JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Personalized Movie Recommendation System Combining Data Mining with the *k*-Clique Method

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

Abstract

Today, most approaches used in the recommendation system provide correct data prediction similar to the data that users need. The method that researchers are paying attention and apply as a model in the recommendation system is the communities' detection in the big social network. The outputted result of this approach is effective in improving the exactness. Therefore, in this paper, the personalized movie recommendation system that combines data mining for the *k*-clique method is proposed as the best exactness data to the users. The proposed approach was compared with the existing approaches like *k*-clique, collaborative filtering, and collaborative filtering using *k*-nearest neighbor. The outputted result guarantees that the proposed method gives significant exactness data compared to the existing approach. In the experiment, the MovieLens data were used as practice and test data.

Keywords

Association Rule Mining, *k*-Cliques, Recommendation System

1. Introduction

The recommendation system is an essential tool for e-commerce [1-3] because it is beneficial for the business owner to predict data and give the correct data to the user. The recommendation system is widely used in the field of online service like shopping online, listening to music online, watching movies online, etc. Most methods used in the recommendation system involve the recorded data of the users regarding purchased goods or the product that a user is interested in to predict the need for data of a user; in the recent technologies, the private data of a user are stored in the social network [4] or websites. Thus, the recommendation system can use this kind of data to increase the exactness of data prediction that meets the need of the users.

Currently, the researchers are interested in the approach that creates a group of social users from the social network communities; they are using this approach as the model in the recommendation system to increase the exactness of data prediction to the users. One of such approaches is the *k*-clique approach, which is the approach used to analyze data in the vast social network [5-7]; the output of this approach is a very satisfactory result, because the effect of data prediction is given high exactness from the

recommendation system. Nonetheless, this approach might provide more increase in data prediction with high precision to the users by combining this approach with the data mining method. Therefore, this paper proposes a personalized movie recommendation system that combines data mining with the *k*-clique method.

The main objective of this paper is to come up with a more excellent efficient approach than the *k*-clique method wherein graphs are correctly connected with k nodes and groups are created from the diagram in a social network [8]. On the other hand, data mining is a technique of analyzing hidden patterns and relationships in data from vast amounts of data by using automated methods from statistics, machine learning, pattern recognition, and mathematics. As a very popular model in the field of recommendation system, data mining uses techniques like classification method, clustering method, and prediction method. Besides that, the association rule mining method, which is one of the data mining techniques, is highly favored in analyzing the frequency of the item in the data transaction.

In the proposed approach, the personalized data of the users are used for checking the similarity of the users with the help of the cosine similarity measure approach. The relationship table will then present the meaning of similarity of the users. After that, the *k*-clique method is used to create several groups of users from the relationship table. Later, the suitable group for the new users will be checked with the help of the cosine measure approach, and the famous movie from the appropriate group will be calculated with the help of collaborative filtering. In the end, the approach of the association rule mining will be used to find the movie that is frequently watched by the member from the group where the new user belongs and recommend such movie to the new users.

The proposed solution enables more effective evaluation. To assess the performance, the proposed approach was compared with existing methods such as *k*-clique process, collaborative filtering using the k-nearest neighbor, and original collaborative filtering approach. The results of the proposed method provide significant exactness data compared to the existing strategy. In this study, the MovieLens data were used as practice and test data.

The rest of this paper is organized as follows: Section 2 presents the related work particularly the definition of some methods used in this paper; Section 3 discusses the proposed method and explains the work in more detail; Section 4, the experimental analysis, presents the experimental results and comparison. Finally, Section 5 presents the conclusions and future works.

2. Related Work

This section defines the *k*-clique approach, provides a brief explanation of collaborative filtering, introduces the recommendation system, data mining, and association rule mining used in this experiment, and presents the experience of using this method.

2.1 Recommendation System

The recommender system is an algorithm that provides relevance and exactness data to the users by analyzing the useful data from big data sets. The recommendation system checks information patterns by learning the activity of the users and creating results relevant to the need or interests of the users [9]. Recommendation systems are increasingly favored and used in broad areas, e.g., products, music, search

terms, books, research articles, news, movies in general, and social tags. They are also used as expert systems for insurance, jokes, financial services, collaborators, clothing, restaurants, and so on. Most recommendation systems use content filtering to make a list of recommendations [10].

