

# Improving sparsity and new user problems in collaborative filtering by clustering the personality factors

Zahra Yusefi Hafshejani<sup>1</sup> · Marjan Kaedi<sup>1</sup> · Afsaneh Fatemi<sup>1</sup>

Published online: 12 January 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** In collaborative filtering recommender systems, items recommended to an active user are selected based on the interests of users similar to him/her. Collaborative filtering systems suffer from the ‘sparsity’ and ‘new user’ problems. The former refers to the insufficiency of data about users’ preferences and the latter addresses the lack of enough information about the new-coming user. Clustering users is an effective way to improve the performance of collaborative filtering systems in facing the aforementioned problems. In previous studies, users were clustered based on characteristics such as ratings given by them as well as their age, gender, occupation, and geographical location. On the other hand, studies show that there is a significant relationship between users’ personality traits and their interests. To alleviate the sparsity and new user problems, this paper presents a new collaborative filtering system in which users are clustered based on their ‘personality traits’. In the proposed method, the personality of each user is described according to the big-5 personality model and users with similar personality are placed in the same cluster using K-means algorithm. The unknown ratings of the sparse user-item matrix are then estimated based on the clustered users, and recommendations are found for a new user according to a user-based approach which relays on the interests of the users with similar personality to him/her. In addition, for an existing user in the system, recommendations are offered in an item-based approach in which the similarity of items is estimated based on the ratings of users similar to him/her in personality. The proposed method is compared to some former collaborative filtering systems. The results demonstrate that in facing the data sparsity and new user problems, this method reduces the mean absolute error and improves the precision of the recommendations.

---

✉ Marjan Kaedi  
kaedi@eng.ui.ac.ir

<sup>1</sup> Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran

**Keywords** Recommender systems · Collaborative filtering · User's personality · Clustering

## 1 Introduction

During the past two decades, recommender systems have played an important role in helping users deal with the problem of information overload, especially in electronic commerce. Recommender systems help users to find their desired items from an enormous mass of available information [1].

There are six main categories of recommender systems [2–4]: content-based systems, collaborative filtering systems, demographic recommender systems, knowledge-based recommender systems, community-based recommender systems, and hybrid systems. In the content-based method, items similar to the ones a given user (called an active user) preferred in the past are recommended to him/her, whereas in collaborative filtering the items are proposed based on the interests of the users similar to the active user [3, 5]. In demographic recommender systems, the demographic information of users are applied for recommending the items. Knowledge-based systems propose items to users based on the knowledge about the usefulness of items for users. In community-based recommender systems, the preferences of the user's friends are considered for recommendation of items to users. In hybrid systems, a combination of the mentioned techniques is applied [6]. Beside the mentioned six categories of recommender systems, deep learning has recently been applied in recommender systems to learn patterns from the users' profiles [7], items contents [8–10], or both [11]. Experiments show that if these sources of data are available, deep learning improves the recommendations quality [7–11]. However, in some applications the data required for deep learning is not available.

Collaborative filtering approach is a widely used technique in many recommender systems and is taken as the basis of many systems [3, 12]. There are two types of collaborative filtering systems: user-based collaborative filtering and item-based collaborative filtering. In the user-based collaborative filtering, users similar to the active user are found and the items which appeal to those users are recommended to the active user. In the item-based approach, the system finds items similar to the items which have been given high ratings by the active user and the similarity of items is measured based on the rating given by the users similar to the active user [13, 14].

In addition to the ratings given by the users, other information resources have also been used to find similar users in collaborative filtering. Users' personality information is one of those resources and recently has attracted considerable attention. Studies demonstrate a significant relationship between individuals' personality traits and their interests [15–17]. According to these findings, it is expected that recommender systems using these factors gain a better understanding of users and provide them with better services [15]. User's personality can be identified through implicit or explicit techniques. So far, some studies have been

conducted on the use of users' personality in collaborative filtering recommender systems. These studies will be reviewed in Sect. 3.

However, like any other system, recommender systems face some issues and challenges. The two major challenges they deal with are: 'new user' and 'rating sparsity'. The *new user* challenge arises due to the lack of enough information about the new-coming user. The data *sparsity* challenge refers to the insufficiency of data about users' preferences and priorities [12, 18].

Due to the wide application of collaborative filtering recommender systems during the past decade, researchers have investigated new methods to cope with the drawbacks of these systems [3, 12]. One group of these methods focuses on clustering the items or users or both of them [12, 19]. Some of the studies conducted on the use of clustering in collaborative filtering systems will be reviewed in Sect. 3. By clustering the users, users similar to one another in some aspects, are detected and placed in the same cluster. Next, the blank elements of information about the preferences and priorities of a user is filled by the information extracted from the users similar to him/her. The features applied for clustering the users directly affect the performance of clustering approach; If these features strongly relate to the preferences and priorities of users with the ability to place the users with similar preferences in the same cluster, the clustering approach becomes highly effective in addressing the 'new user' and 'rating sparsity' problems. Considering the significant correlation between individuals' personality and their interests [15–17], users' personality factors seem to be suitable features for clustering the users to tackle the 'new user' and 'rating sparsity' problems. However, to the best of the authors' knowledge, the users' personality traits have not yet been applied as the attributes for clustering the users in collaborative filtering.

This study is an attempt to develop a new method in which clustering of users is performed based on their personality traits in order to improve the performance of the recommender systems while facing the *new user* and data *sparsity* challenges.

The remainder of the paper is organized as follows: Sect. 2 introduces the big-5 personality model. Section 3 presents the previous works on application of clustering methods in recommender systems as well as the previous studies on application of personality traits in recommender systems. In Sect. 4, the proposed method is explained in detail. In Sect. 5, the proposed method is evaluated and compared to other methods. Finally, in Sect. 6, conclusion and the guidelines for future research are offered.

## 2 Personality estimation using big-5 personality model

Personality can be considered as a pattern of permanent characteristics followed by every person in his/her behavior [20–25]. There exist several studies on estimating the personality of users in an implicit manner. For example, the user's personality is estimated by analyzing the contextual information available in the user's profiles (age, gender, interests, posts, and weblogs) [16–20], the structural information of online social network graph [26–30], or both the contextual and structural information [31–34]. Considering the fact that implicit methods of identifying

users' personality traits are complex and time-consuming and have unreliable precision, this study, like the majority of studies conducted in this connection, uses the explicit method and questionnaire to identify users' personality traits. One of the best-known personality models recognized in psychological studies is the big-5 personality model [30]. A great number of empirical studies have employed this model. Based on big-5 model, individuals' personalities can be studied in five main factors: neuroticism, extroversion, conscientiousness, agreeableness and openness to experience [35, 36]. NEO-PI and NEO-FFI questionnaires are used to estimate the personality according to big-5 model [37]. These questionnaires contain descriptive statements like "I am not a worrier" and "I really enjoy talking to people", which are rated on a Likert-type scale. The NEO-PI consists of 180 items to compute the scores of the five factors and the 30 subfactors (named facets). The NEO-FFI questionnaire contains 60 items selected from the NEO-PI used only to calculate the scores of the five factors (12 questions per aspect). The NEO-FFI is applied in situations where time is limited but reliable, accurate, and global measures about the five personality factors are required [38, 39]; therefore, in the current study, a dataset based on the NEO-FFI questionnaires [37] has been employed. Some sample items from the NEO-FFI questionnaire are presented in "Appendix".

### 3 Literature review

In this section, we first review the previous studies conducted on the use of clustering in the recommender systems. Then we investigate the previous studies related to the application of users' personality in the recommender systems.

#### 3.1 Studies on the use of clustering in the recommender systems

Clustering is a method for grouping data based on their similarity or proximity. In recommender systems, items or users or both of them could be clustered. Various studies have been conducted on the use of clustering in collaborative filtering systems. Some of those studies are reviewed in this section.

