JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Movie Recommendation Algorithm Using Social Network Analysis to Alleviate Cold-Start Problem

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

Abstract

With the rapid increase of information on the World Wide Web, finding useful information on the internet has become a major problem. The recommendation system helps users make decisions in complex data areas where the amount of data available is large. There are many methods that have been proposed in the recommender system. Collaborative filtering is a popular method widely used in the recommendation system. However, collaborative filtering methods still have some problems, namely cold-start problem. In this paper, we propose a movie recommendation system by using social network analysis and collaborative filtering to solve this problem associated with collaborative filtering methods. We applied personal propensity of users such as age, gender, and occupation to make relationship matrix between users, and the relationship matrix is applied to cluster user by using community detection based on edge betweenness centrality. Then the recommended system will suggest movies which were previously interested by users in the group to new users. We show shown that the proposed method is a very efficient method using mean absolute error.

1. Introduction

In recent years, information available on the internet is rapidly growing up; people need more time to select useful information. Social media use continues to grow rapidly, too. There are currently more than 3 billion people around the world using social media each month. The new Global Digital 2018 report has revealed that there are currently more than 4 billion people worldwide using the Internet [1]. Social networks are one of the most popular communications media today and it attracts millions of active users to share and comment on their photos and places with others [2]. To manage the information overload problem, the recommendation system has been developed. A recommendation system is a simple algorithm that allows users to relate and recommend items by filtering user-related data from large data. Recommender systems are software tools that make the recommendation of product or items that are appropriate for a customer's taste based on the analysis of information of products that many customers are interested in and the customer's and their past purchasing activity [3]. The recommendations system can help users to make decisions in multiple contexts. The goal of recommendation system is to find what's likely to be interest to the users. Over the years, many recommender systems were introduced by

---

researchers from different problem areas. We can utilize recommendation systems that use different techniques including collaborative filtering (CF) technique, content-based filtering technique, and hybrid technique. These recommendation systems will recommend popular items that customers liked [4]. CF method is one most successful and widely used in recommendation systems [5]. The basic idea of this algorithm is to use common experiences or similar interests. The important process is to search users that are similar to the target user or find the product that is similar to the predicted product. But, the CF algorithm also has some problems such as a cold-start problem [6]. Cold-start problem is a problem where the system is not able to recommend items to users. For every recommender system, it is required to build a user profile by considering the user's preferences and likes. The user profile is developed by considering her activities and behaviors they perform with the system. On the basis of user's previous history and activities, the system makes decisions and consequently recommends items. The problem arises when a new user or new item enters the system, for such users or items that the system does not have enough information to make decision. For example, if a new user has not rated some items and not yet visited/viewed some items, then it would be difficult for the system to build a model on that basis. This can lead to inaccuracies in estimating similarities between users. Many approaches have been studied to solve the existing problem. Embarak [7] suggested two types of recommendation such as node recommendation and batch recommendation, and then compared the suggested method with three other alternative methods including Naïve Filterbots Method, Media Scout Stereotype Method, and Triadic Aspect Method to solve the cold-start problem. Basiri et al. [8] suggested a new hybrid approach, which focuses on improving performance under cold start problem. This method can give a reasonable and appropriate recommendation.

With the development of technology, the user's behavior and personal information can be tracked and recorded on social networking sites or online shopping sites. This type of technology makes it easier and it is very useful for analyzing user preferences. We analysis social network analysis (SNA) methods [9,10] and introduced the betweenness centrality in SNA into a CF approach. This paper presents a movie recommendation algorithm using SNA to alleviate the cold-start problem. In the proposed method, the user's personal information such as age, gender, and occupation are used to establish a relationship between users. Then the relationship matrix between users will be applied for clustering the user into several communities or groups. In this process, the centrality of SNA is used to detection communities; after that, the system will recommend movies in the group that is similar to the target users by considering CF. The main objective of this article is to develop techniques that can recommend the most suitable movies for target users based on personal characteristics.

The proposed work is briefly described as follows. In Section 2, we will explain about the related work which is the methods used in this paper. Our proposed algorithm for movie recommendation algorithm using SNA to alleviate cold start problem will describe in Section 3. In Section 4, will present experimental analysis and experimental results and finally, the conclusion will be concluded in Section 5.

2. Related Works

In this section, we have a purpose to briefly explain about the relevant research, including the recommendation system, CF, and SNA that are required for the movie recommendation system in this study.

2.1 Recommendation System

In recent years, the recommender system has become more popular. It has been used in many areas

including book, news, movie, music, and products. There are also recommender systems for experts [11], restaurants, collaborators [12], garments, jokes, financial services [13], romantic partners, life insurance, and Twitter pages [14]. These recommendation systems use one type of filtering to predict ratings and user satisfaction, which allows users to purchase products based on their interests or needs. Having information about user's life can give hints about how the user will react when faced with different situations [15]. The recommender system is a useful alternative of search algorithms because it can help users discover what they may not find. It is usually performed using a non-traditional indexing search engine. A recommender system is a technology that makes automatics predictions about the relationship between customers or between items and searches for items that users may need. There are many approaches to make a recommender system, including the following.

