

Movie Recommendation System Based on Users' Personal Information and Movies Rated Using the Method of k -Clique and Normalized Discounted Cumulative Gain

Phonexay Vilakone*, Khamphaphone Xinchang*, and Doo-Soon Park**

Abstract

This study proposed the movie recommendation system based on the user's personal information and movies rated using the method of k -clique and normalized discounted cumulative gain. The main idea is to solve the problem of cold-start and to increase the accuracy in the recommendation system further instead of using the basic technique that is commonly based on the behavior information of the users or based on the best-selling product. The personal information of the users and their relationship in the social network will divide into the various community with the help of the k -clique method. Later, the ranking measure method that is widely used in the searching engine will be used to check the top ranking movie and then recommend it to the new users. We strongly believe that this idea will prove to be significant and meaningful in predicting demand for new users. Ultimately, the result of the experiment in this paper serves as a guarantee that the proposed method offers substantial finding in raw data sets by increasing accuracy to 87.28% compared to the three most successful methods used in this experiment, and that it can solve the problem of cold-start.

Keywords

Association Rule Mining, k -Cliques, Normalized Discounted Cumulative Gain, Recommendation System

1. Introduction

Today, the cyber is growing very fast, and we can feel that everything tends to be based on cyber; everywhere is also cyber [1-3]. Cyber is providing a lot of benefits to people, and the best chance in selling or buying products online, re-meet a childhood friend or meet new friends on the social network like Facebook, Instagram, Twitter, etc. Inside, cyber technology is the recommendation system. The recommendation system becomes the most important technology that helps business owners gain a huge profit from their business over the Internet or social network because this technology can predict what kind of product the users are looking for over the Internet. Almost all methods used in the product recommendation system will utilize the users' behavior information, recommended from the best-selling product or from the same type of product that is attractive to the user.

Nonetheless, these methods suffer from problems of data sparsity, scalability, cold-start, user privacy, etc. As the ordinary routine in the social network system, all users need to provide some necessary

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received April 1, 2019; first revision July 10, 2019; second revision December 26, 2019; accepted January 14, 2020.

Corresponding Author: Doo-Soon Park (parkds@sch.ac.kr)

* Dept. of Computer Sciences and Engineering, Soonchunhyang University, Asan, Korea (xayus@yahoo.com, khamphaphone12@gmail.com)

** Dept. of Computer Software Engineering, Soonchunhyang University, Asan, Korea (parkds@sch.ac.kr)

personal information before they can access any operation [4], and one more extreme trend appearing in the social network is that the number of users is rising rapidly in day by day. For these reasons, many researchers are paying attention to developing various new methods, such as how to classify users into several communities by using the users' personal information or create a technique that can pick up the community by using their connection. One of those methodologies is the k -clique method, which is typically used to analyze data of the users in the vast social network and classify them into various groups depending on their personal information or their connection [5-8]. In the past few years, a researcher used the k -clique method as an algorithm model in the recommendation system, and another researcher used a ranking measure method in the recommendation system, too. This method is used to calculate the rank of the products after recommending to the users those that rank high. The result of both techniques gives a satisfying outcome.

Therefore, to increase the accuracy further compared to the recent result in the recommendation system, avoid the problem of cold-start, and create a new technique instead of using the standard technology in the recommendation system, the movie recommendation system based on the user's personal information and movies rated using the method of k -clique and normalized discounted cumulative gain will be examined in this paper.

The rest of this paper is organized as follows: Section 1 presents the introduction particularly the background and objective of this paper; Section 2 briefly describes the related work done and defines the terms used in this paper; Section 3 is the proposed method architecture; this part explains the more in-depth detail of work in this paper; Section 4 is the experimental analysis, showing the result of the experiments and the comparison results between the proposed methods and existing methods used in the recommendation system; Finally, Section 5 presents the conclusion and future works.

1.1 Problem Statement and Our Contributions

One of the significant challenges in the performance of the recommendation system is how to recommend a product to a new user; there is very little information about the user that is available, and no ratings are usually available for a new item. For that, the new users will need to rate a sufficient number of things to enable the system to capture their preferences accurately and provide reliable recommendations accordingly. Similarly, when new items are added to the system, they need to be rated by a substantial number of users before they could be recommended to users with whose tastes are similar to those of the ones who rated them.

