Multimedia Tools and Applications (2021) 80:2339–2354
https://doi.org/10.1007/s11042-020-09768-8

A hybrid recommendation system based on profile expansion technique to alleviate cold start problem

Faryad Tahmasebi[1] · Majid Meghdadi[1] · Sajad Ahmadian[2] · Khashayar Valiallahi[3]

Received: 21 May 2019 / Revised: 19 August 2020 / Accepted: 28 August 2020 /
Published online: 13 September 2020

Abstract
Recommender systems are one of the information filtering tools which can be employed to find interest items of users. Collaborative filtering is one of the recommendation methods to provide suggestions for target users based on the ratings of like-interest users. This method suffers from some shortcomings such as cold start problem leading to reduce the performance of recommender system in predicting unseen items. In this paper, we propose a hybrid recommendation method based on profile expansion technique to alleviate cold start problem in recommender systems. For this purpose, we take into consideration user's demographic data (e.g. age, gender, and occupation) beside user's rating data in order to enrich the neighborhood set of users. Specifically, two different strategies are used to enrich the rating profile of users by adding some additional ratings to them. The proposed rating profile expansion mechanism has a significant effect on the performance improvement of recommender systems especially when they are facing with cold start problem. The reason behind this claim is that the proposed mechanism makes a denser user-item rating matrix than the original one by adding some additional ratings to it. Obviously, providing a rating profile with further ratings for the target user leads to alleviate cold start problem in recommender systems. The expanded rating profiles are used to calculate similarity values between users and predict unseen items. The results of experiments demonstrate that the proposed method can achieve better performance than the other recommendation methods in terms of accuracy and rate coverage measures.

Keywords Recommender systems · Collaborative filtering · Cold start · Profile expansion · Demographic data

1 Introduction

In recent years, recommender systems are used in different domains particularly in the scope of commercial web sites. The most important task of recommender system is to estimate user's

✉ Majid Meghdadi
meghdadi@znu.ac.ir

Extended author information available on the last page of the article

interests about items in the system. To this end, the opinions of users are acquired explicitly or implicitly to find relevant items for recommending to the users. In addition, these systems can use additional information resources such as demographic data and social information [2, 6, 22]. Therefore, these systems help users to find their interest items and prevent from the wasting of user's time to obtain their relevant information [7, 16, 33, 37].

Collaborative filtering [20, 21, 25, 30] has become one of the most popular techniques which is used in recommender systems. This approach recommends the relevant items to users based on opinions of the other users who are in their neighbors set. This neighbors set is formed based on the similarity values between users and then selecting the users with higher similarity values with the target user. Collaborative filtering approaches are divided into two groups including memory-based and model-based [13]. The memory-based techniques [14] use user-item rating matrix for calculating similarity values between users and predicting unseen items for the target users. In other words, these techniques find nearest neighbors set of users by using a similarity function such as Pearson correlation coefficient [32], Cosine similarity function [13], etc. On the other hand, the model-based techniques [4, 10] make a model based on a subset of data called training set and then use this model to predict the items which are in the remaining data called test set.

Cold start [3, 12] is an important problem in recommender systems which is about new users and new items of the system. In this paper, we focus on the new user problem [31] which occurs when a new user has joined to the system and thus there is an empty or very small rating profile for this user. Hence, the system cannot accurately calculate the similarity values between such cold start users and others. Moreover, identifying the users' neighbors set is a difficult process leading to reduce reliability of recommendations. Several approaches have been proposed based on incorporating additional data resources such as user's demographic data [26, 35] or user's trust relations [8, 24, 39] into the pure collaborative filtering to resolve cold start problem. Formoso et al. [18] proposed a new approach based on profile expansion technique to alleviate cold start problem. This method expands the new user's rating profile by different techniques such as item-local, item-global, and user-local. However, additional information such as demographic data is not considered in these techniques. Therefore, in this paper, we address the new user problem by introducing a novel hybrid recommendation method which is based on profile expansion using user-item rating matrix and demographic information. The proposed method consists of two main phases. At first, the user's rating profile is expanded by using the combination of user-item rating matrix and demographic information. Then, the similarity values between users are calculated based on the new user's rating profile and also the unseen items are predicted for the target users. The proposed method provides the following contributions:

- A novel profile expansion method is proposed based on the combination of user-item rating matrix and demographic information. The rating profiles of users are enriched using the proposed profile expansion method which leads to alleviate cold start problem in recommendation systems. The main advantage of the proposed profile expansion mechanism in comparison to other previously developed approaches is to employ demographic information of users besides the user-item rating matrix.
- Two different strategies including Global Most-Rated (GMR) and Global User-Local Clustering (GUC) are used in the proposed method to expand rating profiles of users by adding some additional ratings to them.

- Different from the pure collaborative filtering-based recommendation models, the proposed method uses the expanded rating profiles of users instead of the original ones to calculate similarity values between users and predict unseen items. This leads to improve the performance of recommender systems especially when they are facing with cold start problem.
- Several experiments are conducted and their results show that the proposed method can effectively alleviate cold start problem and obtain better performance than the other recommendation methods.

The remainder of this paper is organized as follows: Section 2 discusses related works. Section 3 presents the proposed method. Experimental results are reported in Section 4. Finally, Section 5 concludes the paper.