- **Content-based filtering:** Content-based filtering methods depend on the item's features and user preferences [16]. These methods work with data that the user provides. Based on this data, a user profile will be created which will be used to give advice to users. As the user provides more input and accepts recommendations, the engine becomes more and more accurate. These algorithms try to recommend the products that are similar to those that the user liked in the past. In particular, many nominated the products are compared to the products that were previously ranked by the user. This method provides the foundation for data retrieval and data filtering research.

- **Collaborative filtering:** CF method is used to automatically predict or filter information about user interests by collecting settings or tasting data from multiple users (collaborating). The basis of the CF method is assuming that if user U_1 has the same taste with user U_2, user U_1 tends to have the taste of U_2 on issues that differ from those of random users [17]. Collaborative recommender systems receive a list of recommended items by analyzing similarities between users and predicting user ratings of an item based on similar user ratings on the same list [18].

- **Hybrid recommender system:** Hybrid methods can be used in many ways. This approach can be making content-based filtering and collaborative-based filtering predictions separately and it can combine them together. It can be used to enhance content-based capabilities through a collaborative-based approach. It can combine all of them into one format [19].

The purpose of many studies about the referral system is to focus on the ability to recommend products that satisfy customers primarily. CF is the most commonly used method for identifying similarities between items.

2.2 Collaborative Filtering Algorithm

The CF method is derived from collecting and analyzing large amounts of data about user behavior, activities or user preferences, and predicts which users will be similar to other users. To understand what CF is, one can think of a simple question; for instance, if a person wants to read a book, but that person does not know which book is good, what will that person do? Usually, most people like to ask friends to see which books are good. We like to receive a suggestion from friends or people who have the same taste as us. This is the main idea of the CF method [20]. It can be separated into two types: user-based and item-based CF. Item-based CF depends on the similarity between the items calculated using the ratings of people for those items. For example, when users who like item I_1 also like item I_2, the similarity between these two items is are considered similar. User-based CF takes advantage of the similarities between users in the forecast. For example, if Mr. James and Mr. Paul have seen the same movie and they are also giving

the same ratings, the similarities between them is 1. On the other side, assuming that they give the different rating to the item, the similarities will decrease as differences.

In this paper, we adopted a user-based CF algorithm. The algorithm predicts that if users' personal characteristics are similar, then their interests in products or items are also similar. The algorithm searches for the most similar users according to the target user information. Based on the most similar interests or preference, a user's interest can be predicted. The recommender system will carry out information for suggest for relevant users. Cosine similarity is used to measure the similarity of users. Its formula [21] is shown in Eq. (1).

$$\text{sim}(A, B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

In Eq. (1), A and B are two different vectors. A_i is the component of vector A and B_i is the component of vector B, respectively. The result of similarity ranges from -1 to 1. The negative value is the opposite; it refers to two different vectors. A positive value represents two similar vectors. With 0 indicating orthogonally or de-correlation while in-between values indicate intermediate similarity or dissimilarity.

2.3 Social Network Analysis

The SNA is a method used to analyze social network properties. It characterizes networked structures in terms of nodes, such as individual actors, people, or things in the network. Vertices indicate objects or entities while edges indicate links to show relationships or interactions. Both objects and the links may have attributes. Networks are constructed from general, real-world data. They propose several unexpected challenges pending to data domains themselves, e.g., information distillation, pre-processing, and data structures used for displaying knowledge and storage [22]. Social networks represent
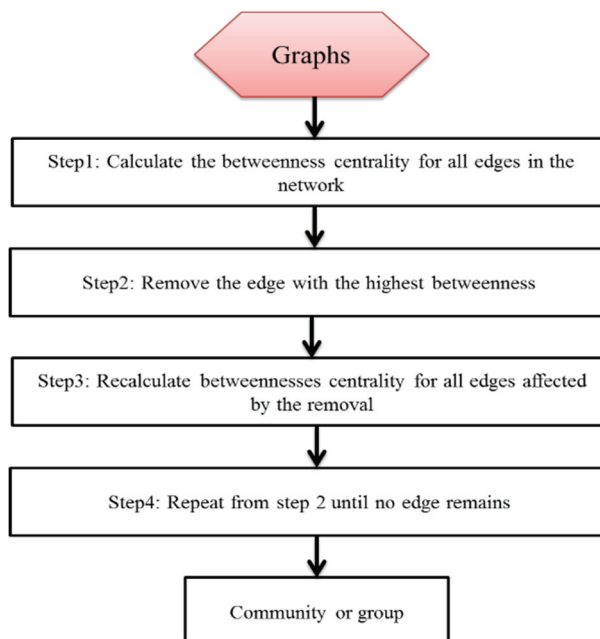