2.2 Data Mining

Data mining is an approach of analyzing hidden patterns and relationships in data from vast amounts of data using automated methods from statistics, machine learning, pattern recognition, and mathematics, such as data mining algorithms and data warehouses, to facilitate business decision making and for efficient analysis. Data mining is known to have the purposes of knowledge discovery and data discovery. The specific benefit of data mining varies according to the goal and the industry; for example, sales and marketing departments can use the customer's information to create one-to-one marketing promotions. Data mining data of customer behaviors and on historical sales patterns can be used to create a prediction model for new products, services, and future sales. In the financial industry, data mining is used to generate the model for detecting fraud and risk. Manufacturing uses data mining as a tool for improving product safety, identifying quality issues, managing the supply chain, and improving operations. Data mining techniques include tracking patterns, classification, association, outlier detection, clustering, regression, and prediction.

Association rule mining is one of the methods in the data mining process. This method is used for finding the rules that may govern correlation, causation, and association structures between sets of items from the transaction databases or calling frequency pattern items in transaction databases. The transaction table, which consists of various issues, is used to find things that are often bought together. For example, beer and nuts are usually bought together because many people love to drink beer while eating nuts. In the same way, beer and diapers, beer and eggs, and cereal and milk are bought together, and so on [11]. Table 1 shows the transaction item table of products purchased by customers.

Table 1. Transaction table

TransactionID	ItemsSet
1	Eggs, Beer, Apple, Cereal
2	Beer, Eggs, Cereal
3	Apple, Diapers, Cereal
4	Eggs, Beer

The association rule mining applications are as follows:
- Fraud detection (supervisor → examiner).
- Loss-leader analysis: they are services that are often at a lower price than the cost by the supplier as a gamble used to encourage further sales.
- Catalog design: it is the selection of items in a business catalog.
- Weblog analysis.
- Cross marketing: it is to work with another business to complement your own instead of being competitors.
- Basket data analysis: it is used to analyze the association of purchases.

The rule form for association rule mining is as follows:

$$A => B \ [s, c] \tag{1}$$

where "s" refers to support; it means the frequency pattern of the association rule mining within transactions. If the result of support is a high value, that means the rule relates to a great part of the transaction database.

$$support(A => B [s, c]) = p(A \cup B) \tag{2}$$

"c" pertains to confidence; it means that the transactions' percentage consists of "A" as well as "B." It is an estimation of a conditional probability.

$$confidence(A => B [s, c]) = sup(A, B) / sup(A) \tag{3}$$

Fig. 1 presents the process of checking the frequency pattern of the item. In C1, item {4} has support = 1, because {4} can be found only at TID100 in database D. In C2, L1 is for checking the frequency pattern of 2 items; we found that {1 5} and {1 2} are equal to 1; it was found in TID300 upon checking the frequency pattern for {2 3 5}. As shown in TID200 and TID300, the support value of {2 3 5} = 2. Finally, the frequency pattern item is {2 3 5}.
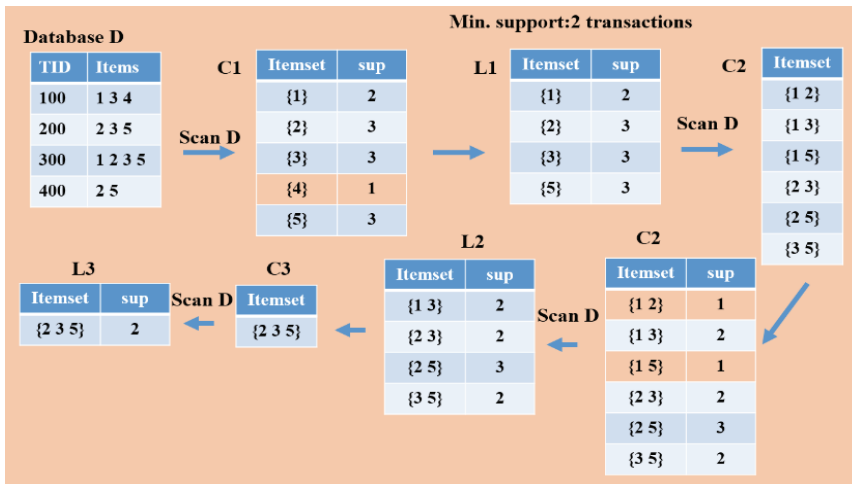

Fig. 1. Example of checking the various pattern item sets.

