Implicit Acquisition of User Personality for Augmenting Movie Recommendations

Wen Wu(✉) and Li Chen

Department of Computer Science, Hong Kong Baptist University,
Kowloon Tong, Hong Kong, China
{cswenwu,lichen}@comp.hkbu.edu.hk

Abstract. In recent years, user personality has been recognized as valuable info to build more personalized recommender systems. However, the effort of explicitly acquiring users' personality traits via psychological questionnaire is unavoidably high, which may impede the application of personality-based recommenders in real life. In this paper, we focus on deriving users' personality from their implicit behavior in movie domain and hence enabling the generation of recommendations without involving users' efforts. Concretely, we identify a set of behavioral features through experimental validation, and develop inference model based on Gaussian Process to unify these features for determining users' big-five personality traits. We then test the model in a collaborative filtering based recommending framework on two real-life movie datasets, which demonstrates that our implicit personality based recommending algorithm significantly outperforms related methods in terms of both rating prediction and ranking accuracy. The experimental results point out an effective solution to boost the applicability of personality-based recommender systems in online environment.

Keywords: Recommender systems · User personality · Implicit acquisition · Collaborative filtering

1 Introduction

Nowadays, recommender systems (RS) have been widely adopted in many online applications for eliminating users' information overload and providing personalized services to them. In order to more accurately learn users' preferences, personality has been increasingly recognized as a valuable resource being incorporated into the process of generating recommendations [18,25]. As a typical work, Tkalcic *et al.* [25] adopted users' personality to enhance the measure of nearest neighborhood in collaborative filtering (CF) systems. Hu and Pu demonstrated that personality can be leveraged into solving the cold-start problem of CF [18]. However, existing studies have mostly relied on personality quiz to explicitly acquire users' personality, which unavoidably demands user efforts. From users' perspective, they may be unwilling to answer the quiz for the sake of saving efforts or protecting their privacy. The application of existing personality-based RSs will thus be limited in real life.

F. Ricci et al. (Eds.): UMAP 2015, LNCS 9146, pp. 302–314, 2015.
DOI: 10.1007/978-3-319-20267-9_25

In this paper, we are motivated to study how to implicitly derive users' personality from their behavior data. Specifically, the main contributions of our work are as follows:

(a) We have first identified a set of features that are significantly correlated with users' personality traits, through experimental validation. Among them, some are domain dependent, such as users' preference for movie genre and movies' diversity, watching duration, and watching motives. Some are domain independent, like users' rating behavior and age info.
(b) We have then integrated all of these features into a regression model to infer users' personality. We have concretely tested three different models, Gaussian Process, Pace Regression, and M5 Rules, and found that Gaussian Process performs the best in terms of inference accuracy.
(c) At the last step, we have developed three variations of personality based CF algorithm, for which the personality is implicitly inferred from user behavior in real-life movie datasets. Through experiment, we have demonstrated that the method combining user personality and ratings is most accurate in terms of both rating prediction and ranking accuracy. The results thus highlight the practical merit of our method in augmenting online movie recommendations.

In the following, we first introduce related work on personality-based recommender systems (Section 2). We then present the details of our feature selection process in movie domain, as well as the experiment on personality inference model in Section 3. In Section 4, we describe how the model is incorporated into the CF based recommending framework. We finally conclude the work and indicate its future directions in Section 5.

2 Related Work

How to use user personality for benefiting recommender systems has attracted increasing attentions in recent years. For instance, Tkalcic *et al.* employed personality to enhance the nearest neighborhood measure in a CF-based image recommender system, and demonstrated that the personality based similarity measure is more accurate than the rating based measure [25]. Hu and Pu [18] also incorporated users' personality into the CF framework. Their experiment indicated that personality-based CF methods are significantly superior to the non-personality based approach in terms of solving the cold-start problem. In addition, they proposed a preference-based approach [17], for which a personality-based interest profile is established for each user to reflect the relationship between her/his personality and preference for music genre [23]. The items that best match the user's profile are recommended. A user study revealed that users in this system perceive the recommended items to be accurate for their friends. Elahi *et al.* [7] developed a novel active learning strategy based on personality, for predicting items that users are able to rate before they get recommendations. They concretely implemented an attribute-based matrix factorization (MF) algorithm, where the attributes include users' age, gender, and personality traits. Through a live user study, they proved that involving these attributes, especially users' personality, can significantly increase the number of items that users can rate.