The main contribution of this paper is summarized as follows:

(1) We proposed a movie recommendation system based on the users' personal information and movies rated using the method of k -clique and normalized discounted cumulative gain. In the proposed approach, the personal data of the users are used for measuring the similarity among them by using the method of cosine similarity measure. After measuring the similarity among the users and completing the result, it will be converted into a relationship table, which in turn will be converted into a network graph later. After that, the approach of the k -clique is used to pick up the group of users from the network graph as the conversion from the relations table. In this process, various groups of users will appear. Later, the new users will check for a suitable group to join by using the approach of the cosine measure to measure the personal information of the new users with the personal data of the users in each group. In some cases, a new user can belong in more than one group, which depends on the value of the similarity measuring

the result. After the new user determines which suitable group to join, the list of movies that got a high-ranking measurement score from the appropriate group, calculated by the normalized discounted cumulative gain, will appear. Finally, the data mining method like association rule mining used for checking the movie that is always watched by the users in a suitable group will recommend the top movies from this process to the new users.

(2) To assess the performance, the proposed approach is compared to the three most successful methods used in the experiment in this study: the approach of the k -clique techniques, the approach of collaborative filtering using the k nearest neighbor combined with the normalized discounted cumulative gain method (CF- k NN+NDCG), and the k -clique with the association rule mining to evaluate the result of the experiment. The process of experimenting with the k -clique methods and k -clique with the association rule mining method is similar to the operation of the proposed method, but there is a difference in the process of listing the recommended movie to the new user. In the k -clique techniques, the process of listing the recommended movie for the users uses the top N method of finding the most popular movie from movies archived from a suitable group. For the k -clique with the association rule mining methods, the process of listing the recommended movies for the new users uses a data mining method like association rule mining to find the movie that is always watched by the users in a suitable group. The workflow of CF- k NN+NDCG processed as the k NN method is used for classifying users into several groups by using the users' personal information. After that, the new user is assigned to a suitable group with the help of the cosine similarity method. Later, the normalized discounted cumulative gain method helps find the movie from a suitable group as the movie to be recommended to the new user. The result of the comparison is very satisfying because it shows that the proposed method result gives more accuracy compared to the three methods.

2. Related Work

To the best of the author's knowledge, there has been some work on classifying users into several communities or picking up the group from the social network or the network graph, but there has been limited work related to the ranking measure used in the recommendation system. Despite that, there is a particular solution that can be used for such requirement, as discussed in Table 1.

Vilakone et al. [9] introduced an improved k -clique as an efficient method in the recommender system. This new algorithm improved the accuracy of prediction, which is dependent on the value of k in the k -clique. Hao et al. [7] presented community detection in the dynamic social network using the k -clique mining based on triadic formal concept analysis. In another work, Hao et al. [10] introduced community detection in the social network using k -clique communities. They also demonstrated picking up the group of the users in the social network using a maximal clique in yet another work [8]. Gregori et al. [11] introduced how to pick up a community on a large-scale network graph based on parallel k -cliques. Palla et al. [12] showed how to identify communities subsequently and determine a community of complex networks in nature and society using the CFinder software. Kumpula et al. [13] presented an improved method of picking up the users using the sequential clique percolation; this method is very significant in community detection with less time taken for execution. Tewari and Priyanka [14] presented the book recommender system based on the CF and association rule mining. Leung et al. [15] proposed how to

avoid the cold-start problem by showing cross-level association rule mining. Tewari and Barman [16] introduced a book recommendation based on the social network and association rule mining. Jomsri [17] presented a book recommendation based on users' personal information with association rule mining for the digital library. Table 1 shows the relation of the technique used in each work.

In Table 1, A refers to the k -clique method, B, the clique method, C, the maximal clique method, D, a formal concept analysis method, E, a social network method, F, a collaborative filtering method, G, a ranking measure method, H, an association rule mining method, and I, a recommendation system.