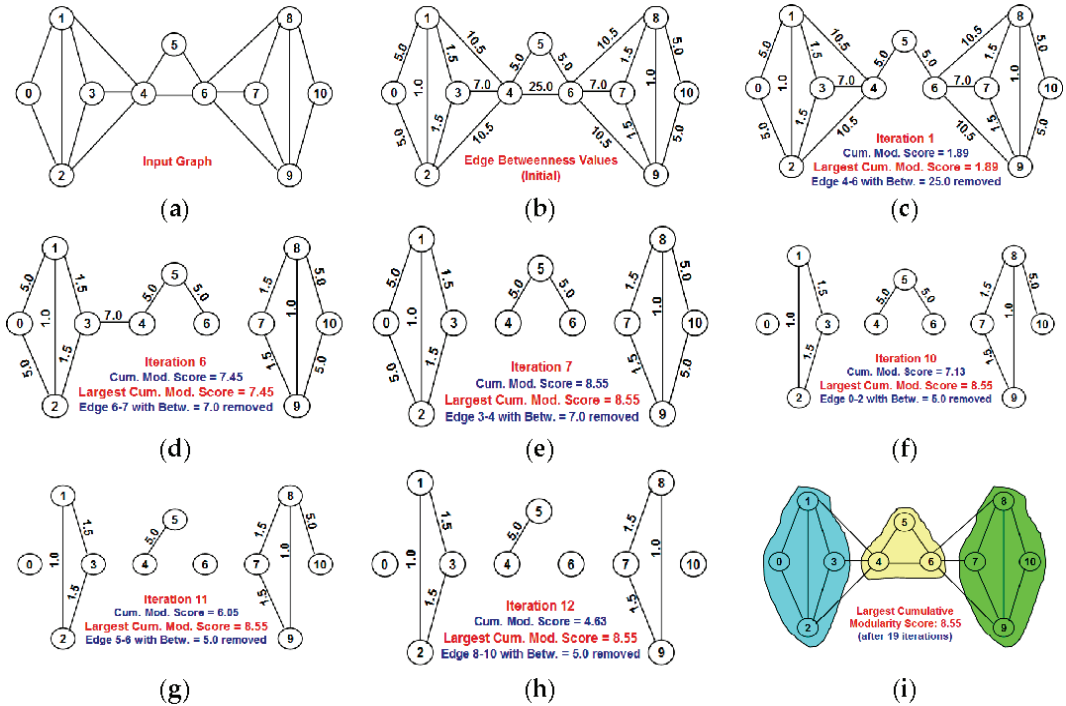**Fig. 1.** Community detect based on edge betweenness algorithm.

Fig. 2. Community structure: (a) input graph, (b) computed betweenness values for the edges in graph, (c) at the end of iteration 1, remaining edges in the graph after removed edge 4–6 with highest betweenness score of 25.0, (d) at the end of iteration 6, remaining edges in the graph after removed edges 6–7 with betweenness score is 7.0, (e) at the end of iteration 7, remaining edges in the graph after removed edges 3 and 4 with betweenness score is 7.0, (f) at the end of iteration 10, remaining edges in the graph after removed edge 0–2 with betweenness score of 5.0, (g) at the end of iteration 11, remaining edges in the graph after removed edges 5–6 with betweenness score is 5.0, (h) at the end of iteration 12, remaining edges in the graph after removed edge 8–10 with betweenness score is 5.0, and (i) at the end of iteration 19, final partitioning of the network graph into communities.

relationships that exist within the community. Social networks provide tools for collaborative education, especially through the theory developed in analyzing social networks [23]. Even in the same community, there may be various types of social networks depending on social relationships as friends, mutual support and cooperation. The similarity is a common standard used to construct social relationship components of a community. Actors or nodes in social networks can be individuals, groups, objects, organizations or events, as long as certain relationships remain together. Centrality is an agent indicator used in SNA. There are many types of centrality including degrees' centrality, closeness centrality, and betweenness centrality.

Betweenness centrality is the measure of the center of the graph based on the shortest path. For all vertices in the connected graph, there is at least one shortest path between the vertices, which is each number of edges that pass through or the sum of the weight of the minimized edge. The betweenness centrality for each vertex is the number of shortest paths that pass through the vertex. Girvan and Newman [24] have presented a community detection algorithm in social networks and biological networks based on edge betweenness to avoid the flaws of hierarchical clustering methods. This algorithm continuously detects communities by removing the edges from the original network, and the connected

components of the network that remain are communities. Instead of trying to create a measure that tells us which edge is most central to the community, but this algorithm focuses on the edge that seems to be "between" communities. The algorithm steps for community detection are summarized as Fig. 1.

An example in Fig. 2 shows the effective algorithm Girvan-Newman for edge community detection based on betweenness.