However, the issue of how to obtain users' personality is still not well solved. The existing studies mainly rely on psychological questionnaires (such as IPIP personality quiz [11] used in [25], and TIPI personality quiz [12] in [7,17,18]) to measure users' personality traits, which unavoidably demand a lot of user efforts and hence impede the system's practical application in real life. Lately, there have been endeavors to derive users' personality from their self-generated data. For instance, Gao *et al.* attempted to derive users' personality from their micro blogs [8]. Golbeck *et al.* identified users' personality traits through their Facebook profile, which includes language features, activities, and personal information [9]. But their main focus has been on social networking sites, rather than on social media domains such as movie. Another limitation is that little work has been done on incorporating the implicitly acquired personality into the process of recommendation generation.

We are thus interested in not only exploring proper behavioral features in movie domain that can be used to infer user personality, but also investigating how to improve recommendation accuracy with the inferred personality.

3 Implicit Acquisition of User Personality

A widely used personality model is the so called big-five factor model, which defines user personality as five traits [14][1]: *Openness to Experience (O)*, *Conscientiousness (C)*, *Extroversion (E)*, *Agreeableness (A)*, and *Neuroticism (N)*. In this section, we first describe how we have identified a set of features and experimentally validated their significant correlations with users' personality traits. We then introduce the personality inference model and its accuracy measurement.

3.1 Feature Identification

User Preference for Movie Genre. The psychological researches conducted in movie domain have shown that users' personality can influence their preference for movie genre. For example, both Cantador *et al.* [1] and Chausson [3] pointed out that more imaginative and creative people (with high O[2]) are more inclined to choose comedy movies, whereas people with low O (i.e., cautious and consistent people) and high E (i.e., outgoing and energetic people) tend to choose romance movies. Therefore, in our work, "movie genre" is taken as an important personality feature.

[1] Due to space limit, the description of these five traits can be found in [14].

[2] High O means that the user has high score on the personality trait "Openness to Experience", which is also applied to the other abbreviations.

Users' Watching Duration. Users' watching duration is another feature that we have considered, because it was reported that people with different personality traits may behave differently in terms of time spent on watching TV programs [21]. For instance, those who are more sensitive and nervous (with high N) tend to spend more time in watching. Considering that movie is similar to TV program in that it is also a type of leisure media, we think that users' personality may also influence their movie watching duration.

Users' Motives Behind Watching Movies. Chamorro-Premuzic *et al.* [2] indicated that users' personality traits can affect their motives when choosing movies. Specifically, the motives include emotional factors (like hedonism and sensation seeking) and cognitive factors (like boredom avoidance and information seeking). It was found that those who possess high O tend to select movies for seeking useful information (i.e., cognitive motives), while individuals with high A and E are likely to watch movies for fun (i.e., emotional motives). Thus, in our work, we take "watching motive" as one personality feature.

User Preference for Movies' Diversity. In our previous work [4,27], we found that users' personality can affect their spontaneous needs for the level of diversity within a list of movies that they will select. For instance, people with high N are more likely to prefer movies with diverse directors. Moreover, the personality trait C is significantly correlated with users' preference for movies' overall diversity when all attributes are considered. Formally, a user's preference for diversity in respect of a specific movie attribute (e.g., genre, director) can be determined via the intra-list metric:

$$Div(attr_k) = \frac{2}{|S_u| \times |S_u - 1|} \sum\nolimits_{m_i \in S_u} \sum\nolimits_{m_j \neq m_i \in S_u} (1 - Sim(m_i, m_j)) \tag{1}$$

where $attr_k$ is the *k-th* attribute, S_u is the set of movies selected by the user, and $Sim(m_i, m_j)$ gives the similarity between two movies in terms of $attr_k$ (i.e., $Sim(m_i, m_j) = \frac{|S_{m_i,attr_k} \cap S_{m_j,attr_k}|}{|S_{m_i,attr_k} \cup S_{m_j,attr_k}|}$, where $S_{m_i,attr_k}$ contains all values of $attr_k$, such as all actors, in movie m_i). The user's preference for movies' overall diversity is further measured as:

$$OverDiv = \sum\nolimits_{k=1}^{n} (w_k \times Div(attr_k)) \tag{2}$$

where w_k indicates the attribute's relative importance to the user ($0 <= w_k <= 1, \sum_{k=1}^{n} w_k = 1$) and n is the total number of attributes.

Users' Rating Behavior. Users' rating behavior might also be indicative of their personality. As shown by Golbeck and Norris [10], two personality traits E and C are both positively correlated with users' ratings on movies. That is, people who are more extrovert, outgoing, cautious, and self-disciplined tend to give higher ratings than others. Besides, Hu and Pu [19] found that C is negatively correlated with the number of ratings that a user gives, which implies that unorganized and impulsive people are likely to rate more items.