Table 1. Comparison of similar works

Author	Description	A	B	C	D	E	F	G	H	I
Vilakone et al. [9]	Efficient movie recommendation algorithm based on improved k -clique	✓	x	x	x	x	x	x	x	✓
Hao et al. [7]	k -cliques mining in dynamic social networks based on triadic formal concept analysis	✓	x	x	✓	✓	x	x	x	x
Hao et al. [10]	K -clique communities detection in social networks based on formal concept analysis	✓	x	x	✓	✓	x	x	x	x
Hao et al. [8]	Detecting bases of maximal cliques in social networks	x	x	✓	x	✓	x	x	x	x
Gregori et al. [11]	Parallel (k)-clique community detection on large-scale networks	✓	x	x	x	✓	x	x	x	x
Palla et al. [12]	Uncovering the overlapping community structure of complex networks in nature and society	x	x	x	x	✓	x	x	x	x
Kumpula et al. [13]	Sequential algorithm for fast clique percolation	x	✓	x	x	x	x	x	x	x
Tewari and Priyanka [14]	Book recommendation system based on collaborative filtering and association rule mining for college students	x	x	x	x	x	✓	✓	✓	✓
Leung et al. [15]	Applying cross-level association rule mining to cold-start recommendation	x	x	x	x	x	x	x	✓	✓
Tewari and Barman [16]	Collaborative book recommendation system using trust-based social network and association rule mining	x	x	x	x	✓	✓	x	✓	✓
Jomsri et al. [17]	Book recommendation system for digital library based on user profiles by using the association rule	x	x	x	x	x	x	x	✓	✓
Proposed method	A movie recommendation system based on users' personal information and movies rated using the method of k -clique and normalized discounted cumulative gain	✓	x	x	x	✓	x	✓	✓	✓

We briefly define the terms used in this paper.

- **Recommendation system:** It is a technology that gives accurate and relevant information to users by analyzing advantage information from large data sets. The system examines the information format by learning users' behavior and generating results relevant to the information that users were looking for [18]. The recommendation system is becoming more popular and widely used such as recommending music, recommending products, words search, approving books, researching articles, endorsing a movie, recommending news, and social tags, It is also an expert system for supporting insurance and recommending jokes, financial services, clothing, restaurants, and so on. Most advice systems use content filtering to create a list of recommendations [19].

- k -Clique technique:** It is a technique used for analyzing information in the big social networks with their very complex community structure. Typically, k -clique is defined as a cluster of users connected in the big social networks graph. The k -clique technique picks up the communities from the network graph with k cliques. This graph consists of the sub-graphs of k users that were completely connected; for example, if $k = 3$ is referring to a triangle or $k = 4$ is referring to a rectangle, this rectangle consists of 3 triangles [7]. An example of the k -clique graph is shown in Fig. 1.

In Fig. 1, A is a group of 6-clique, B is a group of 3-clique, C is a group of 4-clique, and D is a group of 4-clique. Inside the A group are nine groups of 3-clique, and inside the C and D groups are three groups of 3-clique in each group.

- Normalized discounted cumulative gain method:** This technique is an evolution ranking measure used in the Web search engine algorithms [20,21]. The concept of this technique involves using a relevant scale of the score in the document at a search engine output set after examining the gain of a document based on its position in the output table. There are two advantages of this ranking measure: first is that it allows any retrieved document that has scored relevance; the second advantage involves a discount function over the rank. This kind of feature is essential for search engines because users keep documents that are ranked higher than others. This method is expressed as:

$$NDCG = \frac{DCG}{IDCG} \tag{1}$$

$$DCG = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)} \tag{2}$$

$$IDCG = \sum_{i=1}^{|REL|} \frac{2^{rel_{i-1}}}{\log_2(i+1)} \tag{3}$$

where p is accumulated at a particular rank position, IDCG is the ideal discounted cumulative gain, and $|REL|$ refers to the list of relevant documents in the corpus up to position p .

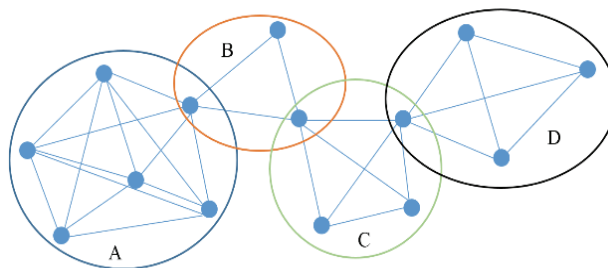


Fig. 1. Example of a graph with k -clique.

3. Proposed Movie Recommendation System

This chapter gives details on the processes of the proposed method. The process starts from number 1 to process number 10. The aspect of the process is shown in Fig. 2.

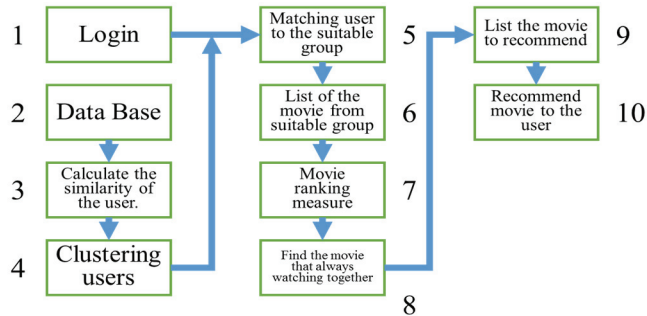


Fig. 2. Workflow of the proposed method.