2 Related works

Collaborative filtering is one of the most widely used techniques in recommender systems [27, 38]. However, this technique suffers from some challenges such as cold start problem. In recent years, several approaches have been proposed to alleviate the cold start problem [36]. Some of these techniques use additional data sources such as demographic data or trust relations beside the collaborative filtering method [1, 5]. Safoury et al. [34] utilized the user's demographic data for calculating similarity values between users in order to resolve the cold start problem. Papagelis et al. [29] proposed a method to alleviate data sparsity and cold start problems using trust inferences as indirect dependency between two users. For example, if user x has co-rated items with user y and also user y has co-rated items with user z, then, both of users x and z have transitive dependency through user y. Moreover, they considered two subjective notations of trust namely confidence and uncertainty properties for all of the directed dependencies. In [17], a POI recommendation method is proposed which integrates the information of user-uploaded and user-favoured photos and the high-order relationship information obtained from user social networks. The main advantage of this method is to alleviate data sparsity problem in POI recommenders by considering different data resources. In [15], a music recommendation system is proposed by employing users' information on music preferences to provide accurate recommendations for users. To improve the accuracy of recommendations, the demographic information of users including age and gender are used in this music recommendation method.

Nguyen et al. [28] employed the user's demographic data such as age, occupation, and gender that are easily available in the user's profile. To this end, they made various α-community space models for grouping the users which α is the user similarity factor. Moreover, the missing α-communities are calculated for the new users that used by rule-based induction process to predict unseen items. Lika et al. [23] proposed an approach to identify neighbors set of users by using classification algorithms and demographic data. Guo et al. [19] presented a recommendation method to incorporate user's trust information in collaborative filtering approach for making better recommendations. For this purpose, the ratings of the trusted neighbors are merged with the ratings of the active user. In addition, the quality of the merged ratings is measured by a confidence metric.

Several recommendation algorithms have been proposed to improve the performance of recommender systems which have not considered any additional data for the users. Anand et al. [11] utilized both local and global similarity measures between users by regarding various α estimation schemes to achieve recommendations with better quality. These various α estimation schemes are based on the overall sparsity of the users and items in the system. Ahn [9] proposed a new similarity measure called Proximity, Impact, and Popularity measure (PIP) for making better recommendations than existing similarity measures such as Pearson correlation coefficient and cosine similarity function. Formoso et al. [18] addressed the limitations of selecting neighbors set for new users by proposing a novel profile expansion approach that includes three types of techniques, called item-global, item-local, and user-local. These methods are based on the query expansion techniques in information retrieval. The item-global technique added most similar items into the user rating profile. On the other hand, the item-local technique consists of two phases. In the first phase, the user's rating profile is expanded with top-I recommendations list according to the items which exist in the user's rating profile. In the second phase, the system generates recommendations to the active user based on the items which are added to the user's rating profile. The local-user method selects the active user's neighbors based on the primary user's profile. Finally, a number of items among the profiles of the neighbors set are selected to expand the active user's rating profile by using different techniques such as local most-rated (LMR) and local user-local clustering (LUC).

3 Proposed method

This section presents the proposed method which consists of two main phases. In the first phase, at first, the similarity values between users are calculated based on the combination of similarities obtained by user-item rating matrix and users' demographic information. The Cosine similarity function is used to calculate the similarity values based on the user-item rating matrix. On the other hand, the users' demographic similarities are calculated through a weighted average of their demographic data [23]. Therefore, the final similarity values between users are determined based on a linear combination of both of the Cosine and demographic similarities. Then, a subset of k most similar users is selected as nearest neighbors set of the target user. Moreover, the rating profile of the active user is expanded using the items of active user's nearest neighbors. Then, in the second phase of the proposed method, the similarity values between users are calculated using the Cosine similarity function based on the expanded rating profile of users and then the unseen items are predicted for the target users. The overview of the proposed method is shown in Fig. (1).

3.1 User's profile expansion

In this subsection, the user's profile expansion phase of the proposed method is described which includes three steps. In the first step, the similarity values between users are calculated using the Cosine similarity function as follows:

$$sim\,(u,v)_{cos} = \frac{\sum_{i\in I_{u,v}} r_{u,i} \times r_{v,i}}{\sqrt{\sum_{i\in I_{u,v}} r_{u,i}^2 \sum_{i\in I_{u,v}} r_{v,i}^2}} \tag{1}$$

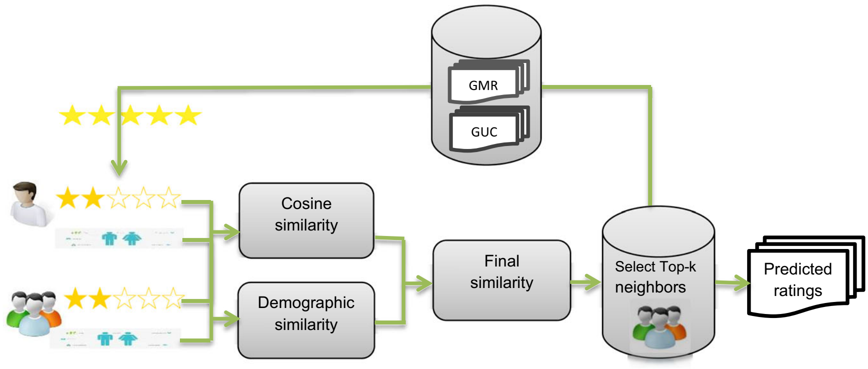

Fig. 1 The overview of the proposed method

where $sim\ (u,v)_{cos} \in [0, 1]$ is the Cosine similarity value between the users u and v, $I_{u,v}$ is the common ratings of both users u and v, and $r_{u,i}$ is rating of the user u to the item i.