3. Movies Recommender Algorithm Using Social Network Analysis to Alleviate Cold-Start Problem

The process of data collection and workflow processes that are sufficient for the recommended system are shown in Fig. 3. ① The system needs to collect user data and movie listings into the database used as a test dataset. ② The system requests that new users log in if they want to join. ③ The system needs to collect new user data and movie listings into the database used as a test dataset. ④ In this process we make a relationship table for a user based on their personal propensity such as age, gender, and occupation; we call this table an adjacency matrix. ⑤ The relationship between the user table or adjacency matrix is used to evaluate a community out of several communities (groups) by analyzing the relationship table of the user and comparing it using community detection based on edge betweenness. ⑥ In this process, after a community (group) of the user has been selected for evaluation, then we modify the group for the new user by computationally comparing the similarity of the new user to other users in
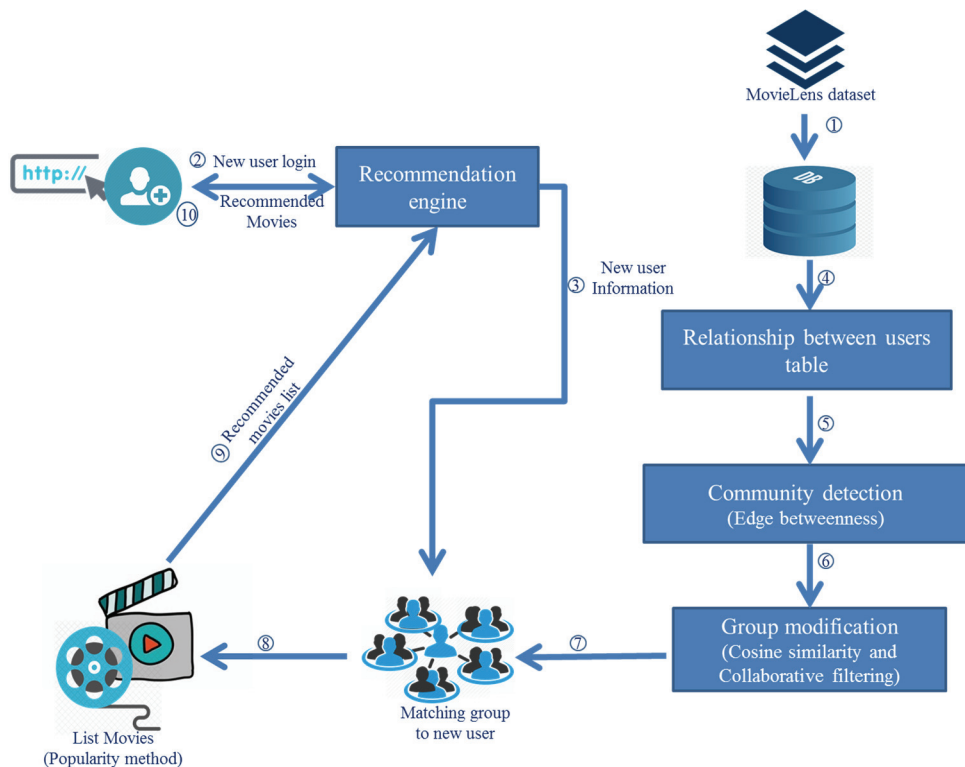

Fig. 3. Movie recommender system configuration diagram.

each group from their personal information. For computing, the similarity of user cosine similarity is used. ⑦ After that we match a group with the new user, the group with the highest similarity will be selected as a group of new users. ⑧ The movies that were watched and rated by users in the matching group of the new user will be arranged in the order of popularity. ⑨ In this process, the recommended system will select the top 5, 10, 20, 30 and 40 most popular movies and finally ⑩ the system will choose the most popular movies that were watched by the members in the group recommend these to the new user.

3.1 Detail of Proposed Processes

3.1.1 New user provides personal information

To recommend the best movies for users, we need to collect some necessary personal information about these users. Users who have the same behavior or same personal propensity may like the same item. Personal propensity that we use for performance in this research includes age, gender, and occupation.

3.1.2 The relationship between users table

In our proposed algorithm, the centrality of SNA was applied. Therefore, we have to make a relationship table between users according to their personal propensity including age, gender, and occupation. We assume that if the users have the same personal information it means they have a relationship to each order. This table can be formalized as a classical mathematical relationship that can be seen as an unspecified graph.

3.1.3 Community detection based on edge betweenness.

After getting the relationship between users table, in this process, we want to cluster users into several groups by applying the relationship between users table. Community detection based on edge betweenness is used for clustering users, a community representation as a group. The concept of detection community based on edge betweenness is that the possibility those edge connections separate modules have the highest betweenness value because the shortest path from one module to another module must cross through those modules. Therefore, if we gradually remove the edge with the highest edge betweenness value, we will get a hierarchical map or a rooted tree, called a dendrogram of the graph. The leaves of the tree are the person's vertices and the root of the tree means the whole graph in the network.

3.1.4 Cosine similarity and Collaborative filtering

Traditional methods to measure the similarity of users just consider the similarities of user ratings. In reality, the similarity of users is not only linked to ratings for items, but is also linked to the preference for certain item categories, that is, user interest for the item category feature. In addition, if two users have similar personal information, these two users are considered highly similar. Therefore, our research modified the group for the new user by using the cosine similarity measure. After users are divided into several groups by community detection based on edge betweenness, then the similarity between the new user and the other users in each group is computed by filtering personal information including age, gender, and occupation. For computing the similarity between users Eq. (1) is used.

3.1.5 The matching group for the new user

After completing the calculation of similarities between new users and other users in each group, we compared the average similarities of new users and each group. The group with the highest similarity was selected as the most similar group for related users.

3.1.6 Ranking the popularity movies

When finding a group that was most similar to a new user, movies watched by group members will be ranked according to their popularity, which was counted from the rating that each member of the group had given to each movie.

3.1.7 Recommended movies to the new user

After ranking popular movies from users that were similar to new users, the most popular movies were selected to be recommended for new users. However, the final decision about which movies the user will watch will depend on the new user.

In this article, we try to combine existing CF techniques with SNA. We use between centrality identification method and the introduction of a movie recommendation system that may give the best predictions about the movie program that the target user might be interested in. The algorithm used in our recommendation system is shown in Algorithm 1.

Algorithm 1. movies recommender system using social network analysis and collaborative filtering

1: Input: user: U_k, user-movie matrix, Rating table: R, Recommended recipient user: N
2: Output**:** Recommended list: RL
3: Procedure:
4: Create the relationships between user table
5: Clustering user U_k by using community detection based on edge betweenness
6: Cosine similarity and collaborative filtering are used to modify the group
7: Matching the group for new user.
8: Count movies rating
9: Recommended movie list.