A new user is required to sign up to the system with their personal information like age, gender, occupation, etc. in the first process of this workflow. Then all input information will be stored in the database in the second process of this workflow. The practical dataset containing 800 users is used to measure the user's similarities based on the cosine similarity method as the third process of this workflow. At the end of this process, the result of a measure is converted into an adjacency relationship matrix table. The element of this table is shown as Y and N, where Y refers to the users who are similar, and N denotes not similar. In the fourth process, the adjacency matrix table is used to convert into a network graph, which is then used to separate the user into groups based on the k -clique method where the number of the group is based on the value of k . The value of k starts from 3 until the number of group is equal to 0; for example, if the value of $k = 15$, a group consisting of fifteen members is not available. Later, the personal information of a user is used to compare with the personal data of the users in all groups based on the cosine similarity measure, as shown in the fifth process. After finish the comparison, the new user will decide the group when the value is equal to 1. Sometimes, a new user can find various suitable groups because the extent of similarity of the new user to the users in each group might be equal to 1 in more than one group. Once the new user finds the suitable group for him/her, the movie from the appropriate group will be presented as a table of the movie in the sixth process. To achieve the recommended movies, the normalized discounted cumulative gain method is used to calculate the top movie in the seventh process. After the list of famous movies is generated from the seventh process; the association rule mining method will find the movie that is always watched in step 8. Finally, the list of recommended movies will be display in step 9. After generating the list of movies in step 9, the first five movies with high score from max to min will be recommend to the users in step 10.

4. Performance Analysis

Dataset

To ensure that the proposed method gives the best accuracy in the recommendation system; the Movie Lens dataset was used in the experimentation part [22]. The dataset used in the experimentation part is divided into two sections. The first part is called a practical dataset, which is composed of 85% of all users and is used for the training of the experiment; the second part is called test dataset, which is composed of 15% of all users and is used for testing the experiment. In the dataset, there are 100,000 ratings of 1,684 movies by 943 users, and the necessary personal information contained in this dataset includes age, gender, and occupation.

Implementation Result

Some results of each step from the experiment in this paper will be presented in this chapter. According to the brief description in step 7 from Fig. 2, to generate the recommended movies, the normalized discounted cumulative gain method was used to calculate the famous movie. For the details of how to calculate the ranking, the measure is shown below.

- **First:** After receiving the list of movies rated by the members in the group where new users belong, the list of movies rated by the member in the group is shown in Table 2.
- **Second:** Calculate DCG. The list of rated Movie1 and the result of DCG are shown in Tables 3–5.

$$DCG_4 = \sum_{i=1}^4 \frac{rel_i}{\log_2(i-1)} = 4 + 1.722 + 1.5 + 1.261 = 8.484 \tag{4}$$

- **Third:** Calculate IDCG.

First, the rated movies are sorted from high to low rate, and the result after sorting is shown in Table 6.

The fourth maximum rated movie is then selected. An example of maximum selection rate is shown in Table 6.

Calculate IDCG of Movie1. The result of the calculation is shown in Tables 7 and 8.

$$IDCG_4 = \sum_{i=1}^4 \frac{rel_i}{\log_2(i-1)} = 4 + 1.722 + 1.5 + 1.261 = 8.484 \tag{5}$$

- **Fourth:** Calculate NDCG. To calculate the result of NDCG for Movie1 to Movie5, Eq. (6) is used; the result is shown in Table 9, and the value of NDCG is between 0 and 1.

$$NDCG_4 = \frac{DCG_4}{IDCG_4} \tag{6}$$

Table 2. Movie IDs and their ratings

MVID	US1	US2	US3	US4	US5	US6	...	USN
MD1	4	2	3	4	0	0
MD2	0	3	4	0	0	3
MD3	3	0	4	2	4	0
MD4	0	0	3	0	3	0
MD5	4	3	4	0	4	0
...
MDN

Table 3. Movie1 and its rating

MVID	US1	US2	US3	US4	US5	US6	...	USN
MD1	4	2	3	4	0	0

Table 4. Table for calculating the value of DCG

i	rel _i	log ₂ (i+1)	rel _i / log ₂ (i+1)
1	4	1	4
2	2	1.585	1.722
3	3	2	1.5
4	4	2.322	1.722
5	0	2.585	0
6	0	2.807	0