Users' Demographic Properties. There are several demographic properties that have been shown related to users' personality. For instance, Chausson [3] found that females usually score higher on personality traits N and A than males. As for age, McCrae *et al.* found that older people have lower scores on E and O, but higher scores on A and C [20]. Regarding education, people with high C (i.e., self-disciplined ones) are more likely to obtain higher education degree [5].

In summary, we have identified six major types of features for implicitly deriving users' personality. Some of them, such as users' rating behavior and demographic properties, can be applied to other domains except for movie, which we call *domain-independent* features, and the others are called *domain-dependent* features.

3.2 Validation Experiment

In order to validate whether the above-mentioned features are truly significantly correlated with users' personality, we have conducted a validation experiment.

User Survey Setup. To do the validation, we first performed a user survey to collect users' interaction behavior with movies and personality values. A total of 148 volunteers (77 females) joined this survey. All of them are Chinese, who are with different education backgrounds (28% with Bachelor, 57% with Master, 7% with PhD, and 8% miscellaneous) and age ranges (61% in the range of 20-25 years old, 32% in the range of 25-30, 4% in the range of 30-40, and 3% in the other ranges).

Each user's personality was assessed via a popular big-five personality quiz that contains 25 questions [14]. Each personality trait's score is the sum of scores on its related 5 questions. For example, one question of assessing *Neuroticism* is *"Do you feel you are always calm or eager?"* which is rated from 1 "calm" to 5 "eager". We also asked each user to freely rate at least 10 movies that s/he has watched by checking movies' information in Douban Movie (movie.douban.com), which is a popular movie reviewing website in China.

Implicit Features. The behavioral features (identified in the previous section) were concretely determined in the following ways. A user's preference for movie genre was obtained by asking her/him to choose three most favorite genres and three least favorite ones (among 15 genres such as action, horror, etc.) in our survey. As for watching duration, we first used the user's answer to the question *"How often do you see a movie?"* (among the options "one week", "one month", "three months", etc.) to represent her/his watching frequency, and then calculated the watching duration by multiplying her/his weekly watching frequency with the movie's average length (i.e., 1.5 hours). The user's main motive behind watching movies was assessed via the question *"At most of times, why do you want to see a movie, for seeking useful information or just for fun?"* In parallel, Eq.1 and Eq.2 were applied to measure the user's preferences for movies' attribute-specific diversity and overall diversity, for which only movies that the user gave positive ratings (i.e., above 3 out of 5) were considered, and attributes' weights (in Eq.2) were obtained through conjoint analysis [13]. Regarding rating behavior, both the average rating and the total number of ratings that a user gave to movies were taken into account.

Correlation Results. The correlations between users' personality traits and their behavioral features were computed via Spearman's rank coefficient because it can be applied to both ordinal and numerical variables [29]. Table 1 shows the results. With respect to user preference for movie genre, it is significantly correlated with all of the five personality traits. For example, people with high O tend to prefer music and animation movies, whereas those who score low on O prefer documentary movies. The genre's preference is also correlated with users' personality traits C, E, A, and N. In terms of watching duration, the personality traits C and A show significantly positive correlations with it, which implies that those who are more organized, patient, friendly and cooperative are more likely to spend more time in watching movies. In addition, those people also tend to watch movies for fun rather than seeking information given the significant correlations between the two traits C and A and users' watching motive. As for rating behavior, people with high C and A are found being more subject to give higher ratings. Furthermore, O and E are significantly negatively correlated with users' preference for movies' diversity w.r.t. genre, indicating that conventional and introvert persons are more inclined to choose diverse genres. The movies' diversity w.r.t country is preferred by suspicious and antagonistic users (with low A). Finally, among users' demographic properties, we observe that age is significantly correlated with O in a negative way, which suggests that young people are more imaginative and creative than elders.

Thus, it can be seen that most of features are empirically proven with significant correlations with users' personality. Particularly, more *domain-dependent* features, such as users' preference for movie genre, their watching duration, and watching motive, exhibit strong correlations, which suggests that the behavioral features related to a specific domain can be more helpful for inferring users' personality. For the next step, we have attempted to combine all of these significant features into a unified inference model.

