Personality-Based Active Learning for Collaborative Filtering Recommender Systems

Mehdi Elahi[1], Matthias Braunhofer[1], Francesco Ricci[1], and Marko Tkalcic[2]

[1] Free University of Bozen-Bolzano, Bozen-Bolzano, Italy
{mehdi.elahi,mbraunhofer,fricci}@unibz.it
http://www.unibz.it
[2] Johannes Kepler University, Linz, Austria
marko.tkalcic@jku.at
http://www.jku.at

Abstract. Recommender systems (RSs) suffer from the cold-start or new user/item problem, i.e., the impossibility to provide a new user with accurate recommendations or to recommend new items. Active learning (AL) addresses this problem by actively selecting items to be presented to the user in order to acquire her ratings and hence improve the output of the RS. In this paper, we propose a novel AL approach that exploits the user's personality - using the Five Factor Model (FFM) - in order to identify the items that the user is requested to rate. We have evaluated our approach in a user study by integrating it into a mobile, context-aware RS that provides users with recommendations for places of interest (POIs). We show that the proposed AL approach significantly increases the number of ratings acquired from the user and the recommendation accuracy.

1 Introduction

Recommender systems (RSs) are information and decision support tools providing users with suggestions for items that are likely to be interesting to them or to be relevant to their needs [20]. A common problem of RSs is cold-start; this occurs when a new user or a new item is added to the system, but the system doesn't have enough information (e.g., ratings, purchasing records, browsing history) about them to reliably recommend any item to this new user or the new item to any user. Several approaches have been recently proposed to deal with this problem [20] but the most direct way is to rely on active learning (AL), i.e., to use an initial data acquisition and learning phase. Here the system actively asks the user to rate a set of items, which are identified using a strategy aimed at best revealing the user's interests and consequently at improving the quality of the recommendations [21,7].

In this paper, we present a novel AL rating request strategy that leverages the knowledge of the user's personality in order to predict which items a user will have an opinion about. More specifically, our approach makes use of one of the most influential models in psychology, namely the Five Factor Model (FFM)

or Big Five dimensions of personality, in which personality is conceptualized in terms of openness, conscientiousness, extraversion, agreeableness and neuroticism [3]. The rationale behind the choice of the FFM for AL is that these factors account for most of the variance among users in terms of trait terms [3]. Previous research has shown that personality influences human behaviours and that there exist direct relations between personality and tastes / interests [19]. Consequently, the incorporation of human personality into AL can help in selecting "good" items to be rated by the user. Moreover, user's personality, as it is shown in this paper, can be acquired with simple and even engaging questionnaires that any user can fill out. The assessed personality can also be illustrated to the user hence making the interaction with the recommender even more rewarding for the user.

We have formulated the following hypotheses: a) the proposed personality-based AL method leads to a higher number of ratings acquired from users compared to a state-of-the-art AL strategy, and b) our proposed AL approach compares favourably to existing state of the art AL methods in terms of achieved recommendation accuracy. To evaluate these hypotheses, we integrated our technique into a context-aware RS that recommends places of interest (POIs) to mobile users [2], and conducted a live user study. The implemented solution takes the personality characteristics of a target user as input to an extended matrix factorization model used for predicting what items the user should be requested to rate. Hence, by improving existing AL strategies, it is able to provide personalized rating requests even to users with few or no past ratings (cold-start).

The remainder of the paper is organized as follows: section 2 reviews the related work. In section 3 we explain the application scenario and in section 4, we present our proposed strategy for AL based on user's personality traits. Section 5 describes the structure of the user study while the obtained results are presented in section 6. Finally, in section 7, conclusions and future work directions are given.

2   Related Work

Active Learning in RSs aims at actively acquiring user preference data to improve the output of the RS [21,7,6]. In [17] six techniques that collaborative filtering systems can use to learn about new users in the sign up process are introduced: *entropy* where items with the largest rating entropy are preferred; *random*; *popularity*; $log(popularity) * entropy$ where items that are both popular and have diverse ratings are preferred; and finally *item-item personalized*, where the items are proposed randomly until one rating is acquired, then a recommender is used to predict the items that the user is likely to have seen. In the considered scenario, the $log(popularity) * entropy$ strategy was found to be the best in terms of accuracy.