Table 5. DCG result

MVID	MD1	MD2	MD3	MD4	MD5
DCG	8.484	3.892	7.408	2.660	9.440

Table 6. Movie1 and its rating after sorting the values from max to min and the selected movie ID and its rating

MVID	US1	US2	US3	US4	US5	US6	...	USN
MD1	4	4	3	2	0	0

Table 7. IDCG calculation

i	rel _i	log ₂ (i+1)	rel _i / log ₂ (i+1)
1	4	1	4
2	4	1.585	1.722
3	3	2	1.5
4	2	2.322	1.261
5	0	2.585	0
6	0	2.807	0

Table 8. IDCG result

MVID	MD1	MD2	MD3	MD4	MD5
IDCG	8.484	3.892	7.408	2.660	9.440

Table 9. NDCG result

MVID	DCG	IDCG	NDCG
1	8.484	8.484	1
2	3.892	3.892	1
3	7.408	7.408	1
4	2.660	2.660	1
5	9.440	9.440	1

Algorithm 1. Normalized discounted cumulative gain algorithm: an algorithm used to calculate the ranking measure of the movie

```

1: Input: MovieList.csv
2: Output: Ranking
3: Initialize Re = MovieList.csv
4: for(i in 1:nrow(Re))
5:   for(j in 1:ncol(Re)){
6:     Re[i,1] = Dataset[i,1]
7:     Re[i,2] = Dataset[i,2]
8:     Re[i,3] = Dataset[i,3]
9:     Re[i,4] = Dataset[i,4]
10:    Re[i,5] = log2(1+1)
11:    Re[i,6] = log2(2+1)
12:    Re[i,7] = log2(3+1)
13:    Re[i,8] = Re[i,2]/Re[i,5]

```



```

14:         Re[i,9] = Re[i,3]/Re[i,6]
15:         Re[i,10] = Re[i,4]/Re[i,7]
16:         Re[i,11] = Re[i,8]+Re[i,9]+Re[i,10] }
17:   Initialize SSorting = Re[,11]
18:   Initialize data = c(SSorting)
19:   Initialize names = c(1:nrow(Re))
20:   Initialize group1 = data.frame("Group_ID"=c(names),"Sum"=c(data))
21:   Initialize group2 = group1[order(-group1$Sum),,drop=FALSE]
22:   for(i in 1:15)
23:     results2[hj,i] = group2[i,1]
  }
24: Write.csv(results2,"results3.csv", row.names=FALSE)

```

After generating the list of movies from the 9th process from Fig. 2, the first of five movies with high score from max to min will be recommended to the users. The list of recommended movies is shown in Table 10.

In Table 10, the values in the element of the “NUID” column mean the new user ID. Likewise, the value in an element of the column “List of recommended movies” refers to the list of movies to recommend to the new user. For example, new user ID “NUID7” received the recommended movies MVID50, MVID286, MVID100, MVID127, and MVID181.

Table 10. Recommended movies for the new users

NUID	List of recommended movies
NUID1	MVID50, MVID258, MVID260, MVID288, MVID294
NUID2	MVID1, MVID258, MVID748, MVID7, MVID50
NUID3	MVID118, MVID258, MVID300, MVID748, MVID1
...
NUID141	MVID50, MVID100, MVID174, MVID258, MVID181
NUID142	MVID258, MVID50, MVID181, MVID288, MVID300
NUID143	MVID50, MVID100, MVID286, MVID127, MVID181

Analysis Result

After the experiment was completed, to guarantee that the proposed method is the best method that offers higher accuracy than some of the methods used in this experiment, the testing dataset was used as a tool to evaluate the accuracy of the result from the proposed method. For the calculation, the mean absolute percentage error (MAPE) equation is used to calculate the value of the accuracy [23-25]. MAPE is widely used in the field of statistics structure for predicting the accuracy of the predictive method; the equation of MAPE is shown below.

$$MAPE = \frac{100\%}{n} \sum_t^n \left| \frac{A_t - F_t}{A_t} \right| \tag{7}$$

In Eq. (7), A_t means the value of the actual result, and F_t means the value of the forecast result.

MAPE's result value using the existing methods was used in this study's experiment, and the proposed method performed calculation by using Eq. (7). If the result value of MAPE for each technique is small, this means that the procedures are useful. The details for the result value of each method are shown below.