To improve the problem of data sparsity, Xue et al. [40] clustered the users based on their ratings using K-means algorithm. Afterwards, based on this clustering, they performed smoothing and neighbor selection and then, they filled the blank elements of the user-item matrix. Evaluations indicated the high scalability of their method. Besides, under the condition of data sparsity, their method showed less error compared to other methods. Pu Wang [41] clustered the users based on their ratings using the K-means algorithm so as to address the sparsity and scalability problems in collaborative filtering despite the nonlinear growth of number of users and items. Ullah et al. [42] developed a method in which, first, the users were clustered based on their demographic information, and ratings, and then, using this clustering, the unknown elements of the user-item matrix were filled. Afterwards, a temporal aware hybrid model was developed for each user based on which the user's neighbors were identified. Finally, in the light of the temporal information, the user's neighbors, and the user-item matrix, some recommendations were offered

to the user. Experiments show that their hybrid method has been able to alleviate the new user and data sparsity problems as well as scalability. Li [43] has clustered both the item and users. In Li's work, the users were clustered using the K-means algorithm on the basis of their ratings, and, afterwards, the unknown ratings for an item in the user-item matrix are filled based on the mean value of rating of the central user for that item. In the next step, the recommendations are offered to users based on the performed clustering. The results demonstrated that this method alleviates the problems of data sparsity and scalability. Altingovde et al. [44] showed that, due to overgeneralization, the cluster-based collaborative filtering methods which only use the information of cluster centroids offer less personalized recommendations and enjoy low precision. For that reason, they suggested that, in order to estimate the ratings, not only the centroids but also all other members within the clusters must be considered. The results demonstrated the high precision and acceptable scalability of their method. To cope with the new user problem Yanxiang et al. [45] clustered users based on characteristics such as age, gender, and occupation. Afterwards, rather than predicting the ratings of each item for each user, they calculated the mean of the ratings given to each item by all the users belonging to each cluster. The experimental results were indicative of the considerable improvement of this method in facing the new user problem in comparison with some former methods. Moradi Dakhel and Mahdavi [46], used K-means clustering and neighbor's voting in collaborative filtering. To offer recommendations to the active user, they chose  $n$  most similar neighbors from the closest cluster to the active user. Afterwards, to predict the ratings of the active user, they used the voting algorithm. The evaluation results demonstrated that this method is faster and more precise compared to some former methods.

As can be seen in the studies reviewed in this section, to the best of the authors' knowledge no previous study has clustered the users based on their personality traits.

### 3.2 Studies on the use of personality in the recommender systems

Users with different personality traits prefer items with different features; thus, personality traits of users could be a valuable information to be exploited for personalization of recommendations [17]. So far some researchers have combined users' personality aspects with recommender systems. The most common objective of these studies are the personalization of items diversity, improving the system's computational complexity, increasing the user satisfaction, and addressing the cold-start and data sparsity problems [17]. In this section, studies related to this subject are briefly reviewed.

Tintarev et al. [47] investigated the effect of openness personality trait on the user's preferred level of novelty and diversity in recommendations. They studied a group of participants who acted as the recommender system. The participants recommended items to a fictitious friend, a favorite item and the personality features of whom were known. The results indicate that people who are low on openness to experience benefit more from inter-theme (thematic) rather than inter-genre (categorical) variation. Cantador et al. [48] studied the correlations between users'

personality types and their preferences in movies, TV shows, music, and books contexts. They analyzed 53,226 Facebook users' profiles composed of their big-5 personality scores and their interests about 16 genres in each of the four mentioned contexts. They extracted some association rules to relate user's personality types to some of the genres in the considered contexts.

Roshchina et al. [49] offered a personality-based recommender system for recommending hotels to users. They estimated the users' personality traits based on the big-5 personality model by investigating the comments written by them. In another study, in order to solve the new user problem, Hu and Pu [50] used users' personality information in collaborative filtering. The three approaches proposed by them were: (1) offering recommendations to users based merely on their personality information, (2) applying a linear combination of personality and ratings of users, and (3) using a cascade approach to cover both the information resources. Afterwards, they compared these methods with the rating-based collaborative filtering and demonstrated their superiority especially that of the cascade approach. In another study, Hu et al. [51] compared the personality-based preference extraction approach with the rating-based approach. Their results demonstrated that there was no difference in the perceived precision in the two systems. However, the required effort, the time spent by the user, and the user's intention to use the system again were considerably better in the personality-based system than the rating-based recommender system.

Furthermore, Hu and Pu [52] compared the performance of rating-based, personality-based, and the hybrid (rating + personality) similarity measures for addressing the new user problem. Their findings indicate that the personality-based similarity measure exhibits better performance in addressing the new user problem when it is used together with the rating based personality measure in user-based collaborative filtering recommender systems.

To compute the similarity of users in a memory-based collaborative filtering system, Tkalcic et al. [53, 54] used personality information based on the big-5 personality model. In a series of experiments, they compared their method with a system which used only ratings, demonstrating that the use of personality to measure the similarities in a memory-based collaborative filtering decreases the computational complexity and increases the efficiency. Jeong et al. [55] used personality factors to address the problems of data sparsity and scalability in recommender systems. They used MBTI (Myers-Briggs Type Indicator) to estimate the user personality. Following investigations, they chose 'age', 'gender', and 'personality' as the most important factors in movie recommendation. Wu et al. [56] offered a method to improve the challenge of variety reduction in recommendations. They paid attention to the fact that the users' diversity requirements are related to users' personality dimensions. Afterwards, they personalized recommendations variety based on the users' personality. Experimental results have demonstrated that this system is more effective in comparison with the common methods in respect of users' subjective perceptions. Hu and Pu [57] proposed a recommender system in which the musical preferences of users are identified based on their personality. They showed that users with less information about music were more willing to use the personality-based recommender system. In another study, Lampropoulos et al.

[58] developed a cascade-hybrid music recommender system for cell phone users, which used both music genre classification and personality information. When a user sends a music file through his/her cell phone, the system offers music files of the same genre which is consistent with the previous preferences of that user as well as preferences of other users with the same personality. Braunhofer et al. [59] showed that to cope with the new user problem, the use of the users' personality information is more effective than the use of their demographic information. They also demonstrated that the use of even one factor of the big-5 personality model can improve the quality of recommendations significantly. To address the new user problem, Wang [60] used the linear combination of user-based collaborative filtering and personality-based and item-based collaborative filtering in order to predict the users' ratings. Their method could be combined with user-based collaborative filtering recommender systems to improve such systems. To tackle the cold-start problem in context-aware recommender systems, Braunhofer et al. [61] combined the demographics-based and content-based context-aware matrix factorization methods; their system switches between these two methods depending on the target recommendation situation. In another study, to address the cold start problem in context aware collaborative filtering, Braunhofer et al. [62] applied the users' personality traits in an active learning approach based on the matrix factorization model. In their proposed method, the user's personality is applied together with rating matrix to present items to the user in a continuous manner in order to obtain a high volume of ratings. In their study, the user personality is acquired through a simple questionnaire at the registration time. Fernández-Tobías et al. [63] examined the incorporation of personality in matrix factorization, active learning, and cross-domain approaches to address the new user problem. They revealed that including the users' personality in cross-domain approach leads to better results, in the condition of availability of a complementary domain.

As reviewed in this section, in a numerous studies the users' personality factors have been applied in recommender systems. However, to the best of the authors' knowledge, in none of those studies have the personality factors been used for clustering the users.

## 4 The proposed method

In this section, a new method is proposed to improve the recommender systems in order to deal with the data sparsity and new user challenges. For this purpose, users' personality information and users' ratings are used as two information resources to alleviate the aforementioned problems in the following steps:

- Upon first entrance into the system, the users fill out the NEO-FFI questionnaire and the users' five personality factors are identified based on the big-5 personality model.
- The users are clustered based on their personality factors using K-means algorithm.