In [18] the authors extend their former work using a rating elicitation approach based on decision trees. The proposed technique is called $IGCN$, and builds a tree where each node is labelled by a particular item to be asked to the

user to rate. According to the user elicited rating for the asked item a different branch is followed, and a new node, that is labelled with another item to rate is determined. It is worth noting that this type of techniques can be applied only to a recommender system that has already acquired a large dataset of ratings. In fact the ratings are necessary to build the above mentioned decision tree. Hence, they could solve the cold-start problem for a new user but they cannot deal with a situation where the entire system must be bootstrapped, because it has not yet interacted with enough users, as in our study.

In [8] three strategies for rating elicitation in collaborative filtering are proposed. The first method, $GreedyExtend$, selects the items that minimize the root mean square error (RMSE) of the rating prediction (on the training set). The second one, named Var, selects the items with the largest $\sqrt{popularity * variance}$, i.e., those with many and diverse ratings in the training set. The third one, called $Coverage$, selects the items with the largest coverage, which is defined as the total number of users who co-rated both the selected item and any other item. They evaluated the performance of these strategies and compared them with previously proposed ones ($popularity$, $entropy$, $entropy0$, $HELF$, and $random$). They show that $GreedyExtend$ outperforms the other strategies. However, despite this remarkable achievement, $GreedyExtend$ is static, i.e., selects the items without considering the ratings previously entered by the user.

In a more recent work of the same authors [9] an adaptive strategy for rating elicitation in collaborative filtering is proposed. Their strategy is based on decision trees where each node is labelled with an item (movie). The node divides the users into three groups based on their ratings for that movie: lovers, who rated the movie high; haters, who rated the movie low; and unknowns, who did not rate the movie. The proposed strategy has shown a significant reduction of RMSE compared with other strategies. It should be noted that their results are again rather difficult to compare with ours. They simulate a scenario where the system is trained and the decision tree is constructed from a large training dataset with millions of ratings. So they assume a large initial collection of ratings. Then, they focus on completely new users, i.e., those without a single rating in the train set. In contrast, our system had a very limited initial rating dataset with only few hundred ratings. Moreover, we analyze the performance of our system in an online study with real users while they executed only an offline study with simulated users.

Moving now to the topic of personality in RSs, we note that earlier studies conducted on the user personality characteristics support the possibility of using personality information in RSs [13,12,22]. In general, user personality can be enquired either explicitly, i.e., asking users to complete a personality questionnaire using one of the personality evaluation inventories, or implicitly, i.e., by observing users' behavioral patterns [16]. However, a previous study has shown that using explicit personality acquisition interfaces yields to better results in terms of user satisfaction, ease of use and prediction accuracy [4].

For this reason we have decided to adopt (as illustrated in the next section) the explicit approach, and in particular we have used the Ten-Item Personality
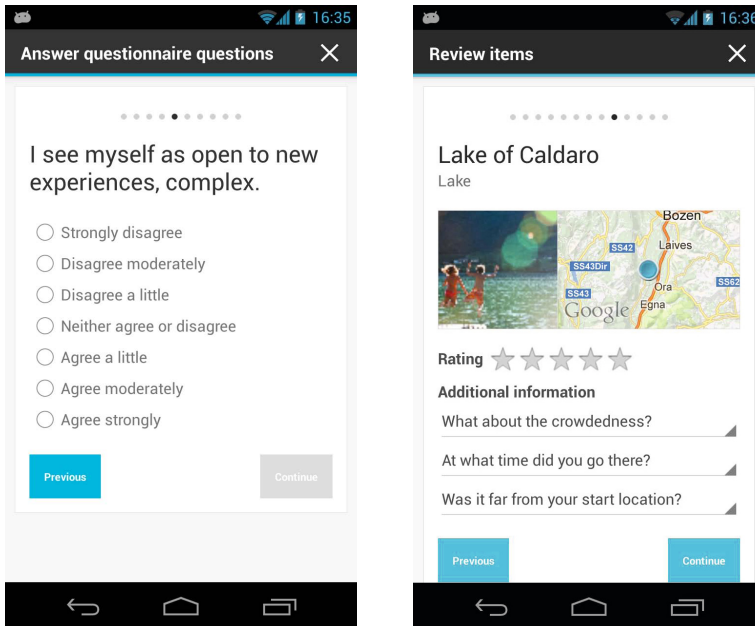
Fig. 1. Sample screenshots of the application

Inventory (10-items TIPI) which takes a few minutes to complete [11]. It represents a sensible option if personality is not the primary topic of research and it is mostly useful when time and space are in short supply (e.g., in a mobile application). We should note here that the personality model we used is a comprehensive model and thus not linked to a specific domain [19].