In addition, we consider demographic similarity between the users in order to find better neighbors for them. The demographic similarity is based on weighted average of demographic attributes such as age, gender, and occupation which can be calculated as follows:

$$sim\ (u,v)_{demo} = \frac{\sum_{j=1}^{l} s_j \quad w_j}{\sum_{j=1}^{l} w_j} \tag{2}$$

where $sim\ (u,v)_{demo} \in [0, 1]$ is the demographic similarity between the users u and v, s_j is the similarity value of the jth attribute of the users u and v, w_j is the weight of jth attribute, and l is the number of all demographic attributes. It should be noted that, if the jth attribute of two users is equal then the value of the similarity s_j will be 1, otherwise the value will be 0.

After calculating the Cosine and demographic similarity values between users, the final similarity values can be calculated as the combination of both Cosine and demographic similarity values using the following equation:

$$sim\ (u,v) = (1-\alpha)\ sim(u,v)_{cos} + \alpha\ sim(u,v)_{demo} \tag{3}$$

where the parameter α determines the weights of the Cosine and demographic similarity values to calculate the final similarity value. In other words, if $\alpha = 0$ then the final similarity value only depends on the Cosine similarity function. On the other hand, if $\alpha = 1$ then the final similarity value only depends on the demographic similarity.

In the second step, a subset of k most similar users is selected as the nearest neighbors set of the target user. Finally, in the last step, the rating profile of the target user is expanded by the items of target user's nearest neighbors. For this purpose, the rating profile of the target user is expanded using two different strategies including Global Most-Rated and Global User-Local Clustering. The descriptions of these strategies are presented in the following.

Global Most-Rated (GMR): In this strategy, the rating profile of the target user is expanded by a subset of items which have been rated by most of the target user's nearest neighbors.

Global User-Local Clustering (GUC): This strategy attempts to find the items which are most similar to the items already existing in the target user's profile and rated by her/his neighbors. In this technique, the similarity value between two items is calculated using the following equation:

$$sim\ (i,j) = \frac{\sum_{n \in N(u)} r_{n,i} \times r_{n,j}}{|N(u)|} \tag{4}$$

where $N(u)$ is the set of neighbors for the target user u, and $r_{n,\ i}$ is the rating of user n to item i.

Algorithm 1. The pseudo-code of the proposed method

Input:

- The set of items: $Y = \{Y_1, \dots, Y_m\}$ where m is the number of items;
- The set of users: $X = \{X_1, \dots, X_n\}$ where n is the number of users;
- The user's demographic data: $X_i = \{X_i^1, \dots, X_i^l\}$ $(i = 1 \dots n)$, n is the number of users and l is the number of demographic attributes;
- The user-item matrix: $R = \{R(X_i, Y_j) | X_i \in X, Y_j \in Y\}$;

Output:

- Predicted ratings for the target users;

Begin

For all $X_i \in X$ do

- Select N ratings randomly;
- Calculate the Cosine similarity values between X_i and the other users using Eq. (1);
- Calculate the demographic similarity values between X_i and the other users using Eq. (2);
- Calculate the final similarity values between X_i and the other users using Eq. (3);
- Find the $top - k$ nearest neighbors for the target user X_i;
- Add I items by the GMR or GUC strategies to the rating profile of the target user X_i;

End for.

For all $X_i \in X$ do

- Calculate the Cosine similarity values between X_i and the other users using Eq. (1);
- Find the $top - k$ nearest neighbors for the target user X_i;
- Predict the unseen items for the target user X_i using Eq. (5);

End for.

End.

3.2 Rating prediction

In this phase, the similarity values between users are calculated using the Cosine similarity function based on the expanded rating profile of the users and then the unseen items are predicted for the target users. Therefore, the Cosine similarity values between the target user and other users are calculated using Eq. (1). Then, the k users who have highest similarity values with the target user are selected as the nearest neighbors of the target user. Finally, the unseen items for the target user can be calculated as follows:

$$pre\ (u,i) = \frac{\sum_{v \in N_u} sim(u,v) \times r_{v,i}}{\sum_{v \in N_u} sim(u,v)} \tag{5}$$

where, N_u is the set of nearest neighbors for the target user u, $sim(u, v)$ is the Cosine similarity value between users u *and* v, and $r_{v,i}$ is the rating of user v to item i. The pseudo-code of the proposed method is presented in Algorithm 1.