3.3 Personality Inference Model and Evaluation

Inference Model. We have compared three regression models for inferring users' personality: Gaussian Process, Pace Regression, and M5 Rules. Formally, a standard form of regression model can be represented as $y = f(x) + \varepsilon$, where x denotes an input vector (in our case, it includes the implicit features like user preference for movie genre, the user's watching duration, etc.), y denotes a scalar output (in our case, it gives the inferred big-five personality scores), and ε is the additive noise. Our purpose is then to estimate the regression function $f(\cdot)$.

Gaussian Process (GP) [22] defines a probabilistic regression based on Bayesian theory and statistical learning theory: $f(x) \sim gp(\mu(x), k(x, x'))$, where

Table 1. Correlations between users' personality traits and implicit features ($^*p < 0.05$ and $^{**}p < 0.01$)

		Openness to Experience (O)	*Conscientiousness (C)*	*Extroversion (E)*	*Agreeableness (A)*	*Neuroticism (N)*
User preference for movie genre		***O***: Music (0.207*), Animation (0.183*), Documentary (-0.167**); ***C***: Animation (-0.201*), Comedy (-0.163*), War (0.192**), Science Fiction (0.176*); ***E***: Mystery (-0.163*), Crime (-0.18*), Romance (0.136**); ***A***: Animation (-0.20*), Science Fiction (0.171*), War (0.192*); ***N***: Romance (0.203*), History (-0.173*), Drama (0.184*), Adventure (-0.106*), Animation (0.202*), Documentary (-0.165*).				
Watching duration		-0.043	**0.176***	0.147	**0.175***	-0.094
Watching motive		-0.091	**0.208***	0.001	**0.181***	-0.139
User preference for movies' diversity	Overall div.	0.038	0.033	-0.010	-0.029	-0.051
	Genre div.	**-0.135***	0.057	**-0.176***	-0.072	0.050
	Country div.	0.069	-0.130	0.069	**-0.189***	0.166
	Release time div.	-0.144	-0.040	-0.104	-0.140	0.085
	Actor/actress div.	-0.003	0.025	0.061	0.001	-0.012
	Director div.	-0.037	-0.013	0.050	0.085	-0.053
Rating behavior	Average rating	-0.134	**0.175***	0.082	**0.262****	-0.018
	Number of ratings	-0.145	0.219	0.040	0.217	-0.012
Demographic property	Gender	0.068	-0.070	0.110	-0.049	0.075
	Age	**-0.201***	0.103	0.121	0.082	-0.057
	Education level	-0.033	0.157	-0.002	0.109	-0.137

$\mu(x)$ stands for the mean function and $k(x, x')$ is the covariance function. In practice, GP can handle datasets with small number of samples and/or many input features. Unlike GP, Pace Regression [26] is a typical form of linear regression analysis. It is applicable when some of the input features are mutually dependent. M5 Rules [16] also assumes a linear distribution of the input features, but it is grounded on the separate-and-conquer strategy to build a decision tree. In comparison to the other models, M5 Rules costs less calculation and can deal with small-scale datasets or datasets with missing values.

Evaluation Procedure. We randomly selected 90% of 148 users who participated in our user survey (see Section 3.2) to train each model and tested it on the remaining 10% users. To avoid any biases, we performed 10-fold cross validation, and measured the accuracy via metrics *Root Mean-Square Error (RMSE)* [15] and *Pearson correlation* [15]. Formally, we define a user's personality as a 5-dimension vector $p_u = (p_u^1, p_u^2, ..., p_u^5)^T$, where each dimension p^k represents one personality trait among O, C, E, A, and N.

Evaluation Results. The results are shown in Table 2, in which *RMSE* values that are returned by the three regression models in respect of each personality trait are all within 10% of its real score. Moreover, Gaussian Process obtains a relatively smaller margin of error than Pace Regression and M5 Rules. In terms of *Pearson correlation*, Gaussian Process produces results with higher correlations than Pace Regression and M5 Rules, regarding all of the personality traits. Particularly, the correlations by Gaussian Process are significant regarding personality traits C ($p < 0.01$), A ($p < 0.05$), and N ($p < 0.05$). We further run

Table 2. RMSE and Pearson correlation results of testing three inference models (*note*: the number in superscript indicates that the model significantly ($p < 0.05$) outperforms the referred one in terms of the corresponding metric)