3   Application Scenario

Our application scenario is a mobile RS for tourists, that recommends interesting places of interests (POIs) in Alto-Adige region in Italy [2]. After the user registration, the system, by using an AL strategy, identifies and presents items to the user in order to collect her ratings (Figure 1, right). The adopted AL strategies analyze the available rating dataset, score the items estimating their ratings usefulness, and the highest scoring items are presented to the user to rate. However, in our application scenario, we encountered a severe cold-start problem, i.e., we needed to bootstrap the RS having just a small dataset with only few hundred ratings. In such a situation, standard AL strategies, as those mentioned in the previous section, fail to select useful and ratable items, and hence they can not elicit any rating from the user.

In order to cope with this problem, we have defined and proposed a novel AL strategy that incorporates an additional source of user information, i.e., user personality information. Hence, in our system before presenting any item to rate,

the Ten-Item Personality Inventory (10-items TIPI) statements are shown to the user and she has to indicate to what extent she agrees with each statement on a 7-point Likert scale (Figure 1, left). Then, the user is assigned to an experimental group, either the random or the compound AL group, and is asked to rate a number of items selected by the random AL strategy or the combination of *log(popularity) * entropy* and an original strategy based on the user's detected personality, as it is explained in section 4. Finally, the user's profile is built and the items with highest predicted ratings can be recommended to her.

3.1   Personality Questionnaire

Personality accounts for the individual differences in people's emotional, interpersonal, experiential, attitudinal and motivational styles [14]. It has been shown that personality affects human decision making and user's interests [19]. In order to learn the personality of the users in a ideal way, the system should have enough time and resources to enquire the personality information using a long questionnaire with many questions (e.g. hundreds) [10]. However, this could be difficult and one is therefore forced to rely on a much shorter questionnaire [11]. For example, in our mobile application, we could not ask the users too many questions and therefore we used the Ten-Item Personality Inventory (10-items TIPI). Figure 1 (left) shows a screen shot of our application where one of the questionnaire statements is illustrated. The full questionnaire includes the ten statements that are listed below:

1. I see myself as extraverted, enthusiastic.
2. I see myself as critical, quarrelsome.
3. I see myself as dependable, self-disciplined.
4. I see myself as anxious, easily upset.
5. I see myself as open to new experiences, complex.
6. I see myself as reserved, quiet.
7. I see myself as sympathetic, warm.
8. I see myself as disorganized, careless.
9. I see myself as calm, emotionally stable.
10. I see myself as conventional, uncreative.

For each statements the user has to indicate to what extent she agrees. When the user completes the questionnaire, a brief explanation of her personality, as those used in [1], is shown to her.

4   Active Learning Strategies

In this section we describe the AL strategies that we have used and compared in the experimental study fully described in the next section.

Log(Popularity) * Entropy scores each item i by multiplying the logarithm of the *popularity* of i (i.e., the number of ratings for i in the training set) with

the entropy of the ratings for i. Then, the top items according to the computed score are proposed to be rated by the user (4 in our experiments).

This strategy tries to combine the effect of the popularity with a score that favours items with more diverse ratings (larger entropy), which may provide more useful (discriminative) information about the user's preferences [17]. Clearly, more popular items are more likely to be known by the user, and hence it is more likely that a request for such a rating will be fulfilled by the user and will increase the size of the rating database. But many popular items in our dataset had no or only one rating, and rating-based popularity scores cannot distinguish such popular items from less popular items with similar number of ratings. Therefore, this strategy may select items that are unpopular and unknown to the user and thus not rateable.

To cope with that problem, we have designed a second strategy that tries to select the items that the user has most likely experienced, by exploiting the personality information of the users. In that sense, it is similar to the popularity strategy, but it tries to make a better prediction of what items the user can rate by leveraging the knowledge of the items that users with similar personalities have rated in the past.