3.3 Computational complexity analysis

There are two main phases in the proposed method. In the first phase, at first, the similarity values between users are calculated using the combination of similarity values obtained by Cosine similarity function and demographic-based similarity function. The computational complexity of the Cosine similarity function is $|U|^2|I|$ while in the case of demographic-based similarity function is $|U|^2l$. It should be noted that $|U|$, $|I|$, and l are the number of users, items, and demographic attributes in the system, respectively. Therefore, the complexity of calculating similarity values between users is $(|U|^2|I| + |U|^2l)$. After calculating the similarity values between users, two different strategies including GMR and GUC can be employed to expand user's rating profile. GMR and GUC strategies need the computational complexity of $|U||I|k$ and $(|I|^2k + |U||I|)$, respectively. k refers to the number of users in the nearest neighbor set of target user. Accordingly, the complexity of the first phase of the proposed method is $(|U|^2|I| + |U|^2l + |U||I|k)$ for GMR strategy and $(|U|^2|I| + |U|^2l + |I|^2k + |U||I|)$ for GUC strategy. In the second phase of the proposed method, the similarity values between users are calculated based on the expanded user's rating profile which its complexity is $|U|^2|I|$. Then, the unknown ratings are predicted based on the final similarity values between users. Predicting unknown ratings for users needs the complexity of $|U||I|k$. Therefore, the complexity of the second phase of the proposed method is $(|U|^2|I| + |U||I|k)$. Finally, the summation of computational complexities related to the two main phases of the proposed method can be considered as the total computational complexity. Therefore, the total computational complexity of the proposed method is $(2|U|^2|I| + |U|^2l + 2|U||I|k)$ for GMR strategy and $(2|U|^2|I| + |U|^2l + |I|^2k + |U||I| + |U||I|k)$ for GUC strategy.

4 Experimental results

In this section, several experiments are performed to evaluate the performance of the proposed method. For this purpose, the proposed method is compared with traditional user-based CF method called NoPE, LMR, and LUC methods [18]. In the following subsections, the detailed descriptions of the used dataset, evaluation measures, and also the results of the experiments are presented.

4.1 Dataset

In the experiments, we used the MovieLens[1] dataset to compare the performance of the proposed method with the other recommendation methods. This dataset is based on a movie recommender system containing a number of users and items. There are 6040 users and 3952 items in this dataset. In this dataset, the users can express their opinions about the items by providing ratings in the range of 1 (bad) to 5 (excellent). The number of all ratings provided by the users to the items is equal to 1,000,000 ratings where each user has rated at least 20 movies in this dataset. In addition to the user-item ratings, this dataset contains demographic

[1] http://grouplens.org/datasets/movielens/

information about the users. There are several demographic attributes including user id, age, gender, and occupation related to the users in this dataset. It should be noted that these demographic attributes are explicitly collected for the users and they are available to employ as input resources for recommender systems.

4.2 Evaluation metrics

We considered accuracy and rate coverage metrics for evaluating the performance of the recommendation methods. The first metric is the Mean Absolute Error (MAE). It calculates absolute differences between predicted and real ratings of test items. The MAE metric is defined as follows:

$$MAE = \frac{\sum_u \sum_i \left| p_{u,i} - r_{u,i} \right|}{n} \tag{6}$$

where, $p_{u,\,i}$ is the predicted rating of user u to item i, $r_{u,\,i}$is the real rating, and n is the number of all ratings in the test set.

Another metric is Rate Coverage (RC) which is based on the ratio of test ratings that predicted by a recommendation method. Therefore, the RC metric can be calculated as follows:

$$RC = \frac{m}{n} \tag{7}$$

where, n is the total number of the test ratings and m is the number of test ratings that are predicted by the recommendation method.

4.3 Experimental setup

In the experiments, a subset of N ratings is randomly selected for each user to evaluate the proposed method in cold start condition. The values of parameter N are set as 2, 3, 4, 5, and 10. Therefore, the number of ratings in the user's rating profile is $N + I$ where I is the number of items added to the user's rating profile. The descriptions of the used parameters in the proposed method are presented in Table 1.

4.4 Results

In this section, different experiments are performed on the Movielens dataset to verify the improvement of the proposed method in comparison to the other recommendation methods. It

Table 1 The descriptions of the used parameters in the proposed method

Parameters	Notation	Values
Size of the basic user's profile	N	2,3,4,5,10
Number of items added to the user's profile	I	5,10,15,20,50
Number of evaluated users	U	6040
Number of neighbors for the target user	k	25
Weight value of the user's age attribute	w_1	0.3
Weight value of the user's gender attribute	w_2	0.1
Weight value of the user's occupation attribute	w_3	0.6
Impact coefficient of cosine and demography similarities	α	0.5

Table 2 MAE and RC evaluation of the proposed method with $N = 3$ and different values of I

	Metric	LMR	GMR	LUC	GUC
I = 5	MAE	0.914	**0.872**	1.054	1.002
	RC	0.056	0.051	0.045	**0.058**
I = 10	MAE	0.906	**0.869**	1.067	1.001
	RC	0.065	0.064	0.065	**0.085**
I = 15	MAE	0.915	**0.886**	1.091	1.000
	RC	0.051	0.053	0.071	**0.105**
I = 20	MAE	1.016	**0.926**	1.143	1.029
	RC	0.050	0.042	0.077	**0.114**
I = 50	MAE	0.995	0.933	**0.906**	0.981
	RC	0.026	0.019	0.077	**0.140**
No PE	MAE	0.894	0. 894	0.894	0.894
	RC	0.018	0.018	0.018	0.018

should be noted that the GMR and GUC are two techniques refer to the proposed method based on the different rating profile expansion strategies. Table 2 reports the results of the experiments based on the MAE and RC measures with $N = 3$ and different sizes of the parameter I. As we can see from these results, the GMR technique obtains the best MAE values based on the different values of the parameter I except for $I = 50$ where the LUC method is the best performer. The MAE value in the case of $I = 50$ for the GMR method is 0.933 while this value is 0.906 for the LUC method. The GUC method is the best performer based on the RC metric in terms of all values of the parameter I. Therefore, it can be concluded that the GUC method can significantly outperform other recommendation methods in alleviating the cold start problem. Increasing the value of the parameter I leads to improve the performance of the GUC method based on the RC metric. On the other hand, the performance of the GMR method is declined based on the MAE metric by increasing the value of the parameter I.