This section discusses how this method applied to the recommendation system. Tewari and Priyanka [12] proposed the book recommendation system using the CF and association rule mining for college students. Leung et al. [13] showed how to avoid cold-start recommendations by applying cross-level association rule mining. Tewari and Barman [14] suggested using association rule mining and social network to form a collaborative book recommendation system. Jomsri [15] showed the use of user profiles with association rule mining in the book recommendation system for the digital library.

2.3 Collaborative Filtering Method

The approach of collaborative filtering involves analyzing user actions and settings in a huge volume of data and subsequently predicting how the similarity affects the other users [16]. As the benefits of collaborative filtering, non-trusted content can be analyzed and items' complexity can be displayed correctly. Popular algorithms used to measure similarity among users or likeness of the subject in

recommendation systems are the Pearson correlation approach and k-nearest neighbor approach. Another collaborative filtering idea is based on supposition; for example, a product that users buy in the past might be bought again by them in the future, and users may probably like the same products that they wanted in the past. When modeling from user actions, differences in prediction between the forecast and actual data collection models often appear. Item-to-item collaborative filtering is a famous example of collaborative filtering. The weaknesses of this method are cold start, sparsity, and scalability; this method also consists of two types: model-based collaborative filtering and memory-based collaborative filtering [17].

2.4 k-Clique Method

The k-clique approach is a popular method of analyzing the data of the complex community structure in massive social networks and is usually always defined as a group of nodes that connected nodes in the system. The k-clique method creates network communities from k cliques, which respond to complete sub-graphs of k nodes. For example, $k = 3$ is a triangle or $k = 4$ is a rectangle with having three triangles inside it [7].

Fig. 2 represents the groups of 6-clique at node {1; 2; 3; 4; 5; 6}, since each node is connected to each other. A set of node {1;2;3;4;5} is a group of 5-clique, whereas sets of {6;2;1} or {6;3;2} or {6;4;3} or {6;5;4} are groups of 3-clique.
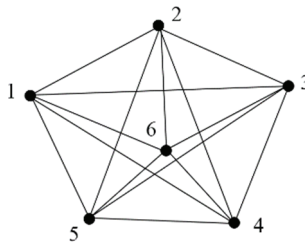
Fig. 2. Example of a graph with k-clique.

This section presents works on how the k-clique method was used in practice. For instance, Vilakone et al. [18] proposed an improved k-clique plan as an efficient method in the recommendation system. Hao et al. [7] presented the k-clique mining method based on triadic formal concept analysis for dynamic social networks. Hao et al. [19] used a formal concept analysis method for the detection of k-clique communities in social networks. Jafarkarimi et al. [10] performed detection using a maximal clique in the social networks, and Gregori et al. [20] showed how to detect community on a large-scale system using parallel k-cliques. Palla et al. [21] showed how to extract subsequently the group communities and determine the k-clique community using the CFinder software. Kumpula et al. [22] demonstrated how to improve the detection efficiency using the sequential clique percolation algorithm.

3. Design of Personalized Movie Recommendation System

This part will provide more details about the work processes of the proposed approach. There are six processes in this work, i.e., process 1 to process 6. The aspect of the method is shown in Fig. 3.

The objective of this approach is to recommend movies from a suitable group to the new user. The personalized data of the user will be used to measure their similarity, and then the users will be separated into various groups. Therefore, the new user, before logging into the system is required to sign up with essential personalized information to the system like age, gender, occupation, and more as the first process in Fig. 3. Then, in the 2nd process, the practical dataset consists of 800 users; it is used to measure the similarities of the users. The cosine similarity method is used to find the similarity of users. Based on the extent of similarity, a value of 1 is set if all features are similar, otherwise, a value of 0 is set if all of the highlights have no similarity. At the end of the process, the adjacency relationship matrix table is created. This table shows the similarities of users, and the element of this table is shown as 0 and 1. A value 1 means the users are similar, whereas a value of 0 means they are not identical.

Fig. 3. Workflow of the proposed method.