- To alleviate the data sparsity, the item-based collaborative filtering algorithm is applied on each cluster separately and the unknown ratings of users in each cluster are estimated based on the ratings of other users in that cluster.
- To alleviate the new user problem, first, the closest cluster to the new user is identified. Then, the user-based collaborative filtering is applied on data of that cluster. This is actually an individualistic strategy described in [44]. Therefore, the favorite items of users whose personalities are similar to the new user's personality are offered to the new user.
- For an active user who already has rated some items in the system, the closest cluster to him/her has been detected previously. Whenever there is a need for offering recommendations to him/her, the item-based collaborative filtering are applied on the ratings of the users belonging to that closest cluster.

The proposed strategy includes two stages: the pre-processing stage and the online stage. The pre-processing stage is performed prior to the need for recommendation and the online stage is performed whenever there is a need for recommendation. These two stages will be introduced later on. The stages of the proposed method are demonstrated in Fig. 1. The input and output data of each stage is depicted in this diagram.

#### 4.1 Pre-processing stage

In the pre-processing stage of the proposed method, the processes of users clustering, estimation of the unknown ratings, and the estimation of similarities of items are performed. These processes will be explained later on. The pre-processing stage is repeated iteratively in relatively long intervals.

##### 4.1.1 Users clustering

In the proposed method for clustering the users based on their personality, the K-means algorithm is used. The five personality factors of each user obtained from the NEO-FFI questionnaire form a five-dimensional vector. Each dimension corresponds to one factor of the big-5 personality model and is represented by a number between 1 and 7. The vector  $p_u = (p_u^1, p_u^2, \dots, p_u^5)$  describes the normalized personality of user  $u$ . The normalized personality vectors of users are considered as the input of the K-means clustering algorithm and the clusters centroids are determined. The studies show that the five factors of big-5 personality model are not independent but they are correlated to each other [64–68]. Therefore, being inspired by the former studies [17, 50, 52], the Pearson's correlation coefficient [69] is taken as a criterion to measure the similarity between a user and a cluster centroid.

Equation (1) measures the similarity between the personality of a user  $u$  and a cluster centroid  $i$  (represented by  $v_i$ ).



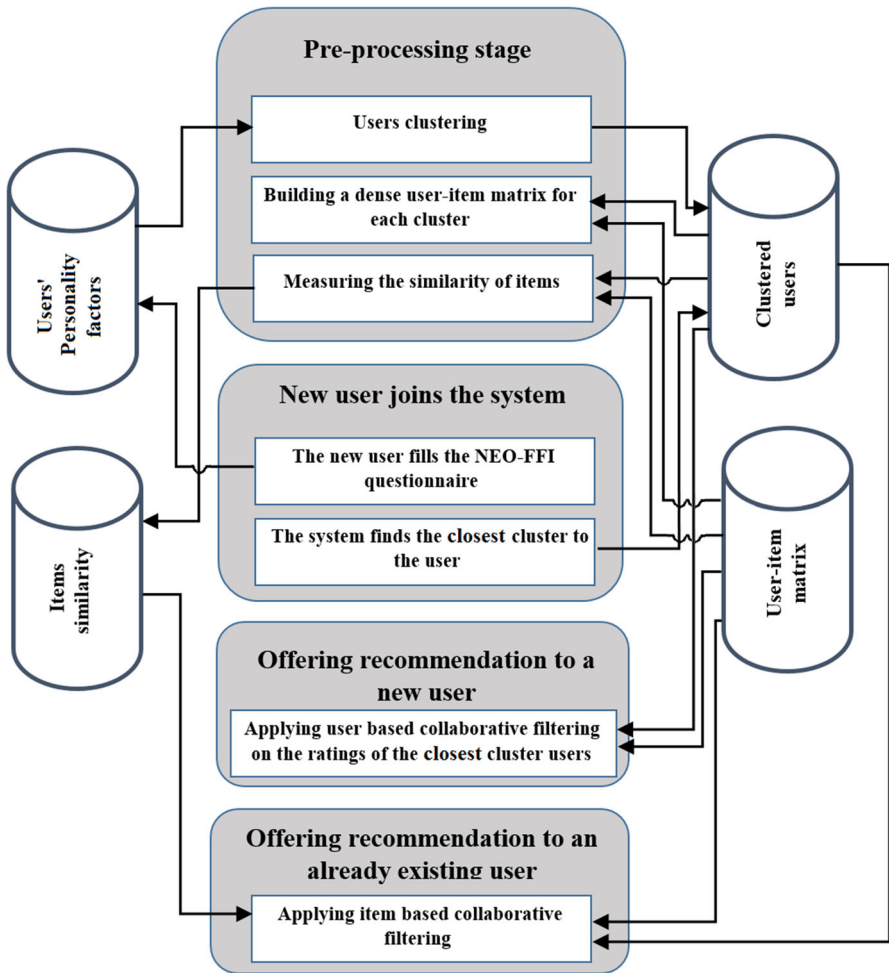


Fig. 1 Stages of the proposed method

$$\text{simp}(u, v_i) = \frac{\sum_{k=1}^5 (p_u^k - \bar{p}_u)(p_{v_i}^k - \bar{p}_{v_i})}{\sqrt{\sum_{k=1}^5 (p_u^k - \bar{p}_u)^2 \sum_{k=1}^5 (p_{v_i}^k - \bar{p}_{v_i})^2}} \quad (1)$$

In the above equation, the similarity between the new user  $u$  and  $v_i$  is represented by  $\text{simp}(u, v_i)$ .  $\bar{p}_u$  stands for the mean of five personality traits of user  $u$  and  $\bar{p}_{v_i}$  stands for the mean of five personality traits of the cluster centroid  $v_i$ .  $p_u^k$  represents the  $k$ th trait of the personality vector of user  $u$  and  $p_{v_i}^k$  represents the  $k$ th trait of cluster centroid  $v_i$ . The value obtained by this equation falls in the interval  $[-1, 1]$  and the closer it is to 1 the greater the similarity.

Actually the main difference between the proposed methods and the previous clustering-based recommender systems lies in the fact that clustering of users in the proposed method is based on the users' personality dimensions.

#### 4.1.2 Building a dense user-item matrix for each cluster

Following the clustering, the next step is to fill in the unknown elements in the user-item matrix. For this purpose, the unknown ratings of the users of each cluster are estimated using the weighted mean of the ratings of the other users belonging to that cluster; thus, the sparse matrix becomes a dense matrix of which, all entries are filled with actual or estimated ratings.

Considering that the estimation of unknown ratings of the users is based on the ratings of users with similar personalities, it is expected that the estimation be more precise than the case where estimation is based on the ratings of all the users.

#### 4.1.3 Measuring the similarity of items

In the proposed method, the similarity of items must be measured to be used in the item-based collaborative filtering approach which is adopted to offer recommendations to the existing users. To speed up the proposed method, this step is performed offline. For this to be done, the adjusted cosine similarity measure [70, 71] is applied on the ratings of each cluster members so as to measure the similarity of each pairs of items based on the ratings of that cluster. Therefore, each pair of items may have different similarity degrees from the viewpoint of members of different clusters. These measured similarities will be used to find the neighbors of user's favorite items in order to offer recommendations.

## 4.2 Online stage

The online stage is run when there is a need for recommending items to an active user. In this stage, the ratings of the active user to the items which have not been rated previously are predicted. Afterwards, the items with highest ratings are recommended to the user. In the online stage, the process of recommendation to the new user is different from that to other users. These two processes will be explained later on.

