number of topics is too small. Third, the proposed methodology cannot be applied if there is only a rating. That is, the methodology can be used only with review data. Finally, Lastly, if the review is not long enough, there is a possibility that topic extraction may not work properly. Therefore, future research needs to consider these limitations.

References

Abdollahi, B. and O. Nasraoui, "Explainable Matrix Factorization for Collaborative Filtering", In *Proceedings of the 25th International Conference Companion on World Wide Web*, (2016), 5~6.

Ahn, H. C., I. G. Han and K. J. Kim, "The Product Recommender System Combining Association Rules and Classification Models: The Case of G Internet Shopping Mall", *Information Systems Review*, Vol.8, No.1 (2006), 181~201.

Bellini, V., A. Schiavone, T. Di Noia, A. Ragone and E. Di Sciascio, "Knowledge-Aware Autoencoders for Explainable Recommender Systems", In *Proceedings of the 3rd Workshop on Deep Learning for Recommender Systems,* (2018), 24~31.

Bennett, J. and S. Lanning, "The Netflix Prize", In *Proceedings of KDD cup and workshop,* Vol.2007, (2007), 35.

Cho, Y. B., Y. H. Cho and S. H. Kim, "Mining Changes in Customer Buying Behavior for Collaborative Recommendations", *Expert Systems with Applications*, Vol.*28*, No.2 (2005), 359-369.

Cho, Y. H., and J. K. Kim, "Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-commerce", *Expert systems with Applications*, Vol.26, No.2 (2004), 233~246.

Cho, Y. H., J. K. Kim and S. H. Kim, "A Personalized Recommender System based on Web Usage Mining and Decision Tree Induction", *Expert systems with Applications*, Vol.23, No.3 (2002), 329~342.

Cho, Y. H., J. K. Kim, D. H. Ahn and H. A. Lee, "An Explanation based Recommender System using Collaborative Filtering: WebCF-Exp", *Korean management Review,* Vol.35, No.2 (2006), 493~519.

Choi, S., Y. Hyun and N. Kim, "Improving Performance of Recommendation Systems using Topic Modeling", *Journal of Intelligence and Information Systems*, Vol.*21*, No.3 (2015), 101~116.

Das, A. S., M. Datar, A. Garg and S. Rajaram, "Google News Personalization: Scalable Online Collaborative Filtering", In *Proceedings of the 16th international conference on World Wide Web,* (2007)*,* 271~280.

Dehuri, S., *Intelligent Techniques in Recommendation Systems: Contextual Advancements and New Methods: Contextual Advancements and New Methods*, IGI Global, 2012.

Filieri, R. and F. McLeay, "E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews", *Journal of Travel Research*, Vol.53, No.1 (2014), 44~57.

Good, N., J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker and J. Riedl, "Combining Collaborative Filtering with Personal Agents for Better Recommendations", *AAAI/IAAI*, (1999), 439.

Gupta, P. and J. Harris, "How E-WOM Recommendations Influence Product Consideration and Quality of Choice: A Motivation to Process Information Perspective", *Journal of Business Research*, Vo.63, No.9-10 (2010), 1041~1049.

He, X., T. Chen, M. Y. Kan and X. Chen, "Trirank: Review-Aware Explainable Recommendation by Modeling Aspects", In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management,* (2015), 1661-1670.

Herlocker, J. L., J. A. Konstan and J. T. Riedl, "Explaining Collaborative Filtering Recommendations", In *Proceedings of the 2000 ACM conference on Computer supported cooperative work,* (2000), 241~250.

Herlocker, J. L., J. A. Konstan, L. G. Terveen and J. T. Riedl, "Evaluating Collaborative Filtering Recommender Systems", *ACM Transactions on Information Systems (TOIS)*, Vol.22, No.1(2004), 5~53.

Blei, D. M. and J. D. Lafferty, "A Correlated Topic Model of Science", *The Annals of Applied Statistics,* Vol.1, No.1 (2007), 17~35.

Jin, M., Luo, X., Zhu, H., and Zhuo, H. H., "Combining Deep Learning and Topic Modeling for Review Understanding in Context-aware Recommendation," In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol 1 (2018), 1605~1614.