In the third process, the adjacency matrix table is used to separate the users into various groups using the *k*-clique method. The number of users in the group is based on the value of *k*. The amount of *k* starts from 3 until it can no longer be assigned; for example, when the cost of *k* is equal to 15, a group consisting of 15 members is not valid because there is no group that includes 15 members. Upon completion of this process, the number of members in the group starts from 3 users until the last value of *k* as shown in Table 2. After that, to check the suitable group, the personalized data of a user is used to compare with the personalized data of all users in the group using the cosine similarity measure method. Finally, the group consisting of a value of 1 refers to the similarity of a new user with that group; it will be decide as a suitable group for the new user. Nonetheless, a new user is always adequate for various groups depending on the similarity of their personalized information measurement; this event happens in the 4th process. Then, when a suitable group is assigned to be the group of a new user, the movie watched or rated by the member of that group is introduced as a recommended movie for the new user. To generate the recommended movies, the collaborative filtering method is used to calculate the famous movie. A popular movie is a movie that is highly rated and which is already sorted from large to small. After a list of famous movies is generated, the association rule mining approach will be used to check the frequency pattern of the popular movies. Finally, the list of recommended movies is generated. In the end, after the list of recommended movies is created, the top 5 movies will be introduced to the new users. Nonetheless,

users will select which among the recommended movies to watch.

In Table 3, the value in the cell of the "New UserID" column refers to the user ID. The value in the "GroupID of new user" column is the group ID. Table 3 also lists the results as to which group the new user can stay in, e.g., NUs4 can stay in groups Gp12, Gp24, Gp26, Gp32, Gp34, Gp37, Gp.

Table 2. Group table wherein the value of k is equal to 11

Group ID	User ID in each group
GP1	US377, US442, US459, US466, US501, US502, US511, US610, US793, US820, US943
GP2	US139, US322, US332, US534, US566, US586, US640, US649, US773, US781, US941
GP3	US361, US377, US442, US459, US466, US501, US502, US511, US610, US793, US820
GP4	US361, US377, US442, US459, US466, US501, US502, US511, US610, US793, US943
GP5	US361, US377, US442, US459, US466, US501, US502, US511, US610, US820, US943
GP6	US361, US377, US442, US459, US466, US501, US502, US511, US793, US820, US943
GP7	US361, US377, US442, US459, US466, US501, US502, US610, US793, US820, US943
GP8	US361, US377, US442, US459, US466, US501, US511, US610, US793, US820, US943
GP9	US361, US377, US442, US459, US466, US502, US511, US610, US793, US820, US943
GP10	US361, US377, US442, US459, US501, US502, US511, US610, US793, US820, US943
….	……
GP725	US76, US139, US322, US332, US566, US586, US640, US649, US773, US781, US941
GP726	US76, US139, US322, US534, US566, US586, US640, US649, US773, US781, US941
GP727	US76, US139, US332, US534, US566, US586, US640, US649, US773, US781, US941
GP728	US76, US322, US332, US534, US566, US586, US640, US649, US773, US781, US941

Table 3. Group table for the new user

New UserID	GroupID new user stay
NUs1	Gp64
NUs2	Gp41
NUs3	Gp42, Gp51, Gp58
NUs4	Gp16, Gp24, Gp26, Gp32, Gp34, Gp37, Gp…
NUs5	Gp12, Gp28, Gp98, Gp99, Gp100, Gp143, Gp…
NUs6	Gp8, Gp12, Gp13, Gp17, Gp18, Gp20, Gp…
NUs7	Gp16, Gp24, Gp26, Gp32, Gp34, Gp37, Gp…
NUs8	Gp16, Gp122, Gp123, Gp124, Gp245, Gp246, …
NUs9	Gp45
NUs10	Gp32
….	…..
NUs140	Gp16, Gp24, Gp26, Gp32, Gp34, Gp37, Gp…
NUs141	Gp10, Gp31, Gp43, Gp44, Gp46, Gp31, Gp…
NUs142	Gp6, Gp11, Gp15, Gp17, Gp22, Gp23, Gp…
NUs143	Gp16, Gp24, Gp26, Gp32, Gp34, Gp37, Gp…

In Table 4, Us1–Us19 refer to the user IDs of the new members in the group. Mov1–Mov1682 refer to the movie ID rated by the member in the group. Number one refers to the movie evaluated by the user in the group, and zero refers to the user with no rating for the movie.