#### 4.2.1 Offering recommendations to an already existing user

For an already existing user, the unknown ratings of the items are estimated tentatively in the pre-processing stage to alleviate the data sparsity. In this stage, however, the ratings are estimated more precisely. As mentioned, in the proposed method the item-based approach is adopted to offer recommendations to an already existing user. The closest cluster to the user was identified in advance (see Sect. 4.1.1). This cluster is called  $c$  here. In this process, at first, among the items which the active user has not encountered, the items similar to his/her favorite items are identified. For this purpose, the similarities measured based on that cluster in

pre-processing stage is used. Afterwards, the active user's ratings for the similar items are predicted. The predicted ratings assigned by user  $u$  to item  $i$ ,  $R'_{u,i}$  is calculated through Eq. (2) [70]:

$$R'_{u,i} = \frac{\sum_{j \in S} (sim_c(i,j) \times R_{u,j})}{\sum_{j \in S} (|sim_c(i,j)|)} \quad (2)$$

where,  $S$  is the set of all the items rated by user  $u$ ,  $R_{u,j}$  is the rates assigned by user  $u$  to item  $j$ , and  $sim_c(i, j)$  is the similarity between items  $i$  and  $j$  calculated based on the ratings of the users belonging to cluster  $c$ . As mentioned in Sect. 4.1.3, the similarity of each pair of items is estimated through the adjusted cosine similarity measure [70, 71] on the users' ratings in the pre-processing stage.

Finally,  $N$  items which have received higher ratings based on this prediction are recommended to the active user.

#### 4.2.2 Offering recommendations to a new user

As mentioned before, in the proposed method, the user-based collaborative filtering is used in order to offer recommendations to a new user. For this purpose, first the closest cluster to the new user is identified and then, considering the rates given by the users in that cluster, the ratings of the new user to items are predicted and some recommendations are offered to the new user. In what follows, these two steps are described.

- Determining the closest cluster to the new user

When the new user fills out the personality questionnaire, his/her 5-dimensional personality vector is determined. The cluster centroids, which have been determined in the pre-processing stage using the K-means algorithm, are representatives of the common personality traits of the users in that cluster. Hence, to find the closest cluster to the new user, the similarity between the normalized personality vector of the new user and each cluster centroid is measured using Pearson's correlation coefficient [69] [see Eq. (1)]. After measuring the similarity between cluster centroids and the new user, the cluster with the highest similarity is selected. This selected cluster is called  $c_u$ .

- Prediction of new user's ratings

Equation (3) is used to predict the rate given to item  $i$  by the new user  $u$ . Using this equation, the weighted mean of rates given by the users in cluster  $c_u$  (i.e., the closest cluster to the new user  $u$ ) to item  $i$  is calculated.

$$r_{u,i} = \frac{\sum_{y \in c_u} w_{yi} \cdot r_{y,i}}{\sum w_{yi}} \quad (3)$$

$r_{u,i}$  is the predicted rate that the new user  $u$  will give to item  $i$ .  $r_{y,i}$  is the rate that user  $y$  from cluster  $c_u$  has given to item  $i$ .  $c_u$  represents the closest cluster to user  $u$ .  $w_{yi}$  is

a coefficient which distinguishes between actual and estimated ratings; If item  $i$  is actually rated by user  $y$ , then  $w_{yi}$  is equal to a pre-defined value  $\alpha$ . If the actual rating of user  $y$  to item  $i$  is unknown but it is estimated in the pre-processing stage for densification of user-item matrix, then  $w_{yi}$  is equal to  $1 - \alpha$ . Parameter  $\alpha$  takes value in the interval  $(0, 1)$  and its appropriate value is assessed in Sect. 5.2. After predicting the ratings for the items not encountered by the new user, a list of  $N$  items with highest predicted ratings, which probably appeal to the new user, is offered to him/her.

## 5 Evaluation

In this section, first, the dataset used for evaluation is introduced. Afterwards, the effect of control parameters on the performance of the proposed method is investigated, and, finally, the proposed method is compared with some previous methods.

### 5.1 Dataset

To evaluate the proposed method, a dataset including both personality traits and users' ratings is required. Therefore, the STS (South Tyrol Suggests) dataset, a dataset collected by a research center at Bozen-Bolzano University in Italy [1, 59, 61, 62], is applied. STS contains the data collected by a context-aware recommender system which offers recommendations to the visitors of South Tyrol region in Italy. It includes 2534 ratings given by 465 users from September 18, 2013 to January 14, 2014 and the number of items here is 249. Many of the ratings accompany the contextual information about the experience. The ratings contextual information and the users' demographic information are not considered in this study; therefore, they are discarded from the dataset. Among all the users 380 users have completed the personality questionnaire. The ratings given by these 380 users (total of 1723 ratings) supplemented with the users' 5 personality dimensions constitute the required data for evaluation of this method; therefore, the dataset applied in our study contains two sets of data:

- The 1723 ratings given by the users to items, within 1–5 range, 1, non-favorability and 5, favorability.
- The 5 personality dimensions of each one of the 380 users according to the big-5 personality model, within 1–7 range for each personality trait.

To evaluate the methods, the proposed method is implemented in Java and the entire dataset is split into training data (70%) and test data (30%), in a random manner. Considering the fact that the personality factors of only 380 users are available in the STS dataset, we split the dataset in term of users' ratings, not in term of users. To do this, being inspired by the experimental setup of former studies [50, 70–72], 70% of the users' ratings has been selected in a random manner and applied as the training data, and the remaining ratings of users are applied as the test

data. Splitting process repeats for 10 times and the average of results are calculated over the 10 runs. Parameter  $N$ , that is the number of recommendations to each user, is considered to be 7.

### 5.2 The effect of control parameters on the precision of the proposed method

The control parameters of the proposed method consist of  $\alpha$ ,  $N_s$ , and  $K$ .  $K$  is the number of clusters in K-mean algorithm. Parameter  $\alpha$  is the weight of the estimated ratings (the weight of actual ratings is equal to  $1 - \alpha$ ) in estimating the ratings through Eq. (3) and  $N_s$  is the number of each item’s neighbors in the item-based collaborative filtering and the number of each user’s neighbors in the user-based collaborative filtering. In this section, the effect of values of these parameters on the precision of the proposed methods is investigated and the best values for these parameters are chosen for the mentioned dataset.

First, the proposed method was applied on the dataset for several values of  $K$ . The sum of squared errors (SSE) was calculated for each value of  $K$ . The most precise results were obtained for 3–7 clusters. Afterwards, the effect of different values of  $\alpha$  and  $N_s$  on the performance of the proposed method was investigated. In this evaluation, the number of clusters was taken as 3 and the value of  $\alpha$  varied from 0.1 to 0.9. Subsequently, the weight of estimated ratings varied from 0.9 to 0.1. For each value of  $\alpha$ , the values 10, 5, 15, and 20 were assumed for  $N_s$ . As illustrated by Fig. 2, the best results were obtained when  $\alpha = 0.4$  and  $N_s = 15$ . To determine the best values for the control parameters, another evaluation was conducted by