– MAPE's result value of the proposed approaches is shown in Table 11 and Fig. 3.

In Table 11, NUS1–NUS143 are the id of the new user from user 1 to user 143, $M(k3)$ to $M(k14)$ is a MAPE of the value of k from 3 to 14 in the k -clique method, and M-average is an average value of each amount of k in the k -clique.

Fig. 3 shows the result of the experiments. The result depends on the values of k . The minimum values of MAPE's achievement are 12.72% when k is equal to 11. Fig. 3 also shows MAPE's results value using the k -clique method and k -clique with the association rule mining method. The process of experiment of the k -clique methods and k -clique with the association rule mining method is similar to the operation of the proposed method, but there is a difference in the process of listing the recommended movies for the new user. In the k -clique methods, during the process of listing, the recommended movies for the users use the top N method of finding the most popular movie from all movies archived from a suitable group to new users. For the k -clique with the association rule mining methods, during the process of listing, the recommended movies for the new users used the data mining method like association rule mining to find the movie that is always watched by the users in the suitable group. In Fig. 3, the blue line refers to the result of the average MAPE of the k -clique methods. The orange line denotes the average MAPE of the k -clique plus association rule mining method, and the gray line refers to the average MAPE of the proposed method.

Table 11. Result value of MAPE for the proposed method

NUSID	M(k3)	M(k4)	M(k5)	M(k6)	M(k7)	...	M(k12)	M(k13)	M(k14)
NUS1	1	1	0	2	0	...	2	0	1
NUS2	2	0	0	0	2	...	1	0	0
NUS3	3	2	1	0	0	...	1	1	0
NUS4	0	1	0	1	1	...	0	1	1
NUS5	4	2	1	4	1	...	1	2	2
...
NUS140	2	1	0	1	4	...	3	0	0
NUS141	0	2	1	4	1	...	1	0	0
NUS142	0	0	1	0	1	...	1	0	1
NUS143	1	1	4	1	2	...	4	2	3
M average	20.00	18.60	15.52	18.60	18.32	...	14.54	14.82	14.68

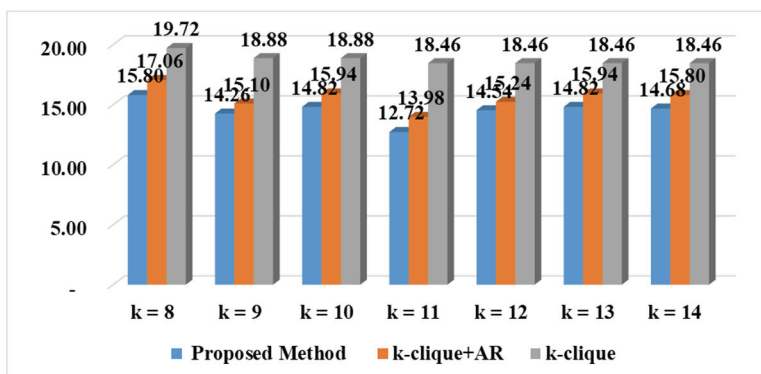


Fig. 3. MAPE result.

The result value of MAPE for collaborative filtering using the *k* nearest neighbor combined with a normalized discounted cumulative gain method (CF-*k*NN+NDCG) is shown in Table 12. The workflow of the CF-*k*NN+NDCG processed as the *k*NN method is used for the users classified into several groups by using the users' personal information. After that, the new user will be assigned to a suitable group with the help of *k*NN. Later, the normalized discounted cumulative gain method helps find the movie from an appropriate group for the new user as the movie to be recommended to the new user.

In Table 12, NUS1–NUS143 are the id of the new user from user 1 to user 143, MAPE CF-*k*NN+NDCG is a MAPE of the CF-*k*NN+NDCG method, and MAPE average is an average value of CF-*k*NN+NDCG.

The average result value of MAPE for the proposed method and the average result value of MAPE for the existing techniques experimented on in this study are compared as shown in Fig. 4.

In Fig. 4, we can guarantee that the best and most accurate method is the proposed method since the result value of the average of MAPE is 12.72%. Next is *k*-clique combined with the association rule mining method with average MAPE value of 13.98%. CF-*k*NN+NDCG also offers high accuracy with average MAPE value of 16.22%, and the original *k*-clique methods give good accuracy in the recommendation system with average MAPE value of 18.46%.