Several experiments are conducted to show the effect of different values of the parameter N on the performance of the recommendation methods. Figs. (2) and (3) show the results of experiments in terms of different values of the parameter N based on MAE and RC metrics, respectively. As we can see from Fig. (2), the MAE values of the recommendation methods are mostly declined when the value of the parameter N is

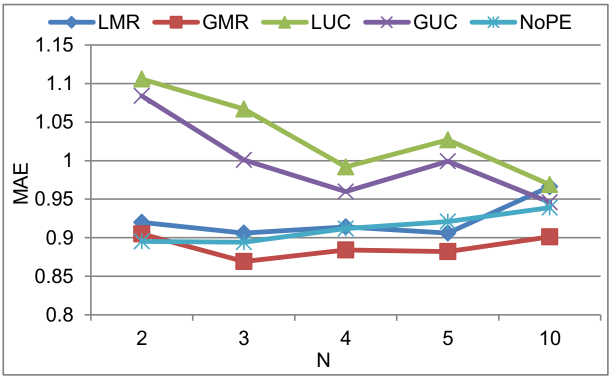

Fig. 2 MAE evaluation of the proposed method with I = 10 and different values of N

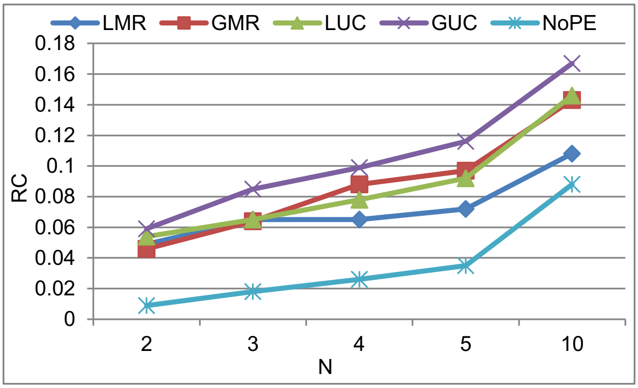

Fig. 3 RC evaluation of the proposed method with I = 10 and different values of N

changed from 2 to 10. Therefore, it can be concluded that the higher value of the parameter N has a positive effect on the performance of the recommendation methods according to the MAE metric. Moreover, these results show that the GMR method is the best performer among all compared recommendation methods based on the MAE metric for different values of the parameter N. On the other hand, the LUC method obtains the worth results based on the MAE metric in comparison to other recommendation methods. Fig. (3) shows the results of experiments based on the RC metric and different values of the parameter N. These results reveal that the performance of the recommendation methods is enhanced by increasing the value of the parameter N. Therefore, the higher value of the parameter N has a positive effect on the performance of the recommendation methods according to the RC metric. Moreover, it can be seen from these results that the GUC method obtains the best RC values in comparison to other recommendation methods for different values of the parameter N. The NoPE method is the worth performer among the compared recommendation methods based on the RC metric. It is expected as the NoPE method performs no rating profile expansion strategy in the recommendation process.

Moreover, several experiments are performed for sensitivity analysis of the parameter α. The parameter α controls the effect of the demographic and Cosine similarity functions (See

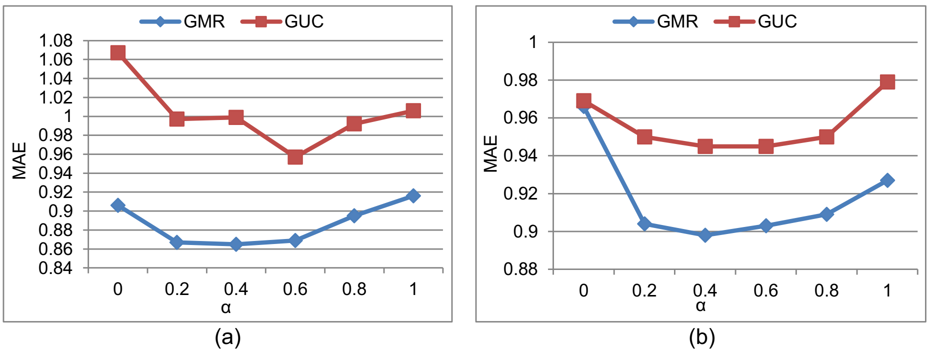

Fig. 4 Sensitivity analysis of the parameter α on the MAE metric for I = 10 (**a**) $N = 3$, (**b**) $N = 10$