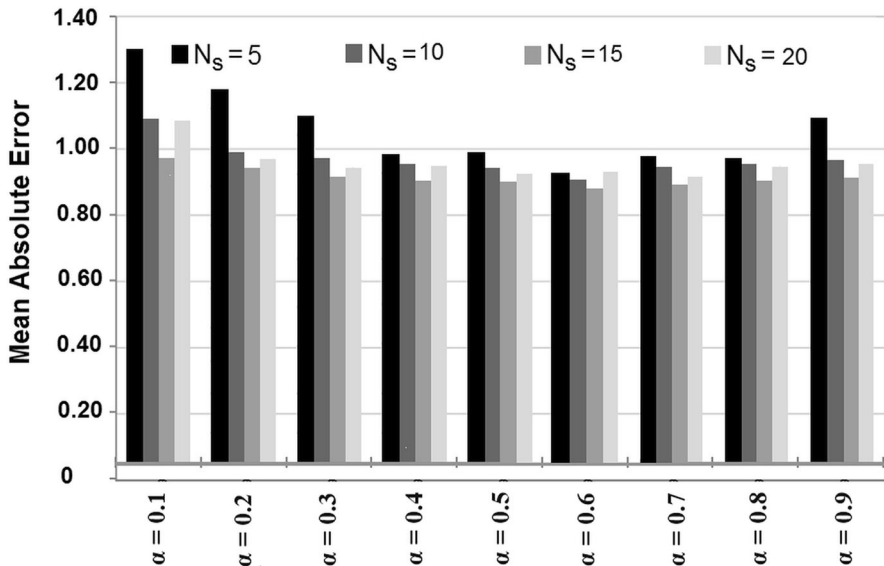
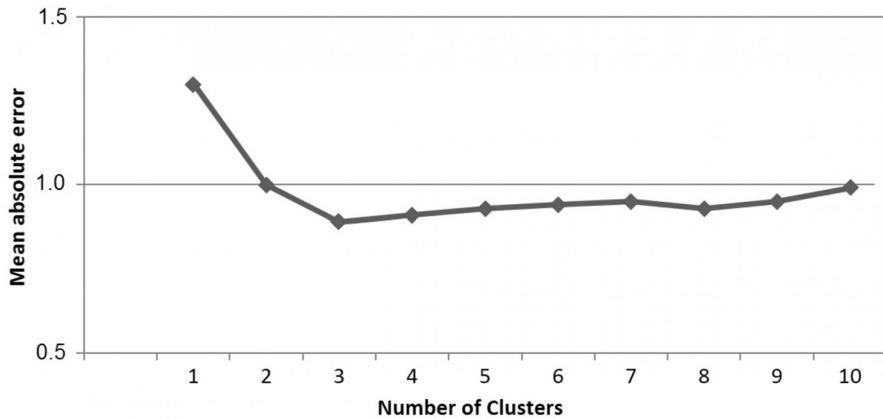


Fig. 2 Mean absolute error of the proposed method for several values of  $\alpha$  and  $N_s$



**Fig. 3** Mean absolute error of the proposed method with respect to the number of clusters ( $K$ )

changing the value of  $K$  in the case where  $\alpha = 0.4$  and  $N_s = 15$ . Figure 3 displays the results of these experiments.

In the light of these evaluations,  $K$  was determined at 3, and  $\alpha$  at 0.4. Also,  $N_s$  was determined at 15 for doing the rest of the experiments.

### 5.3 Comparison with baseline methods

In this section, we compare our proposed method with some other collaborative filtering methods. To do this, we select two traditional collaborative filtering methods and two state-of-the-art methods for comparison. The applied state-of-the-art methods have similarities with our proposed method; one of them is personality-based and the other one is clustering-based. The mean absolute error and precision criteria are used for comparing the performance of these algorithms in alleviating the data sparsity and the new user problems.

In these evaluations, the new method is referred to as CLPCF and the other methods used for comparison are as follows:

- The first method is the traditional user-based collaborative filtering (UbCf) where, ratings are predicted based on the entire collection of ratings. Pearson's correlation coefficient is applied for computing the similarity among users. For handling the new users who have recently joined the system, the average of ratings of all users for any of the items is calculated and used for providing recommendations. For handling the sparsity problem, user-based collaborative filtering algorithm runs in an offline stage and the unknown elements of sparse user-item matrix are filled with the weighted mean of other users' ratings.
- The second method is the traditional item-based collaborative filtering (IbCf) where, Pearson's correlation coefficient is applied as the similarity measure and to provide recommendations for new users, and like as UbCf, the averaged ratings of the items are used. Item-based collaborative filtering algorithm is applied in the offline stage to fill in the sparse user-item matrix.

- The third method is the personality-based collaborative filtering (PbCf) presented by Hu and Pu [49]. The PbCf follows a cascade approach. At the first stage, the 5 personality dimensions of users (according to the big-5 personality model) are applied for predicting the initial ratings for filling the blank elements of user-item matrix. Pearson's correlation coefficient is applied on the users' personality dimensions for computing the similarity among users. Afterwards, the traditional user-based collaborative filtering is applied on the obtained matrix for refining the predations obtained in the first stage. For new users, the users' personality dimensions are used in finding the similar users and predicting the ratings.
- The fourth method is a collaborative filtering where, users are clustered based on their ratings through the K-means algorithm (ClbCf) [41, 44, 70]. The unknown ratings of a user are predicted by an equation based on his/her average ratings and the ratings of the other users belonging to his/her related cluster. In this method, adjusted cosine-based similarity is applied for computing the similarity among items and the Euclidean distance is used for computing the similarity among users. For handling the new users, the averages of other users' ratings are applied.

For comparison, the parameters of the four compared algorithms should be initialized. Hu and Pu [49] evaluated the PbCf method for several values of neighbor size ranging from 5 to 100. They revealed that this method is not significantly sensitive to the number of neighbors. According to their experiments, here, the number of user's neighbors in both the stages of PbCf method is assumed to be 15. According to the experiences conducted by Wang et al. [70], the best value for the number of item's neighbors in ClbCf method is 20. Thus, in our evaluation, this parameter in ClbCf is set to 20. In addition, being inspired by the study of Wang et al. [70], the number of clusters in k-means algorithm is set to 30. The number of user's neighbors in UbCf and the number of item's neighbors in IbCf are assumed to be 5 and 10, respectively. Our experiments revealed that these values led to higher precision, specially in prediction of ratings for the new users. Parameter  $N$  (i.e., the number of recommendations) is a common parameter in all the compared methods. Our experiments showed that assigning values higher than 7 to parameter  $N$  decreases the precision metric for the considered dataset, even though it may have positive effect on some of the other metrics (e.g., recall metric). Considering the fact that our goal is to evaluate the methods based on the precision and MAE metrics, we set parameter  $N$  to be 7 in all the compared methods, as its best value.

### 5.3.1 Comparing the methods in the conditions of data sparsity

To evaluate the performance of this proposed method against different levels of data sparsity (50, 75, and 100%), the sparse user-item matrix is prepared as follows, in accordance with Hu and Pu [52, 73]:

- To have a dataset with the sparsity level of 50%, a dataset is prepared which contains 50% of ratings of all users (except for the recently joined users). In the

same manner, to have a dataset with sparsity level of 75%, a dataset is prepared which contains 25% of ratings of all users (except for the recently joined users). The obtained dataset is split into training data (70%) and test data (30%), in a random manner. Dataset preparation process runs 10 times in a random manner, and the results are averaged over the 10 runs.

- To evaluate the method in the condition of sparsity level of 100%, iteratively, one of the ratings of any of the users is selected in a random manner and applied as the test data, and entire ratings of all of the users are applied as the training data. This evaluation method is named all-but-one [52].

Figures 4 and 5 present the mean absolute error and the precision of the methods in facing different levels of data sparsity. As can be seen, an increase in sparsity level leads to an increase in mean absolute error. However, the results demonstrate that data sparsity has a low effect on the performance of the proposed method.

### 5.3.2 Comparing the methods in facing the new user

Users who have rated few items in the system are considered as new users [17, 50–53, 74]. Here, we define the new users as the users who have given less than 6 ratings. In order to evaluate in the condition of encountering new users, first, the users who have given more than 3 ratings in the system, are selected. Next, among these selected users, some users are chosen in a random manner to play the role of new users (i.e., users who have given few ratings in the system). For these purpose, being inspired by former studies [50, 52, 53], 3, 4, or 5 of their ratings are selected in a random manner as their available ratings (i.e., the training data). The other ratings of these users are considered as test data for which the prediction should be