Table 12. MAPE result of the CF-*k*NN+NDCG method

	MAPE CF- <i>k</i> NN+NDCG
NUS1	2
NUS2	1
NUS3	0
...	...
NUS143	0
MAPE average	16.22

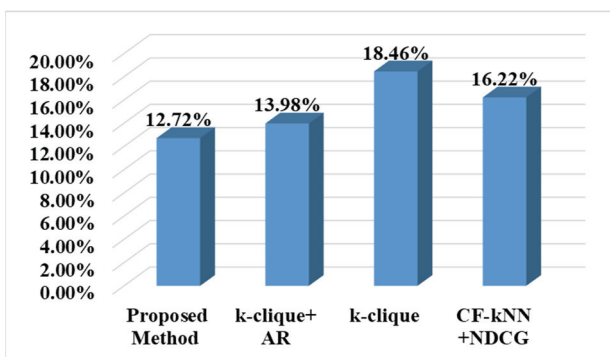


Fig. 4. Comparison of MAPEs of the existing method.

5. Conclusions

The most basic technique in the recommendation system is based on the behavior information of the users or based on the best-selling product, and the result of this method gives satisfaction to the users. Inside it, however, is the problem of cold-start, i.e., it is tough to recommend something to the new user

who has not yet stored his/her information in the system in advance. Therefore, this study proposed a method that helps solve the cold-start problem and realize higher satisfaction than the existing method used in the recommendation system. The idea of this proposed method used the personal information of the users to classify users into several communities with the help of the k -clique, which is a social network analysis method. After that, the system will generate the recommended movies for the new users from the list of the movies in the best suitable community and provide them to the new user by using the normalized discounted cumulative gain method. The result value for MAPE of the existing method used in this paper as shown in Fig. 3 and the excellent result value of MAPE were found at $k = 11$. A comparison with the existing method was also conducted, showing that the proposed method offers higher accuracy and confirming that the proposed method is the best compared to the existing method used in this paper.

For future studies, the algorithm of the k -clique will be modified to decrease the time of experiment and apply it to a vast dataset.

Acknowledgement

This research was supported by National Research Foundation of Korea (No. NRF-2020R1A2B5B01002134).

References

- [1] W. H. Jeong, S. J. Kim, D. S. Park, and J. Kwak, "Performance improvement of a movie recommendation system based on personal propensity and secure collaborative filtering," *Journal of Information Processing Systems*, vol. 9, no. 1, pp. 157-172, 2013.
- [2] P. Viana and J. P. Pinto, "A collaborative approach for semantic time-based video annotation using gamification," *Human-centric Computing and Information Sciences*, vol. 7, article no. 13, 2017.
- [3] D. Lee, "Personalizing information using user's online social networks: a case study of CiteULike," *Journal of Information Processing Systems*, vol. 11, no. 1, pp. 1-21, 2015.
- [4] A. Souril, Sh. Hosseinpour, and A. M. Rahmani, "Personality classification based on profiles of social networks' users and the five-factor model of personality," *Human-centric Computing and Information Sciences*, vol. 8, article no. 24, 2018.
- [5] F. Hao, D. S. Park, and Z. Pei, "When social computing meets soft computing: opportunities and insights," *Human-centric Computing and Information Sciences*, vol. 8, article no. 8, 2018.
- [6] F. Hao, D. S. Sim, D. S. Park, and H. S. Seo, "Similarity evaluation between graphs: a formal concept analysis approach," *Journal of Information Processing Systems*, vol. 13, no. 5, pp. 1158-1167, 2017.
- [7] F. Hao, D. S. Park, G. Min, Y. S. Jeong, and J. H. Park, "k-cliques mining in dynamic social networks based on triadic formal concept analysis," *Neurocomputing*, vol. 209, pp. 57-66, 2016.
- [8] F. Hao, D. S. Park, and Z. Pei, "Detecting bases of maximal cliques in social networks," in *Proceedings of the 11th International Conference on Multimedia and Ubiquitous Engineering (MUE)*, Seoul, South Korea, 2007.
- [9] P. Vilakone, D. S. Park, K. Xinchang, and F. Hao, "An efficient movie recommendation algorithm based on improved k-clique," *Human-centric Computing and Information Sciences*, vol. 8, article no. 38, 2018.
- [10] F. Hao, G. Min, Z. Pei, D. S. Park, and L. T. Yang, "K-clique communities detection in social networks based on formal concept analysis," *IEEE Systems Journal*, vol. 11, no. 1, pp. 250-259, 2017.