Table 4. Relationship of a movie rated by the user and the group of new users

	Us1	Us2	Us3	Us4	Us5	…	Us17	Us18	Us19
Mov1	1	1	0	0	1	…	0	1	0
Mov2	0	0	0	0	0	…	0	0	0
Mov3	0	0	0	0	0	…	0	1	0
Mov4	1	0	0	1	0	…	1	0	0
Mov5	0	1	0	0	0	…	0	0	1
Mov6	0	0	1	0	0	…	1	0	0
…	…	…	…	…	…	…	…	…	…
Mov295	0	0	0	0	0	…	0	0	0
Mov296	1	0	0	1	0	…	0	1	0
Mov297	1	0	1	0	1	…	1	0	1
Mov298	1	1	0	0	1	…	0	1	0
Mov299	0	0	0	0	0	…	0	0	0
Mov300	0	0	1	1	0	…	0	0	1
…	…	…	…	…	…	…	…	…	…
Mov778	0	0	1	0	0	…	0	1	0
Mov779	1	0	0	1	0	…	1	0	0
Mov780	0	1	0	0	0	…	0	1	1
Mov781	0	0	0	0	0	…	0	1	0
…	…	…	…	…	…	…	…	…	…
Mov1680	0	0	1	0	0	…	0	0	1
Mov1681	1	0	0	1	0	…	1	0	0
Mov1682	0	1	0	0	1	…	0	1	0

Table 5. Recommended movies to the new user

New UserID	Recommended movie ID
NUs1	Ms50, Ms258, Ms260, Ms288, Ms294
NUs2	Ms1, Ms258, Ms748, Ms7, Ms50
NUs3	Ms118, Ms258, Ms300, Ms748, Ms1
NUs4	Ms50, Ms100, Ms286, Ms127, Ms181
NUs5	Ms288, Ms294, Ms127, Ms286, Ms328
NUs6	Ms288, Ms294, Ms286, Ms50, Ms258
NUs7	Ms50, Ms286, Ms100, Ms127, Ms181
…	….
NUs141	Ms50, Ms100, Ms174, Ms258, Ms181
NUs142	Ms258, Ms50, Ms181, Ms288, Ms300
NUs143	Ms50, Ms100, Ms286, Ms127, Ms181

In Table 5, the value in the cell of the "New UserID" column refers to the new user ID. The value in the "Recommended movie IF" column is related to the list of movies recommended to the new user. For example, unique user ID "NU4" got M50, M100, M258, M127, and M181 as the recommended movies.

Algorithm 1. Similarity Measure Algorithm (algorithm used to calculate the similarity between users and make a relationship table of users)

Input: practical.csv.

Output: AdjacencyMatrixTable.csv.

Step 1: initialize *mat1, mat2*.

Step 2: for f = 1 to *mat1* do

Step 3:　　for g = 1 to *mat1* do{

Step 4:　　　initialize *an1*

Step 5:　　　for s = 2 to 4 do

Step 6:　　　　if *mat1*[f,s] == *mat1*[f,s]

　　　　　　　an1[s] = 1

　　　　　else　　*an1*[s] = 0

Step 7:　　　　if (*an1*[2]+*an1*[3]+*an1*[4]) == 3

　　　　　　　mat2 [f,g] = 1

　　　　　else　　*mat2* [f,g] = 0 }

Algorithm 2. Group Classification Algorithm (algorithm used for separating users into various groups)

Input: AdjacencyMatrixTable.csv

Output: usersgroup

Step 1: initialize *mat1* = AdjacencyMatrixTable.csv

Step 2: convert *mat1* to graphnetwork

Step 3: initialize *mat2* = *clique(mat1,min=9,max=9)*

Step 4: initialize *k1*

Step 5: for m = 1 to nrow(*mat2*) do　{

Step 6:　　for h = 1 to length(*mat2*) do {

Step 7:　　　*k1*[m,h] = *mat2*[[i][j]]

Step 8: initialize *newcliq*

Step 9: for i = 1 to nrow(*k1*) do {

　　　　　for gh = 1 to ncol(*k1*) do

　　　　　　　newcliq[m,gh] = *k1*[*k1*[m,h,1]

Step 10: write newcliq.csv

Algorithm 3. Association Rule Mining Algorithm (algorithm used to check the frequency pattern of the movie)

Input: Popular.csv, Rating.csv

Output: RecommendedMovies

Step 1: initialize *mat1* = Popular.csv

Step 2: initialize *mat2* = Rating.csv

Step 3: initialize *mat3* = *mat2 – mat1*

Step 4: initialize *mat4 as transpose of mart3*

Step 5: initialize *d1 as data transactions*

Step 6: initialize *frequent* = eclate (tDatas, parameters = list(sup = 0.2, maxleng = 15))

Step 7: inFre1 = as.matrix(inspect(frequent))