Personality-Based Binary Prediction first transforms the rating matrix to a matrix with the same number of rows and columns, by mapping null entries to 0, and not null entries to 1. Hence, the new matrix models only whether a user rated an item or not, regardless of its value. Then, the new matrix is used to train an extended version of the popular matrix factorization algorithm. Our model is similar to the one proposed in [15], and profiles users not only in terms of the binary ratings, but also using known user attributes, in our case, gender, age group and the scores for the Big Five personality traits on a scale from 1 to 5. Given a user u, an item i and the set of user attributes $A(u)$, it predicts ratings using the following rule:

$$\hat{r}_{ui} = \bar{i} + b_u + q_i^\top \cdot (p_u + \sum_{a \in A(u)} y_a), \tag{1}$$

where p_u, q_i and y_a are the latent factor vectors associated with the user u, the item i and the user attribute a, respectively. The model parameters are then learnt, as it is common in matrix factorization, by minimizing the associated regularized squared error function through stochastic gradient descent. Finally, the learnt model predicts and assigns a rateable score to each candidate item i (for each user u), with higher scores indicating a higher probability that the target user has consumed the item i, and hence may be able to rate it.

Both strategies have been used in a "compound" AL strategy in order to select the most useful items for the active user to rate. This has been done by getting a short list of items (4) from each of these strategies and merging them together in a way that the final list includes equal numbers of items from the two strategies. Exploiting a combination of two strategies with different characteristics is beneficial. Firstly, such a combination allowed us to compare their performances, in an offline analysis where we built separated training sets with

the ratings acquired by each individual strategy. Secondly, in a finally deployed RS one can take advantage of both AL strategies to add some diversity to the system rating requests.

Finally, in order to have a baseline we have also considered a control group that was requested to rate items that were selected at **Random**.

5   User Study Evaluation

In order to evaluate the considered elicitation strategies, we conducted a user study involving 108 participants who were randomly assigned either to the compound AL strategy group (n = 54) or the random item selection strategy group (n = 54). Our goal was to study the influence of the rating elicitation strategies on the evolution of the RS's performance. Given a particular AL strategy, the (training) rating matrix evolves by including all the ratings entered by users on the training items elicited so far. The exact test procedure was as follows. After the user has completed the personality questionnaire (as illustrated in section 3.1) she is asked to rate 13 items (see figure 1, right): 5 of these items are test items, i.e., items selected randomly to test the obtained model accuracy, while 8 are train items, i.e., items selected to train the model. For the random strategy group the 8 train items are selected randomly. For the compound AL strategy group, 4 of the train items are selected by log(pop) * entropy strategy, and 4 by personality-based binary prediction strategy.

This evaluation set-up, i.e., one control group and one AL compound strategies group, allowed us to assign a larger number of users to the considered AL strategies and hence to test and compare their performances more reliably. Then, in order to be fair in the comparison of these strategies vs. the random strategy, i.e., to simulate the same number of rating requests for each individual strategy, when evaluating the accuracy of the RS we randomly sampled with probability 0.5 the train ratings acquired by the random strategy (since 8 items was requested to rate at random, vs. 4 using the two AL strategies). Finally, after having trained the prediction model on all the ratings acquired from the users with a specific AL strategy during the time period of the study, the Mean Absolute Error (MAE) on the ratings of the selected random test items was measured. We have selected the test items randomly in order to avoid any bias that may affect the MAE especially when the rating data set is small.

6   Evaluation Result

6.1   Mean Absolute Error

Figure 2 shows the system's MAE after training our extended matrix factorization model with the users' ratings collected by an individual AL strategy, while at the same time ignoring those collected by the other strategies. After having retrained the model with all the users' ratings elicited by a certain AL strategy,

for each user in the user study, we have computed the system MAE on the ratings given to the test items that a user was able to provide and then averaged the MAE of all the users. This allowed us to compute to what extent the prediction accuracy is affected by the addition of new ratings (MAE). Therefore, it indicates the effect of a rating elicitation strategy on the prediction accuracy of the system.