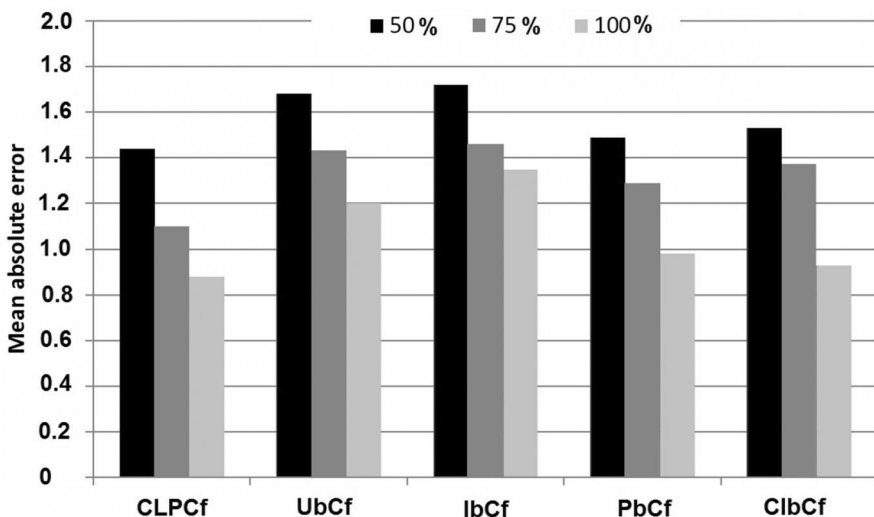


Fig. 4 Mean absolute error of the methods with respect to the density of the user-item matrix



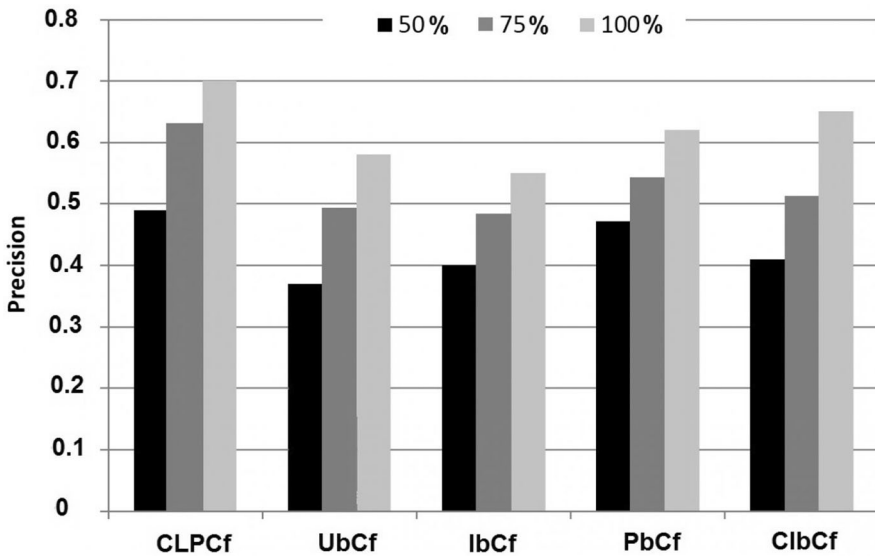


Fig. 5 Precision of the methods with respect to the density of the user-item matrix

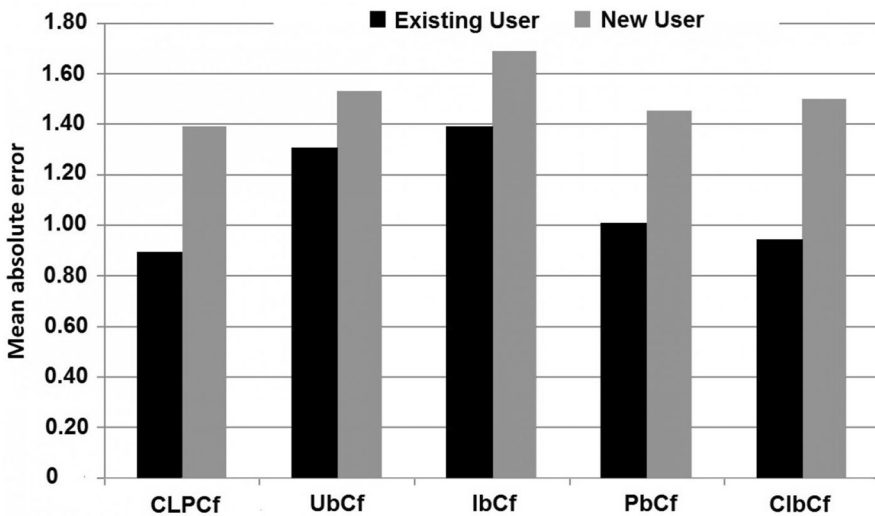
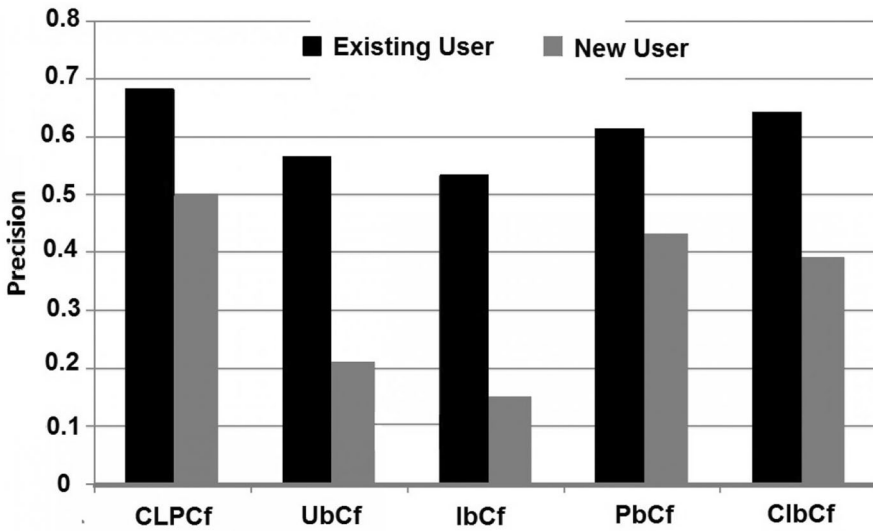


Fig. 6 Mean absolute error of the methods in prediction of ratings for existing users and new users

made. This group of users, which constitute about 10% of the entire users, are applied for the evaluation of methods in the condition of encountering new users. The process of preparing the new users dataset repeats for 10 times and the average results are reported. For evaluation of methods in condition of offering recommendations to the existing users, the entire dataset is split into training data (70%) and



**Fig. 7** Precision of the methods in prediction of ratings for existing users and new users

test data (30%), repeated for 10 times according to the process mentioned in Sect. 5.1.

The mean absolute error of the methods in predicting the ratings of the items for new users and already existing users is presented in Fig. 6 and the precision of the methods is shown in Fig. 7. As illustrated, the precision of the proposed method in facing the new user is higher and its mean absolute error is lower in comparison to the other methods.

## 6 Conclusion

In this paper, a method was proposed to improve the prediction results of recommender systems in facing the data sparsity and new user challenges. In the proposed method, users' personality factors are described based on the big-5 personality model and the users are clustered based on their personality factors. Afterwards, the unknown elements of the user-item matrix (in the condition of data sparsity) as well as the rates given to the items by the users (both the new users and the already existing users) are estimated based on the rates given by other users in the related clusters. The evaluations showed that, with regard to reducing mean absolute error and increasing the precision, the proposed method has obtained better results in comparison with the other methods.

Since personality traits are relatively stable, the need to update the clusters is less than in the case where clustering is based on ratings or any other unstable factors. Besides, finding similar users in a cluster requires less processing time and is done faster than finding them among all users. To improve the evaluations, the scalability of the proposed method could be studied accurately in future studies. Furthermore,

using lie detectors, statistical methods, etc., the accuracy of the collected personality data could be improved by identifying the inaccurate, dishonest, and fake responses [75, 76]. For example, the sabotage index [77], the fixed individualized chance score [78], and the inter-item standard deviation index [79] may be used for this purpose.

The proposed method can be compared to the techniques other than the mentioned four methods, as well. For example, the proposed method can be compared to the recommender systems that uses deep learning to learn rich features for items or users. In addition, it is suggested that other clustering methods than the K-means algorithm be used for clustering the users based on their personality factors and the results be evaluated. Besides, it is recommended that the explicit method be replaced by implicit methods to identify the users' personality. Finally, the current emotion of users can be detected and combined with their personality factors [80, 81] to improve the accuracy of the recommendations.