Step 8: for(a in 1:nrow(maxFreIte1))

maxFreIte[a] = lengths(regmatches(RemoveSymbol[a], gregexpr(",", RemoveSymbol[a])))

Step 9: NewmaxFreIte1 = data.frames("Groups_ID"=c(names),"Sums"=c(datas))

Step 10: maxFreIteO = NewmaxFreIte1[order(-NewmaxFreIte1$Sum),,drop=FALSE]

Step 11: Recommend1 = c(RemoveSymbol[maxFreIteO[1,1]])

Step 12: RemoveCom1 = unlist(strsplit(Recommend1, split=","))

Step 13: MovieID1 = as.matrix(extract_numeric(RemoveCom1))

Algorithm 4. MAPE Algorithm (algorithm used to find the mean absolute percentage error of the experimental result)

Input: result.csv, rating.csv.

Output: MAPE's result.

Step 1: initialize TestingRate, mae.

Step 2: for i = 1 to 143 do{

Step 3:   initializes hav, sco.

Step 4:   for j = 1 to 5 do {

Step 5:   if((TestRate[i,j] == TestRating[k,1]) & (TestRating[k,(i+1)]>0))

Step 6:   hav = hav + 1

Step 7:   sco = sco + TestRating[k, (i+1)] }}

Step 8:   mae[i,1] = i

Step 9:   mae[i,2] = hav

Step 10:   mae[i,3] = sco/hav }

Step 11:   write mae.csv

4. Performance Analysis

4.1 Dataset

To test the effectiveness of the approach, the MovieLens [23] dataset was used for the experimentation. It was separated into two parts: the first part is the practical dataset consisting of 800 users and which were used for the training on the purpose approach; the second part is the test data of 143 users and were used for testing the purpose approach. In the MovieLens dataset consisting of 100,000 ratings, there were 1,684 movies and 943 users, and the necessary data of users are age, gender, and occupation.

4.2 Experimental Setup

For the implementation of the proposed approach, some of the necessary components of hardware and software included the following: Windows 7 ultimate service pack 1 for the operating system; Intel Core i5 750 at 2.67 GHz for the central processing unit; 24 GB random access memory; 64-bit processor type, and 500 GB hard drive memory. For software, we used RStudio version 3.4.0 i386.

4.3 Analysis Result

At the end of this work when the experiment part of the proposed method was completed, the number of movies recommended to the user and movies rated by them from the system was predicted. The

evaluation metric used to estimate the output of the proposed method is the mean absolute percentage error (MAPE). MAPE is popularly used in the field of statistics since it has a structure of predicting the accuracy of the predictive method; the formula of MAPE is shown below [24-26].

$$MAPE = \frac{100\%}{n} \sum_t^n \left| \frac{A_t - F_t}{A_t} \right| \tag{4}$$

A_t is the value of the actual result, and F_t is the value of the forecast result.

MAPE's result value for the collaborative filtering using the k-nearest neighbor method, k-cliques method, and approach of original collaborative filtering is used to evaluate the preciseness of the proposed method. If the MAPE output value result is a low value, that means our proposed plans are useful. Eq. (4) is used to calculate the MAPE result value.

First of all, the mean absolute percentage error for the proposed method is calculated, and MAPE's result value as derived using this method is 13.98% (see Table 6) when k is 11.

$$MAPE = \frac{100\%}{n} \sum_t^n \left| \frac{A_t - F_t}{A_t} \right| = 13.98\% \tag{5}$$

Table 6. MAPE's result using the proposed method

New UserID	MAPE (k3)	MAPE (k4)	MAPE (k5)	MAPE (k6)	MAPE (k7)	…	MAPE (k12)	MAPE (k13)	MAPE (k14)
NUs1	2	0	0	0	1	…	1	1	1
NUs2	0	1	2	2	1	…	2	1	2
NUs3	1	1	1	0	1	…	0	1	1
NUs4	0	0	0	0	0	…	0	0	0
NUs5	4	1	1	5	2	…	2	4	3
NUs6	1	1	2	1	1	…	1	1	1
NUs7	1	0	1	0	1	…	0	1	1
NUs8	4	1	2	0	1	…	1	1	1
…	…	…	…	…	…	…	…	…	…
NUs140	1	2	0	2	0	…	1	1	1
NUs141	0	1	0	1	0	…	3	1	1
NUs142	1	0	2	2	2	…	1	1	1
NUs143	1	2	2	2	2	…	3	1	0
MAPE average	20.84	19.86	16.50	19.86	19.58	…	15.24	15.94	15.80

In Table 6, MAPE (k3) to MAPE (k14) refer to the mean absolute percentage error of k = 3 to 14. NUs1 to NUs143 refer to the new user, with the number in the cell relating to the result of MAPE of each user. The MAPE average refers to the average effect of MAPE of k.