Comparing the results it is clear that our proposed strategy, i.e., personality-based binary prediction, outperforms the other strategies in terms of MAE. As it can be observed in Figure 2, the initial MAE of the system using the 848 training ratings that were available at the beginning of the study was 1.06. The MAE decreased to 0.86 by adding the 125 training ratings acquired by our proposed strategy, whereas it was 0.90 by adding the 112 training ratings collected by the log(popularity) * entropy strategy, and 0.97 by adding the 73 training ratings derived from the random strategy. Hence, the MAE reduction achieved by AL, is 18.8% for personality-based binary prediction, 15.0% for log(popularity) * entropy, and 8.4% for random. Consequently, under application of our proposed strategy, the system's MAE has been reduced the most.
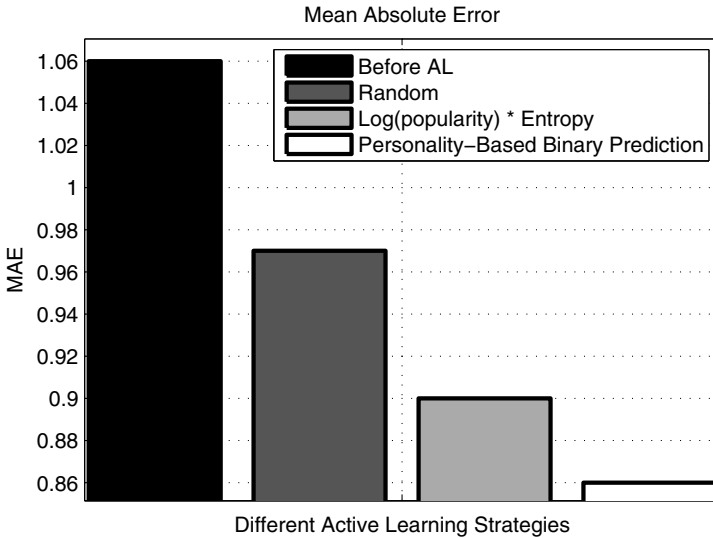

Fig. 2. MAEs of the strategies

6.2   Number of Acquired Ratings

Another important aspect to consider is the number of ratings that are acquired by the considered strategies. As we discussed before, certain strategies can acquire more ratings by better estimating what items are likely to have been experienced or known by the user. Table 1 summarizes our results. Overall,
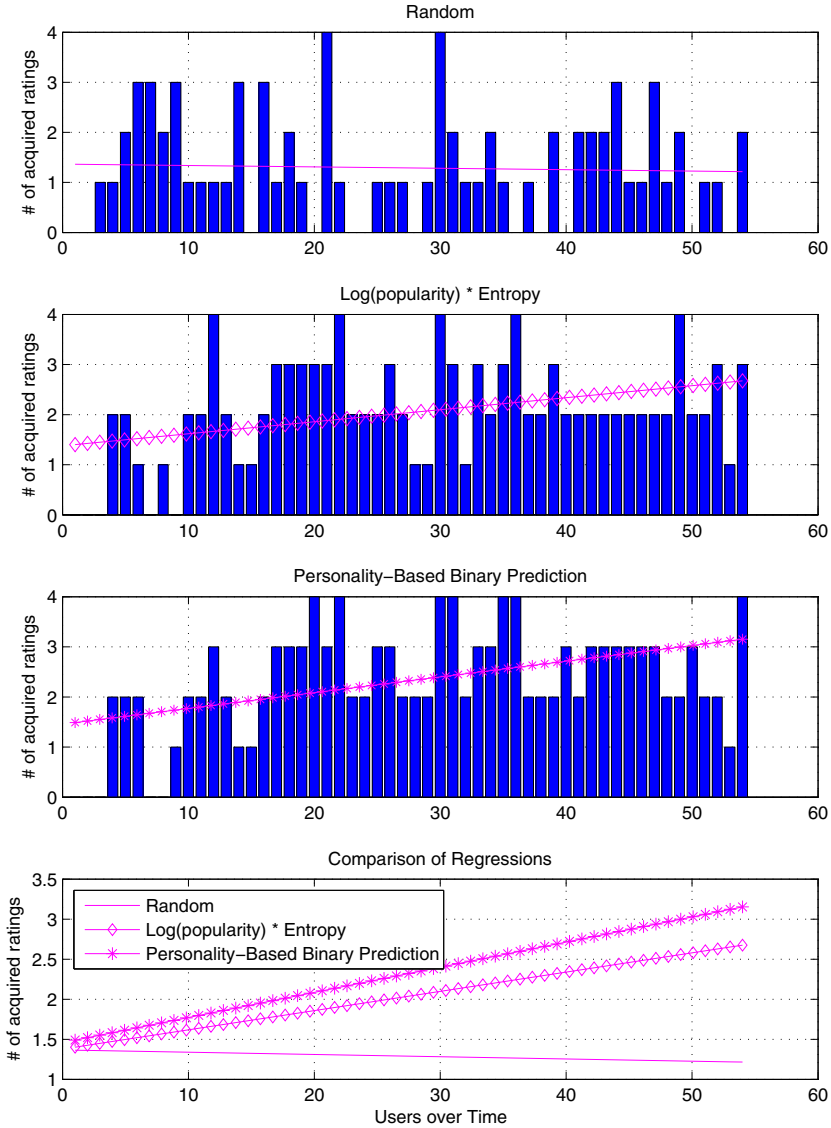
Fig. 3. Number of acquired ratings by different strategies over time