Kim, H. K., J. K. Kim and Y. U. Ryu, "Personalized Recommendation over a Customer Network for Ubiquitous Shopping", *IEEE Transactions on Services Computing*, Vol.2, No.2 (2009), 140~151.

Kim, H. N., A. T. Ji, I. Ha and G. S. Jo, "Collaborative Filtering based on Collaborative Tagging for Enhancing the Quality of Recommendation", *Electronic Commerce Research and Applications*, Vol.9, No.1 (2010), 73~83.

Kim, J. K., H. S. Moon, B. J. An and I. Y. Choi, "A Grocery Recommendation for Off-line Shoppers", *Online Information Review*, Vol.42, No.4 (2018), 468~481.

Lee, J. S. and S. D. Park, "Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering", *Journal of Intelligence and Information Systems*, Vol.13, No.4 (2007), 65~78.

Moon, H. S., J. K. Kim and Y. U. Ryu, "A Sequence-based Filtering Method for Exhibition Booth Visit Recommendations", *International journal of information management*, Vol.33, No.4 (2013), 620~626.

Moon, H. S., M. K. Jung, J. K. Kim and H. K. Kim, "Personalized Exhibition Booth Recommendation Methodology Using Sequential Association Rule", *Journal of Intelligence and Information Systems*, Vol.16, No.4 (2010), 195~211.

Ovsjanikov, M. and Y. Chen, "Topic Modeling for Personalized Recommendation of Volatile Items", In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Berlin, Heidelberg, (2010), 483~498.

Pazzani, M. J. and D. Billsus, "Content-based Recommendation Systems", In *The adaptive web*, Springer, Berlin, Heidelberg, (2007), 325~341.

Ponweiser, M. *Latent Dirichlet Allocation in R (Diploma thesis)*. Institute for Statistics and Mathematics, WU (Wirtschaftsuniversität Wien), Austria, 2012.

Pyo, S. and E. Kim, "LDA-based Unified Topic Modeling for Similar TV User Grouping and TV Program Recommendation", *IEEE transactions on cybernetics*, Vol.45, No.8 (2014), 1476~1490.

Ren, Z., S. Liang, P. Li, S. Wang and M. de Rijke, "Social Collaborative Viewpoint Regression with Explainable Recommendations", In *Proceedings of the tenth ACM international conference on web search and data mining,* (2017), 485~494.

Sarwar, B., G. Karypis, J. A. Konstan and J. T. Riedl, "Analysis of Recommendation Algorithms for E-commerce", In *Proceedings of the 2nd ACM conference on Electronic commerce*, (2000), 158~167.

Shardanand, U. and P. Maes, "Social Information Filtering: Algorithms for Automating "Word of Mouth"", In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (1995), 210~217.

Linden, G., B. Smith and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering", *IEEE Internet computing*, Vol.7, No.1 (2003), 76~80.

Sohn, J. B. and Y. M. Suh, "Using Degree of Match to Improve Prediction Quality in Collaborative Filtering Systems", *Information Systems Review*, Vol.8, No.2 (2006), 139~154.

Suh, K. S., S. Lee, E. K. Suh, H. Kang, S. Lee and U. K. Lee, "Comparisons of Popularity- and Expert-Based News Recommendations: Similarities and Importance", *Asia pacific journal of information systems*, Vol.24, No.2 (2014), 191-210.