The details of MAPE's result value using the proposed approaches are shown in Table 6 above and in Fig. 4. From Fig. 4, we can see that the outcome of the experiments is based on the various values of k. Fig. 4 shows that the minimum MAPE result value is 13.98% when k is equal to 11. Moreover, Fig. 4 shows MAPE's result value of k-clique without the association rule mining method. After completion of the calculation of MAPE's result value using the proposed method, MAPE's result value for collaborative filtering using the k-nearest neighbor method was calculated, and it was found to be 18.88%. In the end, MAPE's result value using the approach of the original collaborative filtering was calculated, and it was found to be 42.00%.
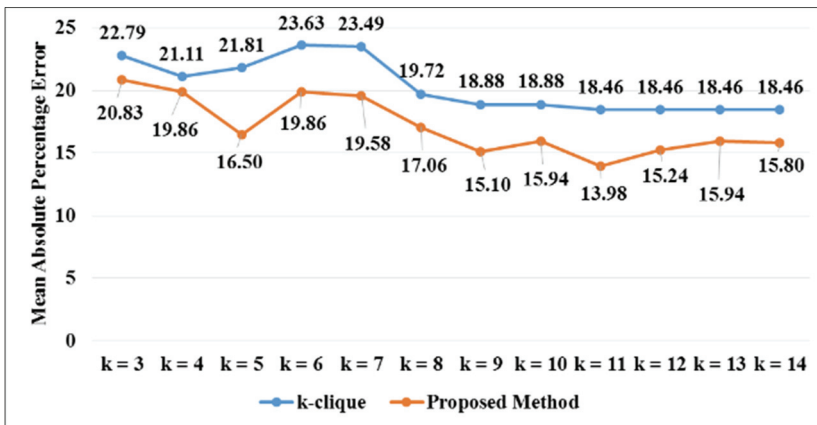
Fig. 4. Result of the proposed method and k-clique method.

In Table 7, MAPE CF-kNN refers to the mean absolute percentage error of collaborative filtering using k-nearest neighbor, and MAPE CF pertains to the mean absolute percentage error of collaborative filtering. NUs1–143 pertain to the new user and number in the cell as a result of the MAPE of each new users. The MAPE average refers to the average effect of the MAPE of both methods.

Table 7. MAPE result using the CF-kNN and maximal clique methods

	MAPE CF-kNN	MAPE CF
NUs1	1	2
NUs2	1	0
NUs3	0	0
NUs4	1	1
NUs5	0	2
NUs6	1	1
NUs7	2	1
NUs8	1	3
Nus9	0	0
NUs10	0	1
…	…	
NUs140	1	3
NUs141	0	1
NUs142	2	2
NUs143	2	1
MAPE average	18.88	42.00

At the last process of this experiment, we conducted a comparison of MAPE's result value using the existing method and the proposed method, i.e., a movie recommendation system based on *k*-clique, based on collaborative filtering using the k-nearest neighbor and based on collaborative filtering. The result of the comparison shows that the existing method is more accurate and useful, but the approach of the collaborative filtering had low accuracy. Fig. 5 presents the proposed method, which was found to be the better method compared to the existing methods.
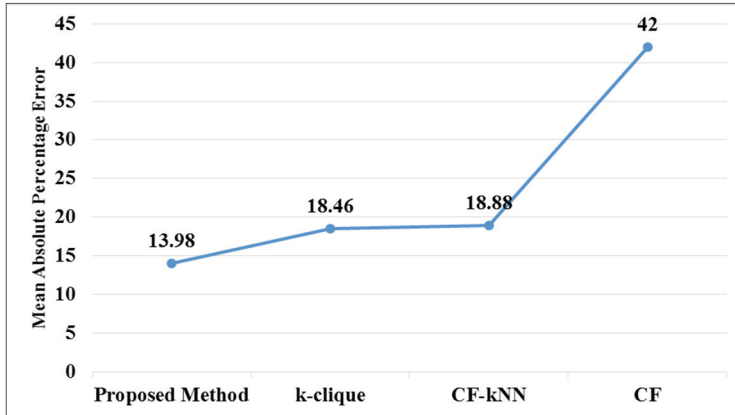
Fig. 5. Comparison with the existing method.