---

4. Experimental Analysis

4.1 Experimental Dataset

In the experiment in this research, we used the MovieLens dataset provided by the GroupLens research group at the University of Minnesota, USA [25]. It comprises data from three orders of scale. Each data set has related user's information, user's ratings, and the movie's information. This dataset consisted of 100,000 ratings; the rating is in range of 1 to 5. This dataset includes 943 users and 1,682 movies. Each user gives a rating to movies at least 20 movies. This dataset also has simple information for each user, such as gender, age and occupation.

To evaluate the quality of our proposed method, the dataset is divided into two parts: the training and testing set. It is very important to evaluate performance using data not involved in formulating the model. To improve the accuracy of the recommendation system, both sets of dataset included 10 random datasets from users who gave ratings to at least 20 movies, users who gave ratings to at least 50 movies, users who gave ratings to at least 100 movies, and users who gave ratings to at least 200 movies. This means we have to implement the method a total of 40 times.

Training dataset is used after a model has been processed. In the setting of this recommender system, partitioning is performed by randomly selecting some users and some ratings from all users. There are 800 users in this dataset. Training set is implemented to build up a model. Testing dataset is used to test the model by making predictions. There are 143 users in this dataset. Some users and some ratings are then randomly selected from all users.

Fig. 4 shows part of the user's personal information provided by MovieLens. Fig. 5 displays a section of rating and matrix data generated from rating data by users who watch certain movies. In Fig. 5, the columns are user numbers or user IDs, while the rows are movie numbers or movie IDs. Table 1 shows the result of user relationships after comparing user's personalization information.

user id	age	gender	occupation	zip code
1	24	M	technician	85711
2	53	F	other	94043
3	23	M	writer	32067
4	24	M	technician	43537
5	33	F	other	15213
6	42	M	executive	98101
7	57	M	administrator	91344
8	36	M	administrator	5201
9	29	M	student	1002
10	53	M	lawyer	90703
11	39	F	other	30329
12	28	F	other	6405
13	47	M	educator	29206
14	45	M	scientist	55106
15	49	F	educator	97301

Fig. 4. MovieLens' user information part

```
196 242 3
186 302 3
22  377 1
244 51  2
166 346 1
298 474 4
115 265 2
253 465 5
305 451 3
6   86  3
62  257 2
```

movies \ users	1	2	3	4	5	6	7	8	9	10
1	5	4			4	4				4
2	3				3					
3	4									
4	3						5			4
5	3									
6	5								5	
7	4				2	5	3	4	4	
8	1				4	5				
9	5				4	5				4
10	3	2				4				
11	2			4		3	3			4

Fig. 5. Users–movies matrix part.

Table 1. The relationship table between users

	u_1	u_2	u_3	u_4	u_5
u_1	1	0	0	1	0
u_2	0	1	0	0	1
u_3	0	0	1	0	0
u_4	1	0	0	1	0
u_5	0	1	0	0	1

4.2 Experimental Environment

Hardware and software used to evaluate the methods proposed in the paper are shown in Table 2.

R programming language is an open source scripting language for predictive analysis and visualization. The R programming language includes functions that support linear modeling, non-linear modeling, classical statistics, classifications, clustering, and more. It has remained popular in academic settings due to its robust features and the fact that it is free to download in source code form under the terms of the Free Software Foundation's GNU general public license. It compiles and runs on UNIX platforms and other systems including Linux, Windows, and Mac OS. Hence, we can easily identify the source code to see what it is doing on the screen. Anyone can fix bugs and add a feature without having to wait for the seller to do it for us. Moreover, it always allows us to integrate with other languages (C, C++). Furthermore, it enables us to interact with many data sources and statistical packages (SAS, SPSS).

Table 2. Resources

Description	Specification
Operating system	Windows 7 Ultimate Service Pack 1
Processor	Intel Core i7 2600k CPU @3.40 GHz
RAM	12 GB
System type	64-bit operating system
Hard disk	500 GB
Data analysis software	R i386 3.4.2

4.3 Experimental Results

Several metrics have been proposed to evaluate the accuracy of the CF method algorithm. Mean absolute error (MAE) is one of the most commonly used tools for measuring the accuracy of the recommender system [26]. MAE evaluates the accuracy of a prediction algorithm by comparing numerical deviation of the predicted rating from the respective actual user rating. Formally, if n is the number of an actual item that is purchase by a target user and MAE is assigned as the mean absolute difference between the pair. And assuming that the predicted rating set of the target user u is $\{p_{u1}, p_{u2}, \ldots, p_{uN}\}$ and actual rating set is $\{r_{u1}, r_{u2}, \ldots, r_{uN}\}$, then MAE is defined as follows:

$$MAE = \frac{\sum_{i=1}^{N} |P_{ui} - r_{ui}|}{N} \tag{2}$$

After performing the proposed method, we want to predict the accuracy of our proposed method. We want to know when our system recommended movies, how many movies were actually watched from the

user. MAE can help measure the level of satisfaction and evaluate the accuracy of the recommender system. Normally, the lower the value of the MAE, the higher the accuracy of the recommendation. To the computation of MAE, the Eq. (2) is used.