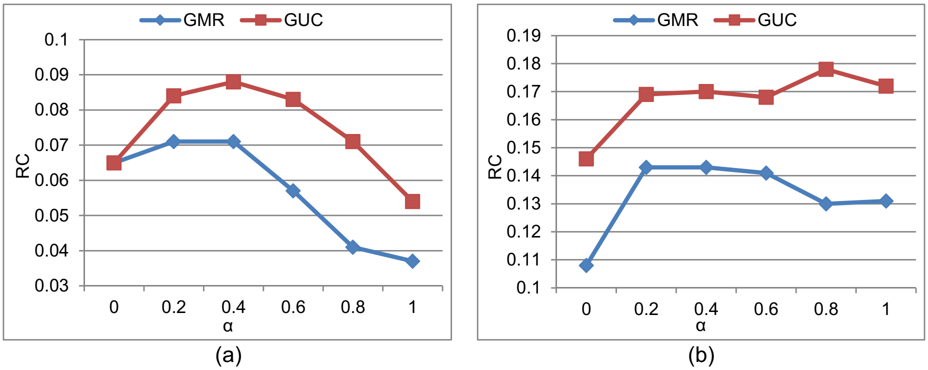

Fig. 5 Sensitivity analysis of the parameter α on the RC metric for I = 10 (**a**) $N = 3$, (**b**) $N = 10$

Eq. (3)). The value of the parameter α can be changed from 0 to 1. Figs. (4) and (5) show the results of the experiments according to different values of the parameter α based on the MAE and RC metrics, respectively. As we can see from Fig. 4(a), the GMR method obtains better accuracy than the GUC method when $N = 3$. Moreover, the best values of the MAE metric for the GMR and GUC methods are resulted when $\alpha = 0.4$ and $\alpha = 0.6$, respectively. Fig. 4(b) indicates that the GMR method obtains better accuracy than the GUC method when the value of the parameter N is equal to $N = 10$. Also, based on these results it can be concluded that both of the GUC and GMR methods obtain the best value of the MAE metric when $\alpha = 0.4$. The experiments are repeated for sensitivity analysis of the parameter α based on the RC metric and the results are reported in Fig. (5). It can be concluded from these results that, the GUC method obtains better results than the GMR method based on the rating coverage metric. Moreover, the best values of the RC metric for both of the GUC and GMR methods are resulted when $\alpha = 0.4$ for $N = 3$ (See Fig. 5(a)). On the other hand, Fig. 5(b) shows that the best values of the RC metric are obtained for the GUC and GMR methods when $\alpha = 0.8$ and $\alpha = 0.2$, respectively.

Several scenarios are considered to experimentally show the sensitivity analysis of the user's demographic data weights (i.e. w_1 for attribute age, w_2 for attribute gender, and w_3 for attribute occupation) on the performance of the proposed method. To this end, a number of experiments are performed based on the different scenarios of the demographic data weights and the MAE and RC metrics are computed for each scenario. Table 3 shows the weights of the user's demographic data in different scenarios. Every combination of the weights determines specific effects of the parameters in which the proposed method pays corresponding attentions to them. For instance, scenario 2 indicates that the system pays more attention on the "age" attribute than the other attributes.

Table 3 the weights of the user's demographic data in different scenarios

Scenarios	Weights
Scenario 1	$w_1 = 0.33$, $w_2 = 0.34$, $w_3 = 0.33$
Scenario 2	$w_1 = 0.6$, $w_2 = 0.3$, $w_3 = 0.1$
Scenario 3	$w_1 = 0.3$, $w_2 = 0.6$, $w_3 = 0.1$
Scenario 4	$w_1 = 0.3$, $w_2 = 0.1$, $w_3 = 0.6$

On the other hand, the importance of the all attributes is equal in scenario 1 and the system has a fair view about all the demographic attributes. Fig. (6) shows the experimental results of the different scenarios for the proposed method where $N = 3$ and $I = 10$. It can be concluded from Fig. 6(a) that, the GMR scheme obtains the best MAE value in the scenario 2. As the attribute "age" has the highest importance in the scenario 2, it can be concluded that this attribute has a significant effect on the improvement of the proposed method based on the MAE metric. Moreover, the MAE values of the GUC scheme are higher than the MAE values of the GMR scheme in all of the used scenarios. Fig. 6(b) shows that the GUC scheme obtains the best RC value in scenario 4 compared to the other scenarios. In scenario 4, the weight of the attribute "occupation" is higher than the other attributes. Thus, these results show that the attribute "occupation" has a positive effect on the performance of the proposed method according to the RC metric.

5 Conclusions

In this paper, we proposed a novel hybrid recommendation method based on profile expansion techniques to improve cold start problem in recommender systems. In the proposed method, the user's demographic data is used beside profile expansion techniques in order to enrich neighbors set of users. To this end, a combined similarity function is considered which is based on the demographic data and user-item rating matrix. Also, the rating profiles of users are expanded using two different techniques to alleviate cold start problem in recommender systems. Several experiments were conducted on the MovieLens dataset to show the efficiency of the proposed method compared to the other recommendation methods. The results of the experiments showed significant improvement of the proposed method in comparison to other recommendation approaches. Some directions can be considered to improve the proposed method as future works. For instance, using a clustering algorithm based on ratings and demographic data of the users can be useful to improve the profile expansion mechanism. Moreover, a novel mechanism can be proposed to use the demographic data with optimal weights for improving the results of using these sources in profile expansion methods.