- [11] E. Gregori, L. Lenzini, and S. Mainardi, "Parallel (k)-clique community detection on large-scale networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 8, pp. 1651-1660, 2013.
- [12] G. Palla, I. Derenyi, I. Farksa, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814-818, 2005.
- [13] J. M. Kumpula, M. Kivela, K. Kaski, and J. Saramaki, "Sequential algorithm for fast clique percolation," *Physical Review E*, vol. 78, no. 2, article no. 026109, 2008.
- [14] A. S. Tewari and K. Priyanka, "Book recommendation system based on collaborative filtering and association rule mining for college students," in *Proceedings of 2014 International Conference on Contemporary Computing and Informatics (IC3I)*, Mysore, India, 2014, pp. 135-138.
- [15] C. W. K. Leung, S. C. F. Chan, and F. L. Chung, "Applying cross-level association rule mining to cold-start recommendations," in *Proceedings of 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops*, Silicon Valley, CA, 2007, pp. 133-136.
- [16] A. S. Tewari and A. G. Barman, "Collaborative book recommendation system using trust based social network and association rule mining," in *Proceedings of 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, Noida, India, 2016, pp. 85-88.
- [17] P. Jomsri, "Book recommendation system for digital library based on user profiles by using association rule," in *Proceedings of the 4th edition of the International Conference on the Innovative Computing Technology (INTECH)*, Luton, UK, 2014, pp. 130-134.
- [18] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*. Boston, MA: Springer, 2011, pp. 1-35.
- [19] H. Jafarkarimi, A. T. H. Sim, and R. Saadatdoost, "A naïve recommendation model for large databases," *International Journal of Information and Education Technology*, vol. 2, no. 3, pp. 216-219, 2012.
- [20] K. Jarvelin and J. Kekalainen, "IR evaluation methods for retrieving highly relevant documents," in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, 2000, pp. 41-48.
- [21] K. Jarvelin and J. Kekalainen, "Cumulated gain-based evaluation of IR techniques," *ACM Transactions on Information Systems*, vol. 20, no. 4, pp. 422-446, 2002.
- [22] F. M. Harper and J. A. Konstan, "The MovieLens Datasets: history and context," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 4, article no. 19, 2015.
- [23] C. Tofallis, "A better measure of relative prediction accuracy for model selection and model estimation," *Journal of the Operational Research Society*, vol. 66, no. 8, pp. 1352-1362, 2015.
- [24] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679-688, 2006.
- [25] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *International Journal of Forecasting*, vol. 32, no. 3, pp. 669-679, 2016.



Phonexay Vilakone <https://orcid.org/0000-0001-5226-6941>

He received his bachelor's degree in Mathematics and Computer Sciences from the National University of Laos, Laos, 2003. He received the master degree of Computer Application (Software System) from Guru Gobind Sigh Indraprastha University, India, 2010. His research interests include data mining and parallel processing. Since March 2017, he is with the Department of Computer Science and Engineering from Soonchunhyang University, South Korea as a PhD candidate.



Khamphaphone Xinchang <https://orcid.org/0000-0002-7387-1777>

She holds a bachelor's degree in Information Technology from National University of Laos, Laos, 2016. Her current research interests include data mining and parallel process. Since March 2017, she is with the Department of Computer Sciences and Engineering from Soonchunhyang University, South Korea as a Master student.



Doo-Soon Park <https://orcid.org/0000-0002-2776-8832>

He received his Ph.D. in Computer Science from Korea University in 1988. Currently, he is a professor in the Department of Computer Software Engineering at Soonchunhyang University, South Korea. He is Director of Wellness Service Coaching Center at Soonchunhyang University and Director of Computer Software Research Group in the Korea Information Processing Society (KIPS). He was President of KIPS from 2015 to 2015, and Director of Central Library at Soonchunhyang University from 2014 to 2015. He was editor in chief of *Journal of Information Processing Systems* (JIPS) at KIPS from 2009 to 2012, and Dean of the Engineering College at Soonchunhyang University from 2002 to 2003. He has served as an organizing committee member of international conferences including, FutureTech 2019, WORLDIT 2019, GLOBALIT 2019, CSA 2018, BIC 2018, MUE 2018, WORLDIT 2018, GLOBALIT 2018, CUTE 2017, FutureTech 2017, MUE 2017, WORLDIT 2017, GLOBALIT 2017. His research interests include data mining, big data processing, and parallel processing. He is a member of IEEE, ACM, KIPS, KMS, and KIISE.