Personality trait	**[1]Gaussian Process**	**[2]Pace Regression**	**[3]M5 Rules**
RMSE			
Openness to Experience (O)	0.0354^{3}	0.0357^{3}	0.0369
Conscientiousness (C)	0.0350^{3}	0.0382^{3}	0.0380
Extraversion (E)	0.0375^{3}	0.0378^{3}	0.0410
Agreeableness (A)	0.0484^{3}	0.0494^{3}	0.0504
Neuroticism (N)	0.0568^{3}	0.0584^{3}	0.0605
Pearson correlation coefficient ($^{*}p < 0.05, ^{**}p < 0.01$)			
Openness to Experience (O)	$0.1424^{2,3}$	0.0458	0.0537
Conscientiousness (C)	$\mathbf{0.3319}^{**,2,3}$	0.0600	$\mathbf{0.1942}^{*}$
Extraversion (E)	$0.1085^{2,3}$	0.0839	0.0871
Agreeableness (A)	$\mathbf{0.2177}^{*,2,3}$	0.1204	$\mathbf{0.1704}^{*}$
Neuroticism (N)	$\mathbf{0.2634}^{*,2,3}$	0.0553	0.1375

Note: Here, we normalized each personality value into [0-1] scale via the logarithmic form of normalization: $\overline{X} = \frac{\log_{10} X}{\log_{10} max}$, where X is the original score of each personality trait, and max gives the maximum value among all of the samples. Moreover, the standard deviations of the five personality traits at the normalized 0-1 scale are respectively: $\boldsymbol{O}(0.035)$, $\boldsymbol{C}(0.038)$, $\boldsymbol{E}(0.037)$, $\boldsymbol{A}(0.048)$, $\boldsymbol{N}(0.058)$.

a pairwise t-Test to identify the significance level of differences between these models. It shows that Gaussian Process significantly outperforms M5 Rules in terms of both *RMSE* and *Pearson correlation* ($p < 0.05$) in respect of all of the personality traits, and is significantly better than Pace Regression in terms of *Pearson correlation* ($p < 0.05$). Pace Regression is significantly better than M5 Rules in terms of *RMSE* ($p < 0.05$). As for the reason why Gaussian Process performs the best, we think it is because the non-linear relationship between input and output as defined in this model may better fit the characteristic of our data. It also avoids bringing the noise of input to output by performing a probabilistic function.

The above results hence demonstrate that a user's personality traits can be inferred by unifying some implicit features. Given that Gaussian Process shows the best accuracy among the three compared models, we adopt it for the next step of recommendation generation.

4 Recommendation Based on Implicit Personality

The next question we are interested in solving is: *how to employ the implicitly acquired personality for improving the real-life movie recommendation?* In the following, we present our algorithm development and evaluation results.

4.1 Algorithm Development

We first apply Gaussian Process to derive users' personality from their interaction behavior with movies (i.e., implicit features). The inferred personality is then incorporated into the collaborative filtering (CF) process. Concretely, we develop three variations of the personality-based recommending method: pure personality-based CF, and two hybrid CF methods that integrate personality with ratings.

Personality-Based CF. Similar to the approach proposed in [18], we can use personality to calculate user-user similarity. Specifically, we adopt the 5-dimension vector $p_u = (p_u^1, p_u^2, ..., p_u^5)^T$ (derived from the previous step; see Section 3.3) to define a user's personality, and then compute the personality-based similarity between two users u and v via *Cosine* measure:

$$Simp(u,v) = \frac{\sum_{k=1}^{5}(p_u^k \times p_v^k)}{\sqrt{\sum_{k=1}^{5} {p_u^k}^2}\sqrt{\sum_{k=1}^{5} {p_v^k}^2}} \tag{3}$$

The rating predicted to an unknown item i for the target user u is then calculated as:

$$PreScore_{personality}(item_i) = \overline{r_u} + k\sum_{v\in\Omega_u} Simp(u,v) \times (r_{v,i} - \overline{r_v}) \tag{4}$$

where $r_{v,i}$ is user v's rating for item i, $\overline{r_u}$ and $\overline{r_v}$ are respectively user u's and v's average ratings, k is equal to $\frac{1}{\sum_{v\in\Omega_u}|Simp(u,v)|}$, and Ω_u is the set of u's neighbors who rated item i (note that the similarity threshold is set as 0.7 in our experiment). We call this method PB.

Hybrid CF. We further implement two hybrid CF methods. One is called $RPBL_{score}$, where users' implicit personality scores and ratings are combined into the CF framework in a linear weighted way:

$$PreScore_{hybrid}(item_i) = \alpha \times PreScore_{rating}(item_i) + (1-\alpha) \times PreScore_{personality}(item_i) \tag{5}$$

where $PreScore_{rating}(item_i) = \overline{r_u} + k\sum_{v\in\Omega_u} Simr(u,v) \times (r_{v,i} - \overline{r_v})$, which returns the predicted rating based on the traditional rating-based similarity measure (i.e., $Simr(u,v) = \frac{\sum_{i\in I}(r_{u,i}\times r_{v,i})}{\sqrt{\sum_{i\in I} {r_{u,i}}^2}\sqrt{\sum_{i\in I} {r_{v,i}}^2}}$), $PreScore_{personality}(item_i)$ is the personality-based rating prediction (Eq.4), and α is a weighting parameter used to control the relative contributions of the two types of prediction (that is tuned through experimental trials).