As described in Section 4.1 we performed experiments in four cases: random testing datasets from users who gave ratings to at least 20, 50, 100, 200 movies. To reduce inaccuracy of the MAE values, we performed 10 experiments using each dataset. In total, of an experiment are 40 times. The average results of the movie recommendation system using SNA and CF for 10 experiments are shown in Fig. 6.

Fig. 5 shows the average of all experiments from 10 random times; the best result for the number of movie recommendations was 5. Therefore, we averaged the result from four cases; the average result from four cases was 3.55 when 5 movies were recommended. The average result from four cases was 3.79 when 10 movies were recommended. The average result from four cases was 4.34 when 20 movies were recommended. The average result from four cases was 4.59 when 30 movies were recommended. The average result from four cases was 4.78 when 40 movies were recommended. The maximum MAE value is 10. Therefore, we can say that our proposed method is very effective and can solve the cold-start problem. From these results, we can also interpret that the user was interested in more than 3 out of 5 movies recommended by the system.
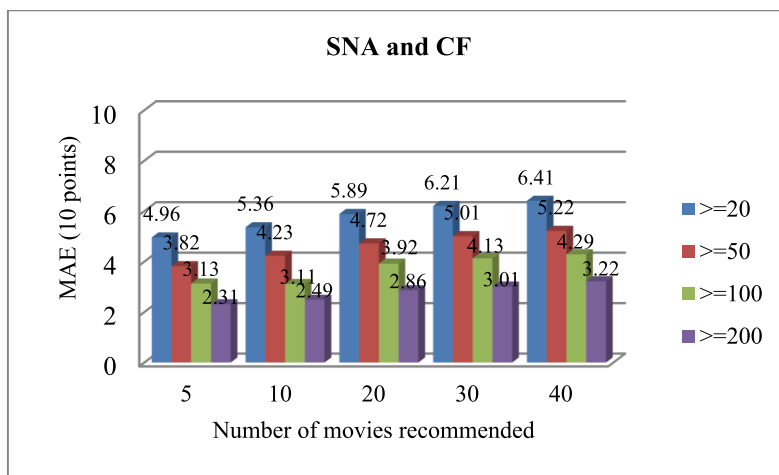

Fig. 6. The result of social network analysis (SNA) and collaborative filtering (CF) method.

In order to confirm the effectiveness of the methods that are presented in this paper, we have compared with other methods including density-based on clustering [27] method, the CF with k-NN, and CF method.

The results of the movie recommendation system using density-based on clustering by used the same dataset as a proposed method are shown in Fig. 7.

In Fig. 6, the average result from four cases was 4.19 when 5 movies were recommended. The average result from four cases was 4.76 when 10 movies were recommended. The average result from four cases was 5.28 when 20 movies were recommended. The average result from four cases was 5.44 when 30 movies were recommended. The average result from four cases was 5.63 when 40 movies were recommended. The maximum MAE value is 10. The best number of movies recommended in this method was also 5.
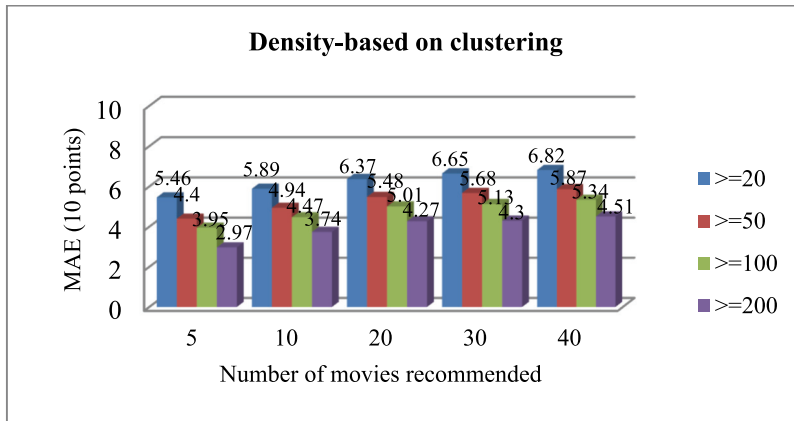
Fig. 7. The result of density-based on clustering method.

Fig. 8 showed the results of the movie recommender system using k-NN and CF, the average result from four cases was 3.60 when 5 movies were recommended. The average result from four cases was 3.96 when 10 movies were recommended. The average result from four cases was 4.35 when 20 movies were recommended. The average result from four cases was 4.61 when 30 movies were recommended. The average result from four cases was 4.85 when 40 movies were recommended. The maximum MAE value is 10. The best number of movies recommended in this method was also 5, as in previous methods.
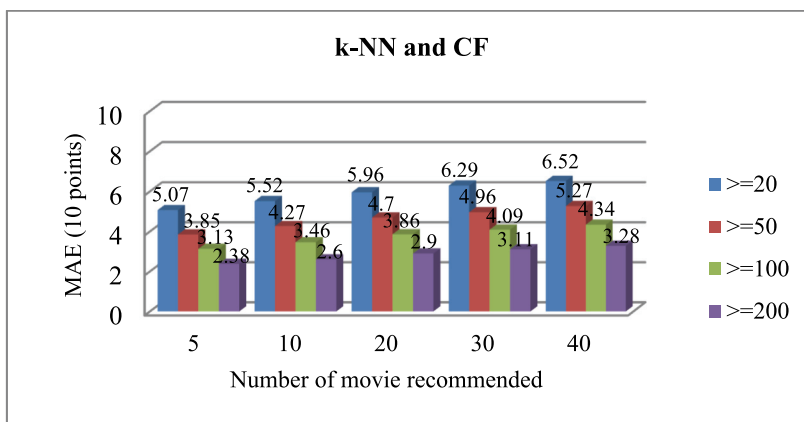

Fig. 8. The result k-NN and collaborative filtering method.