**Acknowledgements** The authors would like to thank Dr. Mohammad Ali Nematbakhsh for his support and assistance to improve the manuscript.

## Appendix

The NEO-FFI questionnaire comprises 60 items, 12 items to assess each of the five personality factors (six items are positively worded and six items are negatively worded for each personality factor). For each item, the respondent should describe how accurately that item describes him/her. The respondent selects the response based on the five-point Likert scale (ranging from strongly disagree to strongly agree). Then, the following scores are assigned to the answers of positively worded items:

Strongly disagree (1), Disagree (2), Neither agree nor disagree (3), Agree (4), Strongly agree (5) and the following scores are used for the negatively worded items:

Strongly disagree (5), Disagree (4), Neither agree nor disagree (3), Agree (2), Strongly agree (1).

Finally, the personality factors are assessed by summing up the scores. The NEO-FFI is a copyrighted questionnaire. Because of copyright compliance policy, only few sample statements are listed below:

Personality factor	Sample statement	Positive/negative
Neuroticism	I worry about things	Positive
Extraversion	I think a lot before I speak or act	Negative
Conscientiousness	I get chores done right away	Positive

## References

1. Elahi, M., Braunhofer, M., Ricci, F., & Tkalcic, M. (2013). Personality-based active learning for collaborative filtering recommender systems. In *AI\* IA 2013: Advances in Artificial Intelligence*, pp. 360–371, Springer.
2. Burke, R. (2007). Hybrid web recommender systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive web*, Volume 4321 of the series Lecture Notes in Computer Science (pp. 377–408). Springer, Berlin/Heidelberg.
3. Bilge, A., & Polat, H. (2013). A scalable privacy-preserving recommendation scheme via bisecting k-means clustering. *Information Processing and Management*, 49, 912–927.
4. Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39, 10059–10072.
5. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749.
6. Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (2nd ed., Vol. 54, pp. 1–34). Boston, MA: Springer US.
7. Elkahky, A., Song, Y., & He, X. (2015). A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *International World Wide Web Conference Committee (IW3C2)*, Florence, Italy, May 18–22.
8. den Oord, A. V., Dieleman, S., & Schrauwen, B. (2013). Deep content based music recommendation. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, (Eds.), *Advances in Neural Information Processing Systems 26*, pp. 2643–2651.
9. Wang, H., Wang, N., & Yeung, D. (2015). Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, Australia, August 10–13, pp. 1235–1244.
10. Wang, X., & Wang, Y. (2014). Improving content-based and hybrid music recommendation using deep Learning. In *Proceedings of the 22nd ACM International Conference on Multimedia*, Orlando, Florida, USA, November 3–7, pp. 627–636.
11. Cheng, H., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., et al. (2016). Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, Boston, MA, USA, September 15, pp. 7–10.
12. Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132.
13. Cacheda, F., Carneiro, V., Fernández, D., & Formoso, V. (2011). Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web (TWEB)*, 5, 2.
14. Dang, T. T., Duong, T. H., & Nguyen, H. S. (2014). A hybrid framework for enhancing correlation to solve cold-start problem in recommender systems. In *Seventh IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, pp. 1–5.
15. Hu, R. (2010). Design and user issues in personality-based recommender systems. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, Como, Italy, pp. 357–360.
16. Tsai, C. F., & Hung, C. (2012). Cluster ensembles in collaborative filtering recommendation. *Applied Soft Computing*, 12, 1417–1425.
17. Tkalcic, M., & Chen, L. (2015). Personality and recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (2nd ed., Vol. 54, pp. 715–739). Boston, MA: Springer.
18. Xie, F., Xu, M., & Chen, Z. (2012). RBRA: A simple and efficient rating-based recommender algorithm to cope with sparsity in recommender systems. In *26th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 306–311.
19. Ancy, S., & Unnikrishnan, V. (2015). A study on clustering and collaborative filtering. In *International Conference on Computation of Power, Energy Information and Communication (ICCPCEIC)*, Chennai, Tamilnadu, India, April 22–23, pp. 0039–0042.
20. Feist, G. J. (2013). The scientific personality. In G. J. Feist & M. Gorman (Eds.), *Handbook of the psychology of science* (pp. 95–122). New York: Springer.

21. Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, 26(6), 1289–1295.
22. Kalish, Y., & Robins, G. (2006). Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure. *Social Networks*, 28(1), 56–84.
23. Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on twitter. *Journal of Research in Personality*, 46(6), 710–718.
24. Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805.
25. Skowron, M., Ferwerda, B., Tkalčič, M., & Schedl, M. (2016). Fusing social media cues: Personality prediction from Twitter and Instagram. In *WWW'16 Companion, Proceedings of the 25th International Conference Companion on World Wide Web*, Montreal, Quebec, Canada, April 11–15, pp. 107–108.
26. Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Extended Abstracts Volume*, Vancouver, BC, Canada, May 7–12, pp. 253–262.
27. Olguun, D. O., Gloor, P. A., & Pentland, A. S. (2009). Capturing individual and group behavior with wearable sensors. In *Proceedings of the 2009 AAAI Spring Symposium on Human Behavior Modeling*, Stanford, CA.
28. Roberts, S. G., Wilson, R., Fedurek, P., & Dunbar, R. (2008). Individual differences and personal social network size and structure. *Personality and Individual Differences*, 44(4), 954–964.
29. Golkar, I., & Kaedi, M. (2016). Developing a model for estimating the extraversion degree of social network members using the information extracted from the graph structure. *Journal of Modeling in Engineering*, 13(43), 91–106.
30. Golkar Amnieh, I., & Kaedi, M. (2015). Using estimated personality of social network members for finding influential nodes in viral marketing. *Cybernetics and Systems*, 46(5), 355–378.
31. Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from Twitter. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, Boston, MA, USA, October 9–11, pp. 149–156.
32. Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., & Crowcroft J. (2012). The personality of popular Facebook users. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, Seattle, WA, USA, February 11–15, pp. 955–964.
33. Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011). Our twitter profiles, our selves: Predicting personality with twitter. In *IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, Boston, MA, USA, October 9–11, pp. 180–185.
34. Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., et al. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction (Special Issue on Personality in Personalized Systems)*, 26(2), 109–114.
35. McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 175–215.
36. Oren, T., & Ghasem-Aghaee, N. (2003). Personality representation processable in fuzzy logic for human behavior simulation. In *Proceedings of the 2003 Summer Computer Simulation Conference*, Montreal, PQ, Canada, July 20–24, pp. 11–18.
37. Costa, P. T., & MacCrae, R. R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): Professional manual*. Odessa: Psychological Assessment Resources.
38. Rosellini, A. J., & Brown, T. A. (2011). The NEO Five-factor inventory: Latent structure and relationships with dimensions of anxiety and depressive disorders in a large clinical sample. *Assessment*, 18(1), 27–38.
39. Sajadinejad, M. S., Molavi, H., Asgari, K., Kalantari, M., & Adibi, P. (2012). Personality dimensions and type D personality in female patients with ulcerative colitis. *Journal of Research in Medical Sciences*, 17(10), 898–904.
40. Xue, G. R., Lin, C., Yang, Q., Xi, W., Zeng, H.-J., Yu, Y., et al. (2005). Scalable collaborative filtering using cluster-based smoothing. In *Proceedings of the 28th Annual International ACM SIGIR*