국문요약

고객 선호 변화를 고려한 토픽 모델링 기반 추천 시스템

강소영* · 김재경** · 최일영*** · 강창동****

추천 시스템은 사용자가 다양한 옵션 중에서 최선의 선택을 할 수 있도록 도와준다. 그러나 추천 시스템이 상업적으로 성공하기 위해서는 극복할 몇 개의 문제점이 존재한다. 첫째, 추천시스템의 투명성 부족 문제이다. 즉, 추천된 상품이 왜 추천되었는지 사용자들이 알 수 없다. 둘째, 추천시스템이 사용자 선호의 변화를 즉각적으로 반영할 수 없는 문제이다. 즉, 사용자의 상품에 대한 선호는 시간이 지남에 따라 변함에도 불구하고, 추천시스템이 사용자 선호를 반영하기 위해서는 다시 모델을 재구축해야 한다. 따라서 본연구에서는 이러한 문제를 해결하기 위해 토픽 모델링과 순차 연관 규칙을 이용한 추천 방법론을 제안하였다. 토픽 모델링은 사용자에게 아이템이 왜 추천되었는지 설명하는데 유용하며, 순차 연관 규칙은 변화하는 사용자의 선호를 파악하는데 유용하다. 본 연구에서 제안한 방법은 크게 토픽 모델링 및 사용자 프로파일 생성 등 토픽 모델링에 기반한 사용자 프로파일 생성 단계와 토픽에 사용자 선호 확인 및 순차 연관 규칙 발견 등 순차 연관 규칙에 기반한 추천 단계로 구분된다. 벤치마크 시스템으로 협업 필터링 기반 추천 시스템을 개발하고, 아마존의 리뷰 데이터 셋을 이용하여 제안한 방법론의 성능을 비교 평가하였다. 비교 분석 결과, 제안한 방법론이 협업 필터링 기반 추천시스템보다 뛰어난 성능을 보였다. 따라서 본 연구에서 제안하는 추천 방법을 통해 추천 시스템의 투명성을 확보할 수 있을 뿐만 아니라, 시간에 따라 변화하는 사용자의 선호를 반영할 수 있다. 그러나 본 연구는 토픽과 관련된 상품을 추천하기 때문에, 토픽에 포함된 상품의 수가 많을 경우 추천이 정교하지 못하는 한계점이 있다. 또한 토픽의 수가 적기 때문에 토픽에 대한 순차 연관 규칙이 너무 적은 문제점이 있다. 향후 연구에서 이러한 문제점을 해결한다면 좋은 연구가 될 것으로 판단된다.

주제어 : 추천 시스템, 토픽 모델링, 순차 연관 규칙, 순차 분석, 토픽 패턴, 설명가능성

논문접수일 : 2020년 4월 1일 논문수정일 : 2020년 5월 7일 게재확정일 : 2020년 5월 14일
원고유형 : 일반논문 교신저자 : 강창동

* 경희대학교 대학원 경영학과
** 경희대학교 경영학과
*** 경희대학교 경영대학원
**** 교신저자 : 강창동
경희대학교 대학원 경영학과

저 자 소 개

So Young Kang

So Young Kang obtained her M.S. in the Big Data Management and her B.S. in Business Administration from Kyung Hee University. Her current research interests include data mining, recommender system, and deep learning.

Il Young Choi

Il Young Choi is a lecturer at Graduate School of Business Administration, Kyung Hee University. He obtained his MS and Ph.D. in MIS, and his B.S. in Economics from Kyung Hee University. His current research interests focus on personalized service, bigdata analysis, social network analysis, and business analytics. He has published numerous papers which have appeared in Information Technology and Management, International Journal of Information Management, Online Information Review, and so on.

Jae Kyeong Kim

Jae Kyeong Kim is a professor at School of Management, Kyung Hee University. He obtained his M.S. and Ph.D. in Management Information Systems (MIS) from KAIST (Korea Advanced Institute of Science and Technology), and his B.S. in Industrial Engineering from Seoul National University. His current research interests focus on business analytics, recommender systems, big data analysis, and deep learning. He has published numerous papers which have appeared in IEEE transactions on Systems, Men and Cybernetics, IEEE transactions on Services Computing, Information & Management, International Journal of Information Management, Applied Artificial Intelligence, Artificial Intelligence Review, International Journal of Human-Computer Studies, Expert Systems with Applications, and Group Decision and Negotiations.

Chang Dong Kang

Chang Dong Kang is a CEO at dooco, Inc.. He obtained his B.S. in business management at Kyonggi University and his M.S. in business management at Kyung Hee University. His current research interests focus on big data analytics, recommendation system, business intelligence, SCM, and car dash cam.