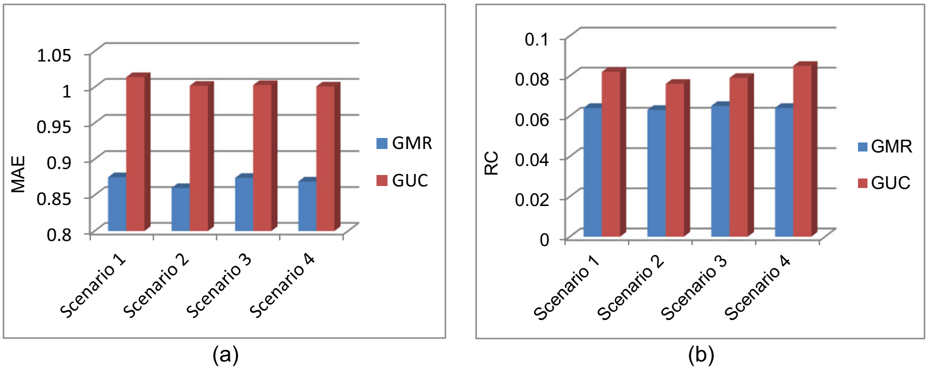

Fig. 6 The effect of different scenarios on the performance of the proposed method when I = 10 and $N = 3$ (**a**) MAE metric (**b**) RC metric

References

1. Aghdam MH, Analoui M, Kabiri P (2016). Modelling trust networks using resistive circuits for trust-aware recommender systems. J Inf Sci:0165551516628733
2. Ahmadian S, Afsharchi M, Meghdadi M (2019) An effective social recommendation method based on user reputation model and rating profile enhancement. J Inf Sci 45(5):607–642
3. Ahmadian S, Afsharchi M, Meghdadi M (2019) A novel approach based on multi-view reliability measures to alleviate data sparsity in recommender systems. Multimed Tools Appl 78(13):17763–17798
4. Ahmadian S, Joorabloo N, Jalili M, Meghdadi M, Afsharchi M, Ren YA, Temporal Clustering Approach for Social Recommender Systems (2018) IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). Barcelona, Spain: IEEE 2018:1139–1144
5. Ahmadian S, Joorabloo N, Jalili M, Ren Y, Meghdadi M, Afsharchi M (2020) A social recommender system based on reliable implicit relationships. Knowl-Based Syst 192:1–17
6. Ahmadian S, Meghdadi M, Afsharchi M (2018) A social recommendation method based on an adaptive neighbor selection mechanism. Inf Process Manag 54(4):707–725
7. Ahmadian S, Meghdadi M, Afsharchi M (2018) Incorporating reliable virtual ratings into social recommendation systems. Appl Intell 48(11):4448–4469
8. S Ahmadian, P Moradi, Akhlaghian F (2014). An improved model of trust-aware recommender systems using reliability measurements. 6th Conference on Information and Knowledge Technology (lKT 2014). Shahrood University of Technology, Tehran, Iran: IEEE; . p. 98–103
9. Ahn HJ (2008) A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. Inf Sci 178(1):37–51
10. Al-Shamri MYH, Bharadwaj KK (2007). A compact user model for hybrid movie recommender system. Conference on Computational Intelligence and Multimedia Applications, 2007 International Conference on: IEEE. p. 519–24
11. Anand D, Bharadwaj KK (2011) Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities. Expert Syst Appl 38(5):5101–5109
12. Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. Knowl-Based Syst 46:109–132
13. Breese JS, Heckerman D, Kadie C (1998). Empirical analysis of predictive algorithms for collaborative filtering. Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence: Morgan Kaufmann Publishers Inc.. p. 43–52
14. Candillier L, Meyer F, Fessant F (2008). Designing specific weighted similarity measures to improve collaborative filtering systems. Industrial Conference on Data Mining: Springer. p. 242–55
15. Z Cheng, J Shen, L Nie, TS Chua (2017). Kankanhalli M. Exploring user-specific information in music retrieval. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Shinjuku, Tokyo, Japan: ACM. p. 655–64
16. Corbellini A, Mateos C, Godoy D, Zunino A, Schiaffino S (2015) An architecture and platform for developing distributed recommendation algorithms on large-scale social networks. J Inf Sci 41(5):686–704
17. Cui C, Shen J, Nie L, Hong R, Ma J (2017) Augmented collaborative filtering for sparseness reduction in personalized POI recommendation. ACM Trans Intell Syst Technol 8(5):1–23
18. Formoso V, FernáNdez D, Cacheda F, Carneiro V (2013) Using profile expansion techniques to alleviate the new user problem. Inf Process Manag 49(3):659–672
19. Guo G, Zhang J, Thalmann D (2014) Merging trust in collaborative filtering to alleviate data sparsity and cold start. Knowl-Based Syst 57:57–68
20. Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) 22(1):5–53
21. Jalili M, Ahmadian S, Izadi M, Moradi P, Salehi M (2018) Evaluating collaborative filtering recommender algorithms: a survey. IEEE Access 6:74003–74024
22. Lai C-H, Liu D-R, Liu M-L (2015) Recommendations based on personalized tendency for different aspects of influences in social media. J Inf Sci 41(6):814–829
23. Lika B, Kolomvatsos K, Hadjiefthymiades S (2014) Facing the cold start problem in recommender systems. Expert Syst Appl 41(4):2065–2073
24. Liu Y, Lin Z, Zheng X, Chen D (2015). Incorporating social information to perform diverse replier recommendation in question and answer communities. J Inf Sci :0165551515592093
25. Luo X, Xia Y, Zhu Q (2012) Incremental collaborative filtering recommender based on regularized matrix factorization. Knowl-Based Syst 27:271–280
26. Mazhari S, Fakhrahmad SM, Sadeghbeygi H (2015). A user-profile-based friendship recommendation solution in social networks. J Inf Sci :0165551515569651