- Conference on Research and Development in Information Retrieval*, Salvador, Brazil, August 15–19, pp. 114–121.
41. Wang, P. (2012). A personalized collaborative recommendation approach based on clustering of customers. *Physics Procedia*, 24, 812–816.
  42. Ullah, F., Sarwar, G., Lee, S. C., Park, Y. K., Moon, K. D., & Kim, J. T. (2012). Hybrid recommender system with temporal information. In *International Conference on Information Networking (ICOIN)*, Bali, Indonesia, February 1–3, pp. 421–425.
  43. Li, X. (2011). Collaborative filtering recommendation algorithm based on cluster. In *International Conference on Computer Science and Network Technology (ICCSNT)*, December 24–26, pp. 2682–2685.
  44. Altıngövdü, I. S., Subakan, O. N., & Ulusoy, O. (2013). Cluster searching strategies for collaborative recommendation systems. *Information Processing and Management*, 49, 688–697.
  45. Yanxiang, L., Deke, G., Fei, C., & Honghui, C. (2013). User-based clustering with top-N recommendation on cold-start problem. In *Third International Conference on Intelligent System Design and Engineering Applications (ISDEA)*, Hong Kong, China, January 16–18, pp. 1585–1589.
  46. Dakhel G. M., & Mahdavi, M. (2011). A new collaborative filtering algorithm using K-means clustering and neighbors' voting. In *11th International Conference on Hybrid Intelligent Systems (HIS)*, Malacca, Malaysia, December 5–8, pp. 179–184.
  47. Tintarev, N., Dennis, M., & Mastho, J. (2013). User modeling, adaptation, and personalization. In *Lecture Notes in Computer Science*, 7899, pp. 190–202.
  48. Cantador, I., Fernández-tobías, I., & Bellogín, A. (2013). Relating personality types with user preferences in multiple entertainment domains. In *EMPIRE 1st Workshop on Emotions and Personality in Personalized Services*, Rome, June 10.
  49. Roshchina, A., Cardiff, J., & Rosso, P. (2011). A comparative evaluation of personality estimation algorithms for the twin recommender system. In *Proceedings of the 3rd International Workshop on Search and Mining User-Generated Contents*, Glasgow, United Kingdom, October 24–28, pp. 11–18.
  50. Hu R., & Pu, P. (2011). Enhancing collaborative filtering systems with personality information. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, Chicago, IL, USA, October 23–27, pp. 197–204.
  51. Hu R., & Pu, P. (2009). Acceptance issues of personality-based recommender systems. In *Proceedings of the Third ACM Conference on Recommender Systems*, pp. 221–224.
  52. Hu, R., & Pu, P. (2010). Using personality information in collaborative filtering for new users. In *Proceedings of the 2nd ACM RecSys'10 Workshop on Recommender Systems and the Social Web*, New York, October 22–25, pp. 17–24.
  53. Tkalcic, M., Kunaver, M., Tasic, J., & Kosir, A. (2009). Personality based user similarity measure for a collaborative recommender system. In *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real World Challenges*, pp. 30–37.
  54. Tkalčić, M., Kunaver, M., Košir, A., & Tasič, J. (2011). Addressing the new user problem with a personality based user similarity measure. In F. Ricci, G. Semeraro, M. de Gemmis, P. Lops, J. Masthoff, F. Grasso, & J. Ham, (Eds.), *Joint Proceedings of the Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA 2011) and the 2nd Workshop on User Models for Motivational Systems: The Affective and the Rational Routes to Persuasion (UMMS 2011)*, CEUR Workshop Proceedings, 740, pp. 19–26.
  55. Jeong, W. H., Kim, S. J., Park, D. S., & Kwak, J. (2013). Performance improvement of a movie recommendation system based on personal propensity and secure collaborative filtering. *Journal of Information Processing Systems*, 9, 157–172.
  56. Wu, W., Chen, L., & He, L. (2013). Using personality to adjust diversity in recommender systems. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, Paris, France, May 1–3, pp. 225–229.
  57. Hu, R., & Pu, P. (2010). A study on user perception of personality-based recommender systems. In *User Modeling, Adaptation, and Personalization*, Volume 6075 of the series Lecture Notes in Computer Science, pp. 291–302, Springer.
  58. Lampropoulos, A. S., Lampropoulou, P. S., & Tsihrintzis, G. A. (2012). A cascade-hybrid music recommender system for mobile services based on musical genre classification and personality diagnosis. *Multimedia Tools and Applications*, 59, 241–258.
  59. Braunhofer, M., Elahi, M., & Ricci, F. (2015). User personality and the new user problem in a context-aware point of interest recommender system. In I. Tussyadiah, & A. Inversini (Eds.),

- Information and Communication Technologies in Tourism, Proceedings of the International Conference in Lugano, Switzerland*, February 3–6, pp. 537–549.
60. Wang, J. (2015). A collaborative filtering systems based on personality information. In *2015 International Industrial Informatics and Computer Engineering Conference*, Shaanxi, China, January pp. 10–11.
  61. Braunhofer, M., Codina, V., & Ricci, F. (2014). Switching hybrid for cold-starting context-aware recommender systems. In *RecSys '14, Proceedings of the 8th ACM Conference on Recommender Systems*, Silicon Valley, USA, October 6–10, pp. 349–352.
  62. Braunhofer, M., Elahi, M., Ge, M., & Ricci, F. (2014). Context dependent preference acquisition with personality-based active learning in mobile recommender systems. In P. Zaphiris, & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration*. Springer. Volume 8524 of the series Lecture Notes in Computer Science, Heraklion, Crete, Greece, June 22–27, pp. 105–116.
  63. Fernández-Tobías, I., Braunhofer, M., Elahi, M., Ricci, F., & Cantador, I. (2016). Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction (Special Issue on Personality in Personalized Systems)*, 26(2), 221–255.
  64. Sadr, M. M. (2016). The role of personality traits predicting emotion regulation strategies. *International Academic Journal of Humanities*, 3(4), 13–24.
  65. Revelle, W., Wilt, J., & Rosenthal, A. (2010). Individual differences in cognition: New methods for examining the personality-cognition link. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of individual differences in cognition* (pp. 27–49). New York: Springer.
  66. Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, 134(1), 138–161.
  67. Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, 93(5), 1042–1052.
  68. Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92(5), 1270–1285.
  69. Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009, 4–23.
  70. Wang, Q., Cao, W., & Liu, Y. (2014). A Novel Clustering Based Collaborative Filtering Recommendation System Algorithm. In *Advanced Technologies, Embedded and Multimedia for Human-centric Computing*, Volume 260 of the series Lecture Notes in Electrical Engineering, pp. 673–680, Springer.
  71. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, Hong Kong, China, May 1–5, pp. 285–295.
  72. Amatriain, X., & Pujol, J. M. (2015). Data mining methods for recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (2nd ed., Vol. 54, pp. 227–262). Boston, MA: Springer US.
  73. Ghazarian, S., & Nematbakhsh, M. A. (2015). Enhancing memory-based collaborative filtering for group recommender systems. *Expert Systems with Applications*, 42, 3801–3812.
  74. Rubens, N., Elahi, M., Sugiyama, M., & Kaplan, D. (2015). Active Learning in Recommender Systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (2nd ed., pp. 809–846). Boston, MA: Springer US.
  75. Scandell, D. J., & Wlazelek, B. G. (1996). Self-presentation strategies on the Neo-five factor inventory: Implications for detecting faking. *Psychological Reports, Part 2*, 79(3), 1115–1121.
  76. Fronczyk, K. (2014). The identification of random or careless responding in questionnaires: The example of the NEO-FFI. *Annals of Psychology*, 2, 457–473.
  77. Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the sixteen personality factor questionnaire (16PF)*. Champaign, IL: IPAT.
  78. Haertzen, C. A., & Ross, F. E. (1978). Using four chance profiles to estimate carelessness. *Psychological Reports*, 41, 1079–1087.
  79. Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences*, 84, 79–83.

80. PourMohammadBagher, L., Kaedi, M., Ghasem-Aghaee, N., & Ören, T. I. (2007). Anger evaluation for fuzzy agents with dynamic personality. *Mathematical and Computer Modelling of Dynamical Systems*, 15(6), 535–553.
81. Ghasem-Aghaee, N., PoorMohamadBagher, L., Kaedi, M., & Ören, T. I. (2007). Anger filter in agent simulation of human behavior, In *Proceedings of the 18th IASTED International Conference: Modelling and Simulation*, Montreal, QC, Canada, May 30–June 1, pp. 44–46.