An alternative hybrid method is called $RPBL_{similarity}$ [18], which emphasizes combining personality scores and ratings for computing user-user similarity:

$$Simpr(u,v) = \beta \times Simr(u,v) + (1-\beta) \times Simp(u,v) \tag{6}$$

where β manipulates the weights of the two kinds of similarity measure, i.e., $Simr(u,v)$ and $Simp(u,v)$. $Simpr(u,v)$ is then used to predict an unknown item i's rating for the target user u (by replacing $Simp(u,v)$ in Eq.4).

4.2 Experiment Setup

In the experiment, we compared the three personality-based approaches, i.e., pure personality-based CF (PB) and two hybrid CF methods ($RPBL_{score}$ and $RPBL_{similarity}$), with the standard rating-based CF [18] (shortened as RB) on two real-life movie datasets: HetRec[3] and Yahoo! Movie[4]. Both datasets contain users' movie ratings, and movies' descriptive information such as their genres, actors, directors, and so on. Yahoo! Movie dataset's rating sparsity level[5] is relatively higher (98.8% vs. 95.5% in HetRec dataset). We performed 10-fold cross-validation to measure each method's performance. Concretely, we split the ratings provided by each user into two parts, 80% used for inferring the user's personality, and 20% for testing the recommendation result.

The implicit features used for constructing the personality inference model (see Section 3.2) are identified through the following ways. To obtain user preference for movie genre, we used the proportion of each genre that appears in movies that a user gave positive ratings (i.e., above 3 out of 5), to represent her/his preference for that genre. As for watching duration, we multiplied the average number of movies that the user has rated per week with the movie's average length. The user's preferences for movies' diversity in respect of genre and country were calculated via Eq.1. We also calculated each user's average rating. As for their age information, it is provided in HetRec dataset, but not in Yahoo! Movie. The feature "watching motive" was not considered in this experiment because it is not available in both datasets. We then adopted Gaussian Process (see Section 3.3) to unify these features for inferring each user's big-five personality traits.

The recommendation accuracy was measured in terms of both rating prediction and ranking performance: 1) *Mean Absolute Error (MAE)* and *Root Mean-Squared Error (RMSE)* [15] for determining the tested method's rating prediction accuracy; and 2) *Normalized Discounted Cumulative Gain (nDCG)* and *Mean Average Precision (MAP)* [28] used for determining the method's top-N ranking accuracy ($N = 10$ in our experiment). Due to space limit, the details of these four metrics are referred to [15, 28].

4.3 Results

The optimal values of parameters involved in these algorithms, including the neighborhood size k for each method, α in Eq.5, and β in Eq.6, were identified through experimental trials. Table 3 shows each method's performance (with its optimal parameter value(s)) and the pairwise t-Test statistical analysis results.

It can be seen that the two hybrid CF methods, $RPBL_{score}$ and $RPBL_{similarity}$, both significantly outperform the pure personality-based approach (PB) and the standard rating-based method (RB), in terms of rating prediction (i.e., *MAE* and *RMSE*) in the two datasets. They also achieve significantly higher ranking accuracy (*nDCG-5* and *MAP*) than RB. The

---

[3] HetRec dataset contains 1,872 users, 10,095 movies, and 848,169 ratings.

[4] Yahoo! Movie dataset contains 800 users, 9,128 movies, and 90,832 ratings.

[5] The rating sparsity level is calculated via $Level_{sparsity} = 1 - \frac{|nonzero_{entries}|}{|total_{entries}|}$ [24].