Traditional datasets that are not tuned have been tested using traditional CF algorithm. For the movie recommendation system based on original CF, the neighbor for the user who has the same taste as the new user was found. The result of the movie recommender system using CF is shown in Fig. 9. When 5 movies were recommended the result was 5.75. When 10 movies were recommended the result was 5.84. When 20 movies were recommended the result was 6.06. When 30 movies were recommended the result was 6.29. When 40 movies were recommended the result was 6.48. The best number of movies recommended using this method was also 5; same as other methods.

After that, we will show the efficiency of each method by comparing the total average of the MAE results of the four methods. The total average result of the movie recommendation system that is

proposed to use k-NN and CF shows better results than the total average result of the movie recommendation system that is proposed to use density-based clustering. The k-NN and CF method also shows better results than the results of the movie recommendation that is proposed to use original CF. The total average result of the method we propose, which is a movie recommendation system that is proposed to use SNA and CF, is more accurate than the other three methods as shown in Fig. 10. We also have to argue that the efficiency of our method is better than the other three methods. We also have to argue that the efficiency of the three methods is the best of our method.
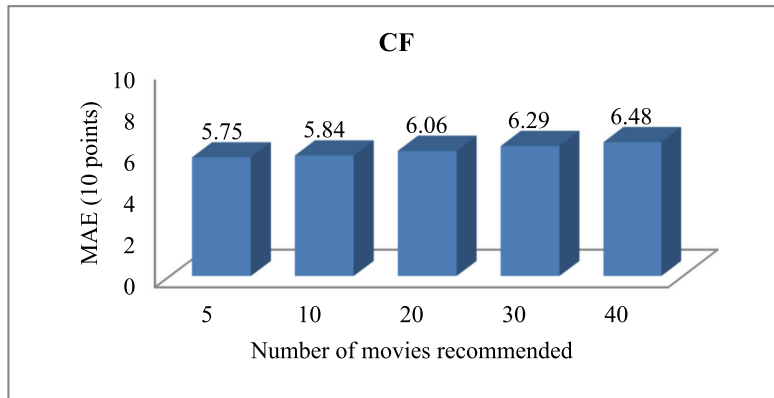

Fig. 9. The result of collaborative filtering method.
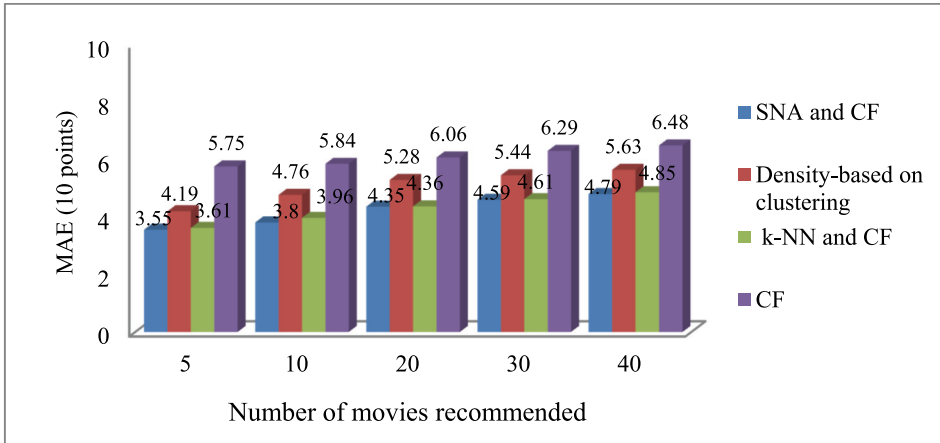

Fig. 10. Comparative the result of the experimental.

5. Conclusions and Future Work

This paper aims to solve the problem of CF by using SNA in the recommender system. We design and implement in R programming which is an open source scripting language for predictive analytics and data visualization. The recommendation system is a way to help the users to find the information that they want easily. CF is one of the most popular uses one and a successful method used in the

recommender system. However, it has weakness such as cold start problem. To overcome this problem, in this paper we proposed an alternative approach for the recommender system using both SNA and CF. We found the community or group for the user based on edge betweenness centrality. The method that we proposed here is very effective for making movie recommendations. Analyzing results showed that the total number of 20 movies recommended was better than 40 and the best number of movies recommended was 5. In addition, the method presented in this paper showed the best performance, followed by k-NN and CF, Density-based clustering, and CF.

However, the implementation processed in this paper by using R programming took a very long time. Therefore, we propose to reduce the experiment time and to improve the accuracy and effectiveness of the recommender system by applying various types of datasets in future research.

Acknowledgement

References

[1] We Are Social Ltd., "Digital in 2018: world's Internet users pass the 4 billion mark," 2018; https://weare social.com/blog/2018/01/global-digital-report-2018.

[2] S. M. Albladi and G. R. Weir, "User characteristics that influence judgment of social engineering attacks in social networks," *Human-centric Computing and Information Sciences*, vol. 8, article no. 5, 2018.

[3] 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.