the average number of ratings acquired per user in our train set is 1.9 (out of 4 items that were requested to rate). Ultimately, the personality-based binary prediction strategy has elicited by average 2.31 ratings from each user, whereas the log(popularity) * entropy strategy elicited 2.07 ratings, and random strategy elicited 1.35 ratings. Conducting t-test, we observed that the personality-based binary prediction strategy can elicit significantly more ratings than either

log(popularity) * entropy strategy (p-value = 0.0054) or random strategy (p-value<0.0000). Our strategy can obtain more than 57% of the requested ratings. Conversely, log(popularity) * entropy can acquire 51% and random 33% of the requested ratings. Interestingly, these are the results that we conjectured in one of our previous works [5].

In figure 3 we plot the number of ratings that each strategy has acquired from the users over time together with the corresponding regression lines. Observing the results, we can see that over time the average number of ratings acquired by the random strategy is approximately constant. However, this number is increasing when the personality-based binary prediction and log(popularity) * entropy strategies are used. Overall, our proposed strategy, i.e., the personality-based binary prediction, achieves the best result. The number of ratings that this strategy acquires is increasing faster than using the other strategies (its regression line has the largest slope). This means that our proposed strategy is better learning to estimate which items may have been experienced by the users and the users are able to rate. This is an important factor since if a strategy selects the items that are informative but not experienced by the users, it can not acquire their ratings (which is typical for the random strategy). Therefore, a good strategy must focus not only on the quality but also on the quantity of the ratings, as our proposed strategy does.

Finally, when looking at the relationships between user's personality traits and the number of ratings provided, we have observed that people with higher openness to experience can rate significantly more items than those with lower openness to experience (p-value=0.044). This means that the personality of the people can affect their rating behaviors, which is the conjecture that initially motivated this research. This type of relationships opens a new line of research, showing that user's personality can be useful to select and assign a convenient AL strategy to a group of users, when eliciting the largest number of ratings is in order.

Table 1. Pairwise comparison of the number of ratings acquired over 4 requests by different strategies

Pair of strategies	Means	p-value	# of ratings
Random / log(popularity) * entropy	1.35 / 2.07	0.0003	73 / 112
Random / personality-based binary prediction	1.35 / 2.31	0.0000	73 / 125
Personality-based binary prediction / log(popularity) * entropy	2.31 / 2.07	0.0054	112 / 125

7   Conclusions and Future Work

In this paper, we have presented a novel AL technique for addressing the cold-start problem in RSs. The proposed technique uses the Five Factor Model (FFM) of personality traits as its basis in order to provide a user with personalized rating requests, without completely relying on explicit feedback (e.g., ratings)

or implicit feedback (e.g., item views or purchases) which is usually more difficult to obtain and are not available in cold-start situations.

We have developed the following two research hypotheses: a) our proposed AL method leads to a higher increase in the number of acquired user ratings in comparison to a state-of-the-art rating elicitation strategy, and b) the prediction accuracy of the recommendation model improves more when utilizing our proposed AL strategy than when using another popular and effective state of the art AL strategy. In a live user study, we successfully verified both hypotheses and we have shown that user personality has an important impact in her rating behaviour.