Table 3. Overall comparison results (*note*: the number in superscript indicates that the method significantly ($p < 0.05$) outperforms the referred one in terms of the corresponding metric; and the value inside the parenthesis indicates the improvement percentage against the baseline RB approach)

HetRec				
	^1RB ($k = 200$)	^2PB ($k = 300$)	$^3RPBL_{score}$ ($\alpha = 0.3$)	$^4RPBL_{similarity}$ ($k = 300$, $\beta = 0.8 * \frac{\lvert I_u \cap I_v \rvert}{\lvert I_u \cap I_v \rvert + 0.5}$)
MAE	0.6969	0.6835 (1.92%)	$0.6643^{1,2}$(4.67%)	$0.6738^{1,2}$(3.32%)
RMSE	0.9382	0.9362 (0.21%)	$0.8919^{1,2}$(4.93%)	$0.9038^{1,2}$(3.66%)
nDCG-5 (top-10)	0.8421	0.8475 (0.06%)	$0.8689^{1,2}$(3.18%)	0.8593^{1}(2.05%)
MAP (top-10)	0.5348	0.5509 (3.02%)	0.5639^{1}(5.44%)	0.5578^{1}(4.29%)
Yahoo! Moive				
	^1RB ($k = 150$)	^2PB ($k = 200$)	$^3RPBL_{score}$ ($\alpha = 0.3$)	$^4RPBL_{similarity}$ ($k = 200$, $\beta = 0.8 * \frac{\lvert I_u \cap I_v \rvert}{\lvert I_u \cap I_v \rvert + 0.5}$)
MAE	1.0308	1.0089^{1}(2.13%)	$0.9912^{1,2}$(3.84%)	$0.9752^{1,2}$(5.39%)
RMSE	1.5620	1.5515 (0.67%)	$1.4767^{1,2}$(5.46%)	$1.4646^{1,2}$(6.23%)
nDCG-5 (top-10)	0.8178	0.8419 (2.94%)	0.8683^{1}(6.16%)	0.8606^{1}(5.22%)
MAP (top-10)	0.7028	0.7110 (1.17%)	$0.7436^{1,2}$(5.81%)	0.7342^{1}(4.45%)

comparison between $RPBL_{score}$ and $RPBL_{similarity}$ shows that, although they are not significantly different, $RPBL_{score}$ is slightly better as for most measures. The optimal value of weighting parameter α in $RPBL_{score}$ (Eq.5) is always 0.3, suggesting that the personality-based prediction takes more contribution to the final prediction result than the rating based prediction. In comparison, as $RPBL_{similarity}$ emphasizes enhancing user-user similarity by integrating personality with ratings, it is shown that the personality is more helpful for accomplishing its rating prediction task in sparser rating situation (because of the relatively better MAE and RMSE scores in Yahoo! Movie dataset).

Hence, through this experiment, we demonstrate that the personality scores as derived from implicit features can take positive effect on augmenting real-life movie recommendations. In particular, combining them with users' ratings can boost both rating prediction and ranking accuracy to a higher level, relative to the method that considers the implicit personality or users' ratings alone.

5 Conclusion and Future Work

In this paper, we presented an approach to deriving users' personality implicitly from their behavior in movie domain, and furthermore used the derived personality to augment online movie recommendations. Specifically, we first validated the significant correlations between multiple features and users' personality traits through user survey. We then compared three regression models in terms of their ability of unifying these significant features into automatically inferring a user's personality, among which Gaussian Process shows better performance than Pace Regression and M5 Rules. We further implemented three variations of CF method which are all based on the implicitly acquired personality: one is purely based on the inferred personality to enhance user-user similarity in CF process, and the other two combine the personality with users' ratings in either generating the final item prediction score ($RPBL_{score}$) or computing user-user similarity ($RPBL_{similarity}$). The experimental results indicate that the algorithms incorporated with both implicit personality and ratings significantly outperform not only the non-personality approach but also the pure personality-based approach, in terms of both rating prediction and ranking accuracy.

The main limitation of our study is that, as there are not users' explicit personality values in the two real-life movie datasets (HetRec and Yahoo! Movie), it is infeasible to validate our personality inference model with them. We hence plan to investigate other datasets as well as trying to increase the diversity and scale of our user survey's population, so as to further consolidate the recommender approach's practical value. On the other hand, we will investigate more types of behavioral features with the goal of further enhancing our personality inference model. For instance, it may be interesting to study whether people with different personality values would possess different propensity for being influenced by the item's word-of-mouth (WOM) when they give its rating [6]. We will also establish the inference model for other domains, such as music, as motivated by the literatures that show the relationship between music features and user personality [23]. In addition, we believe that our work will be beneficial to solve the cold-start issue in cross-domain recommendation, for which the personality as inferred from user behavior in one domain (e.g., movie) may be used to generate recommendations in another domain (e.g., music).

Acknowledgments. We thank grants ECS/HKBU211912 and NSFC/61272365.