27. P Moradi, F Rezaimehr, S Ahmadian, Jalili M (2016). A trust-aware recommender algorithm based on users overlapping community structure. Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer). Negombo, Sri Lanka: IEEE. p. 162–7
28. Nguyen A-T, Denos N, Berrut C (2007). Improving new user recommendations with rule-based induction on cold user data. Proceedings of the 2007 ACM conference on Recommender systems: ACM. p. 121–8
29. Papagelis M, Plexousakis D, Kutsuras T. (2005). Alleviating the sparsity problem of collaborative filtering using trust inferences. International Conference on Trust Management: Springer. p. 224–39
30. H. A. Rahmani, M. Aliannejadi, S. Ahmadian, M. Baratchi, M. Afsharchi, Crestani F. (2019). LGLMF: Local geographical based logistic matrix factorization model for POI recommendation. Asia Information Retrieval Symposium, AIRS 2019: Information Retrieval Technology: Springer. p. 66–78
31. Rashid AM, Karypis G, Riedl J (2008) Learning preferences of new users in recommender systems: an information theoretic approach. ACM SIGKDD Explorations Newsletter 10(2):90–100
32. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J (1994). GroupLens: an open architecture for collaborative filtering of netnews. Proceedings of the 1994 ACM conference on Computer supported cooperative work: ACM. p. 175–86
33. Ricci F, Rokach L, Shapira B (2011). Introduction to recommender systems handbook. Springer
34. Safoury L, Salah A (2013) Exploiting user demographic attributes for solving cold-start problem in recommender system. Lecture Notes on Software Engineering 1(3):303
35. Said A, Plumbaum T, De Luca EW, Albayrak S (2011). A comparison of how demographic data affects recommendation. User Modeling, Adaptation and Personalization (UMAP).7
36. Son LH (2016) Dealing with the new user cold-start problem in recommender systems: a comparative review. Inf Syst 58:87–104
37. Sridevi MM, Rao RR, Rao MV (n.d.). A Survey on Recommender System
38. Zheng N, Li Q, Liao S, Zhang L (2010) Which photo groups should I choose? A comparative study of recommendation algorithms in Flickr. J Inf Sci 36(6):733–750
39. Ziegler C-N, Lausen G (2004) .Analyzing correlation between trust and user similarity in online communities. International Conference on Trust Management: Springer. p. 251–65

Faryad Tahmasebi received his B.S. degree in computer engineering from Azad University, Ilam, Iran, in 2011, and the M.S. degree in computer engineering from University of Zanjan, Zanjan, Iran in 2016. His research interests include recommender systems and social networks analysis.

Majid Meghdadi received the B.S. degree in Mathematics and Applied Computer from University of Tehran, Iran in 1991 and the M.S. degree in Computer Engineering from University of Sharif, Iran in 1994 and the PhD degree in Electronic and Computer Education in 2010 from Gazi University, Turkey. His current research interests include data mining, recommender systems, and social networks analysis. He is as an Associate Professor of the Department of Computer Engineering, University of Zanjan, Zanjan, Iran.

Sajad Ahmadian received his B.S. degree in computer engineering from Razi University, Kermanshah, Iran, in 2011, and the M.S. degree in computer engineering, artificial intelligence from University of Kurdistan, Sanandaj, Iran in 2014. Moreover, he has been received his Ph.D. degree in computer engineering, artificial intelligence from University of Zanjan, Zanjan, Iran, in 2018. He conducted a part of his Ph.D. research work in the laboratory of complex networks, RMIT University, Melbourne, Australia, form February 2018 to July 2018. His research interests include recommender systems, social networks analysis, data mining and machine learning.

Khashayar Valiallahi received his B.S. degree in computer engineering from Azad University of Malayer, Iran, in 2011, and the M.S. degree in computer engineering from University of Kurdistan, Sanandaj, Iran in 2014. His research interests include recommender systems and data minig.

Affiliations

Faryad Tahmasebi [1] · **Majid Meghdadi** [1] · **Sajad Ahmadian** [2] · **Khashayar Valiallahi** [3]

Faryad Tahmasebi
faryadtahmasebi1367@gmail.com

Sajad Ahmadian
s.ahmadian239@gmail.com

Khashayar Valiallahi
khashayar.valiallahi@gmail.com

1 Department of Computer Engineering, University of Zanjan, Zanjan, Iran

2 Faculty of Information Technology, Kermanshah University of Technology, Kermanshah, Iran

3 Department of Computer Engineering, Kurdistan University, Sanandaj, Iran