Our future work includes the further analysis of the data obtained from the study in order to understand potential performance differences among the compared AL strategies that are due to different personality traits. We would like to test our proposed method on larger training and test sets. We would also like to understand the impact of using different rating prediction models, such as context-aware ones, on the performance of our proposed AL approach.

References

1. Short personality quiz - psych central. Based upon the Ten-Item Personality Inventory (TIPI) (February 2013)
2. Baltrunas, L., Ludwig, B., Peer, S., Ricci, F.: Context relevance assessment and exploitation in mobile recommender systems. Personal and Ubiquitous Computing 16(5), 507–526 (2012)
3. Costa, P., McCrae, R.: Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In: The Five-Factor Model of Personality: Theoretical Perspectives, pp. 51–87 (1996)
4. Dunn, G., Wiersema, J., Ham, J., Aroyo, L.: Evaluating interface variants on personality acquisition for recommender systems. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 259–270. Springer, Heidelberg (2009)
5. Elahi, M., Repsys, V., Ricci, F.: Rating elicitation strategies for collaborative filtering. In: Huemer, C., Setzer, T. (eds.) EC-Web 2011. LNBIP, vol. 85, pp. 160–171. Springer, Heidelberg (2011)
6. Elahi, M., Ricci, F., Rubens, N.: Adapting to natural rating acquisition with combined active learning strategies. In: Chen, L., Felfernig, A., Liu, J., Raś, Z.W. (eds.) ISMIS 2012. LNCS, vol. 7661, pp. 254–263. Springer, Heidelberg (2012)
7. Elahi, M., Ricci, F., Rubens, N.: Active learning strategies for rating elicitation in collaborative filtering: a system-wide perspective. ACM Transactions on Intelligent Systems and Technology 5(1) (2014)
8. Golbandi, N., Koren, Y., Lempel, R.: On bootstrapping recommender systems. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1805–1808. ACM (2010)
9. Golbandi, N., Koren, Y., Lempel, R.: Adaptive bootstrapping of recommender systems using decision trees. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, pp. 595–604. ACM (2011)
10. Goldberg, L.R.: The development of markers for the big-five factor structure. Psychological Assessment 4(1), 26–42 (1992)

11. Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the big-five personality domains. Journal of Research in Personality 37, 504–528 (2003)
12. Hu, R., Pu, P.: A comparative user study on rating vs. personality quiz based preference elicitation methods. In: Proceedings of the 14th International Conference on Intelligent User Interfaces, IUI 2009, pp. 367–372. ACM, New York (2009)
13. Hu, R., Pu, P.: Enhancing collaborative filtering systems with personality information. In: Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys 2011, pp. 197–204. ACM, New York (2011)
14. John, O.P., Srivastava, S.: The big five trait taxonomy: History, measurement, and theoretical perspectives. In: Handbook of Personality: Theory and Research, vol. 2, pp. 102–138 (1999)
15. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer 42(8), 30–37 (2009)
16. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 2–5 (March 2013)
17. Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., Mcnee, S.M., Konstan, J.A., Riedl, J.: Getting to know you: Learning new user preferences in recommender systems. In: Proceedings of the 2002 International Conference on Intelligent User Interfaces, IUI 2002, pp. 127–134. ACM Press (2002)
18. Rashid, A.M., Karypis, G., Riedl, J.: Learning preferences of new users in recommender systems: an information theoretic approach. ACM SIGKDD Explorations Newsletter 10(2), 90–100 (2008)
19. Rentfrow, P.J., Gosling, S.D., et al.: The do re mi's of everyday life: The structure and personality correlates of music preferences. Journal of Personality and Social Psychology 84(6), 1236–1256 (2003)
20. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B.: Recommender Systems Handbook. Springer (2011)
21. Rubens, N., Kaplan, D., Sugiyama, M.: Active learning in recommender systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P. (eds.) Recommender Systems Handbook, pp. 735–767. Springer (2011)
22. Tkalcic, M., Kosir, A., Tasic, J.: The ldos-peraff-1 corpus of facial-expression video clips with affective, personality and user-interaction metadata. Journal on Multimodal User Interfaces 7(1-2), 143–155 (2013)