[4] L. T. Ponnam, S. D. Punyasamudram, S. N. Nallagulla, and S. Yellamati, "Movie recommender system using item based collaborative filtering technique," in *Proceedings of 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, Pudukkottai, India, 2016, pp. 1-5.

[5] W. Li, X. Li, M. Yao, J. Jiang, and Q. Jin, "Personalized fitting recommendation based on support vector regression," *Human-centric Computing and Information Sciences*, vol. 5, article no. 21, 2015.

[6] S. K. Tiwari and S. K. Shrivastava, "An approach for recommender system by combining collaborative filtering with user demographics and items genres," *International Journal of Computer Applications*, vol. 128, no. 13, pp. 16-24, 2015.

[7] O. H. Embarak, "A method for solving the cold start problem in recommendation systems," in *Proceedings of 2011 International Conference on Innovations in Information Technology*, Abu Dhabi, United Arab Emirates, 2011, pp. 238-243.

[8] J. Basiri, A. Shakery, B. Moshiri, and M. Z. Hayat, "Alleviating the cold-start problem of recommender systems using a new hybrid approach," in *Proceedings of 2010 5th International Symposium on Telecommunications*, Tehran, Iran, 2010, pp. 962-967.

[9] 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.

[10] 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.

[11] H. H. Chen, I. I. Ororbia, G. Alexander, and C. L. Giles, "ExpertSeer: a keyphrase based expert recommender for digital libraries," 2015; https://arxiv.org/abs/1511.02058.

[12] H. H. Chen, L. Gou, X. Zhang, and C. L. Giles, "Collabseer: a search engine for collaboration discovery," in *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, Ottawa, Canada, 2011, pp. 231-240.

[13] A. Felfernig, K. Isak, K. Szabo, and P. Zachar, "The VITA financial services sales support environment," in *Proceedings of the 19th National Conference on Innovative Applications of Artificial Intelligence*, Vancouver, Canada, 2007, pp. 1692-1699.

[14] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. B. Zadeh, "WTF: the who to follow system at Twitter," in *Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, Brazil, 2013, pp. 505-514.

[15] A. Souri, S. 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.

[16] C. C. Aggarwal, *Recommender Systems*. Cham: Springer International Publishing, 2016.

[17] L. Terveen and W. Hill, "Beyond recommender systems: helping people help each other," in *Human-Computer Interaction in the New Millennium*. Boston, MA: Addison-Wesley, 2001, pp. 487-509.

[18] H. T. Zou, Y. He, S. Zheng, H. L. Yu, and C. Hu, "Online group recommendation with local optimization," *Computer Modeling in Engineering & Sciences*, vol. 115, no. 2, pp. 217-231, 2018.

[19] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge & Data Engineering*, vol. 17, no. 6, pp. 734-749, 2005.

[20] Y. Wang and L. Zhu, "Research on collaborative filtering recommendation algorithm based on Mahout," in *Proceedings of 2016 4th International Conference on Applied Computing and Information Technology/3rd International Conference on Computational Science/Intelligence and Applied Informatics/1st International Conference on Big Data, Cloud Computing, Data Science & Engineering (ACIT-CSII-BCD)*, Las Vegas, NV, 2016, pp. 400-405.

[21] H. Tan and H. Ye, "A collaborative filtering recommendation algorithm based on item classification," in *Proceedings of 2009 Pacific-Asia Conference on Circuits, Communications and Systems*, Chengdu, China, 2009, pp. 694-697.

[22] M. Vedanayaki, "A study of data mining and social network analysis," *Indian Journal of Science and Technology*, vol. 7, no. S7, pp. 185-187, 2014.

[23] J. Palau, M. Montaner, B. Lopez, and J. L. De La Rosa, "Collaboration analysis in recommender systems using social networks," in *Cooperative Information Agents VIII*. Heidelberg: Springer, 2004, pp. 137-151.

[24] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821-7826, 2002.

[25] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan, and J. Riedl, "MovieLens unplugged: experiences with an occasionally connected recommender system," in *Proceedings of the 8th International Conference on Intelligent User Interfaces*, Miami, FL, 2003, pp. 263-266.

[26] F. Ye and H. Zhang, "A collaborative filtering recommendation based on users' interest and correlation of items," in *Proceedings of 2016 International Conference on Audio, Language and Image Processing (ICALIP)*, Shanghai, China, 2016, pp. 515-520.

[27] S. G. Moghaddam and A. Selamat, "A scalable collaborative recommender algorithm based on user density-based clustering," in *Proceedings of the 3rd International Conference on Data Mining and Intelligent Information Technology Applications*, Macao, China, 2011, pp. 246-249.

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

She received B.S. in School of Information Technology from National University of Laos, Laos, 2016. Since March 2017, she is with the Department of Computer Sciences and Engineering from Soonchunhyang University, Korea as a Master student. Her current research interests include data mining and parallel processing.

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

He received his B.S. 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. Since March 2017, he is with the Department of Computer Science and Engineering from Soonchunhyang University, Korea as a PhD candidate. His research interests include data mining and parallel processing.

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, 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 2018, WORLDIT 2018, GLOBAL IT 2018, CSA 2017, BIC 2017, MUE 2017, WORLDIT 2017, GLOBALIT 2017, CUTE 2016, FutureTech 2016, MUE 2016, WORLDIT 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.