5. Conclusions

Currently, most researchers are interested in communities created from social networks, using this approach to enhance the exactness of predicting the needs of the users. Thus, using this approach is effective in improving accuracy. Nonetheless, the recommendation system that uses community detection in the social network as a system model can provide increased exactness of data prediction to the users when combined with another data mining method. To make the recommendation system more accurate, this study proposed the plan on the personalized movie recommendation system that combines data mining with the k-clique method. 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 process. After that, the system will generate the recommended movies for the new users from the list of movies in the most suitable community for the new user by using the data mining method. Based on the result value of the experiment as shown in Fig. 4, the best accuracy was found when the cost of $k = 11$.

The following methods were used in this experiment for evaluation: the movie recommendation system using the k-clique method; the movie recommendation system using collaborative filtering using the k-nearest neighbor; the movie recommendation system using collaborative filtering; and the proposed method. The outputted results showed that the proposed method yielded higher accuracy of providing the information matching the need of a new user compared to the other techniques used in this paper.

In future studies, the Normalized Discounted Cumulative Gain (NDCG) method will be combined with community detection in the social network method (k-clique) to enhance accuracy and effectiveness. The algorithm of the k-clique approach will then be re-modified; the purpose of modifying the algorithm of this method is to reduce the duration of execution and use it with a large dataset.

Acknowledgement

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 movie 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. Souri, 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, Korea, 2017.

[9] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*. Boston, MA: Springer, 2011, pp. 1-35.

[10] H. Jafarkarimi, A. T. H. Sim, and R. Saadatdoost, "A naive recommendation model for large databases," *International Journal of Information and Education Technology*, vol. 2, no. 3, pp. 216-219, 2012.

[11] S. K. Solanki and J. T. Patel, "A survey on association rule mining," in *Proceedings of the 5th International Conference on Advanced Computing & Communication Technologies*, Haryana, India, 2015, pp. 212-216.

[12] 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.

[13] 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 the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops*, Silicon Valley, CA, 2017, pp. 133-136.

[14] A. S. Tewari and A. G. Barman, "Collaborative book recommendation system using trust based social network and association rule mining," in *Proceedings of the 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, Noida, India, 2016, pp. 85-88.

[15] 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.

[16] G. D. Linden, B. R. Smith, and N, K. Zada, "Automated detection and exposure of behavior-based relationships between browsable items," U.S. Patent 9070156, 2015.

[17] N. Rubens, M. Elahi, M. Sugiyama, and D. Kaplan, "Active learning in recommender systems," in *Recommender Systems Handbook*. Boston, MA: Springer, 2015, pp. 809-846.

[18] 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.

[19] F. Hao, G. Min, Z. Pei, D. S. Park, and L. T. Yang, "K-clique community detection in social networks based on formal concept analysis," *IEEE Systems Journal*, vol. 11, no. 1, pp. 250-259, 2015.

[20] 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.

[21] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814-818, 2015.

[22] 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.

[23] 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, 2016.

[24] 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.

[25] 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.

[26] 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 are data mining and parallel processing. Since March 2017, he is with the Department of Computer Science and Engineering from Soonchunhyang University in 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 in Korea as a Master student.

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

He got a Ph.D. in 1988 from Computer Science from Korea University. Today, he is a professor in the Computer Software Engineering of Soonchunhyang University, Korea. Besides that, he is director-general of Wellness Service Coaching Center at Soonchunhyang University. He also works for the KIPS as a Director of Computer Software Research Group. He was President of the Korea Information Processing Society from 2015 to present. He is a Director of Central Library from 2014 to 2015 of the Soonchunhyang University. He was editor in chief of Journal of Information Processing Systems at KIPS from 2009 to 2012, and Dean of the Engineering College at Soonchunhyang University from 2002 to 2003. He is an organizing committee member of international conferences including, FutureTech 2018, WORLD-IT 2018, GLOBAL-IT 2018, CSA 2017, BIC 2017, MUE 2017, WORLD-IT 2017, GLOBAL-IT 2017, CUTE 2016, FutureTech 2016, MUE 2016, WORLD IT-2016, GLOBAL-IT 2016. His research interests include data mining, big data processing, and parallel processing. He is a member of IEEE, ACM, KIPS, KMS, and KIISE.