References

1. Cantador, I., et al.: Relating personality types with user preferences in multiple entertainment domains. In: UMAP Workshops (2013)
2. Chamorro-Premuzic, T., et al.: Understanding individual differences in film preferences and uses: a psychographic approach. Journal of the Social Science of Cinema 87 (2014)
3. Chausson, O.: Who watches what?: assessing the impact of gender and personality on film preferences. Paper published online on the MyPersonality project website http://mypersonality.org/wiki/doku.php (2010)
4. Chen, L., et al.: How personality influences users' needs for recommendation diversity? In: CHI Extended Abstracts, pp. 829–834. ACM (2013)
5. De, R.B., Schouwenburg, H.C.: Personality in learning and education: A review. European Journal of Personality **10**(5), 303–336 (1996)
6. Duan, W., et al.: The dynamics of online word-of-mouth and product sales - an empirical investigation of the movie industry. Journal of Retailing **84**(2), 233–242 (2008)
7. Elahi, M., Braunhofer, M., Ricci, F., Tkalcic, M.: Personality-based active learning for collaborative filtering recommender systems. In: Baldoni, M., Baroglio, C., Boella, G., Micalizio, R. (eds.) AI*IA 2013. LNCS, vol. 8249, pp. 360–371. Springer, Heidelberg (2013)

8. Gao, R., et al.: Improving user profile with personality traits predicted from social media content. In: Proc. 7th ACM RecSys, pp. 355–358. ACM (2013)
9. Golbeck, J., et al.: Predicting personality with social media. In: CHI Extended Abstracts, pp. 253–262. ACM (2011)
10. Golbeck, J., Norris, E.: Personality, movie preferences, and recommendations. In: Proc. 5th ASONAM, pp. 1414–1415. ACM (2013)
11. Goldberg, L.R., et al.: The international personality item pool and the future of public-domain personality measures. Journal of Research in Personality **40**(1), 84–96 (2006)
12. Gosling, S.D., et al.: A very brief measure of the big-five personality domains. Journal of Research in Personality **37**(6), 504–528 (2003)
13. Green, P.E., et al.: Thirty years of conjoint analysis: Reflections and prospects. Interfaces **31**(3 suppl.), S56–S73 (2001)
14. Hellriegel, D., Wlldman, R.: Organizational Behavior (Eleventh edition). South-western college, Cincinnati, Ohio (2010)
15. Herlocker, J.L., et al.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) **22**(1), 5–53 (2004)
16. Holmes, G., Hall, M., Prank, E.: Generating rule sets from model trees. In: Foo, N. (ed.) Advanced Topics in Artificial Intelligence. LNCS, vol. 1747, pp. 1–12. Springer, Heidelberg (1999)
17. Hu, R., Pu, P.: A study on user perception of personality-based recommender systems. In: De Bra, P., Kobsa, A., Chin, D. (eds.) UMAP 2010. LNCS, vol. 6075, pp. 291–302. Springer, Heidelberg (2010)
18. Hu, R., Pu, P.: Enhancing collaborative filtering systems with personality information. In: Proc. 5th RecSys, pp. 197–204. ACM (2011)
19. Hu, R., Pu, P.: Exploring relations between personality and user rating behaviors. In: UMAP Workshops (2013)
20. McCrae, R.R., et al.: Age differences in personality across the adult life span: parallels in five cultures. Journal of Developmental Psychology **35**(2), 466 (1999)
21. McIlwraith, R.D.: "I'm addicted to television": The personality, imagination, and TV watching patterns of self-identified TV addicts. Journal of Broadcasting and Electronic Media **42**(3), 371–386 (1998)
22. Rasmussen, C.E.: Gaussian Processes for Machine Learning. MIT Press (2006)
23. Rentfrow, P.J., Gosling, S.D.: 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 (2003)
24. Sarwar, B., et al.: Analysis of recommendation algorithms for e-commerce. In: Proc. 2nd EC, pp. 158–167. ACM (2000)
25. Tkalcic, M., et al.: Personality based user similarity measure for a collaborative recommender system. In: Emotion in HCI Workshops, pp. 30–37 (2009)
26. Wang, Y.: A new approach to fitting linear models in high dimensional spaces. Ph.D. thesis, The University of Waikato (2000)
27. Wu, W., et al.: Using personality to adjust diversity in recommender systems. In: Proc. 24th HT, pp. 225–229. ACM (2013)
28. Yilmaz, E., et al.: A simple and efficient sampling method for estimating AP and NDCG. In: Proc. 31st SIGIR, pp. 603–610. ACM (2008)
29. Zar, J.H.: Significance testing of the spearman rank correlation coefficient. Journal of the American Statistical Association **67**(339), 578–580 (1972)