International Review of Business 19 (2019) 71-86

Relationship between User Rating Behavior and Personality in Recommender Systems

Yoshinori HIJIKATA*

Abstract

Recommender systems have been widely used in one-to-one marketing, in which items are recommended to users based on their interests and preferences. Users are required to input their preference information in recommender systems to receive high-quality recommendations. The users are generally asked to rate several items before using the system. Recently, it has become popular to use information about the personality of the users in recommender systems. Some studies have proposed methods for recommending items based on personality. However, no study has assessed the relationship between the rating behavior of a user and their personality. This study examines the relationship between several features of user behavior for rating items and several factors of personality. A user experiment was conducted in an experimental system, and it was observed that users with high extraversion stop rating at an early stage and those with high neuroticism give high scores to items.

Keywords: One-to-one marketing, Recommender system, Rating, Psychology, Personality, User behavior

1. Introduction

One-to-one marketing is a type of marketing method that adapts its marketing strategies to each user based on his/her needs and demographic attributes (Peppers, Rogers and Dorf 1999). Recommendation is a key technique for realizing adaptation. With this approach, the user is shown products or contents (hereinafter called "items") that he/she is likely to be interested in. This process is usually executed by computers, and the system that realizes the recommendation is called a "recommender system."

Basic ideas and some prototypes of recommender systems were proposed in the early 1990s (Goldberg et al. 1992, Resnick et al. 1994, Shardanand and Maes 1995). The term "information filtering" was initially used for this type of system (Loeb and Terry 1992), but the term "recommender systems" became popular after the ACM (Association for Computing Machinery) edited a special issue called "recommender system" in their journal

* Associate Professor, School of Business Administration, Kwansei Gakuin University

(Resnick and Varian 1997). Early research on recommender systems focused on improving recommendation accuracy. However, it has been reported that accuracy does not necessarily satisfy users and that overall usefulness should be improved in recommender systems (Herlocker et al. 2004). Thus, various types of evaluation metrics such as diversity, novelty and serendipity, as well as algorithms for improving the metrics, were proposed in the late 2000s (Ziegler, C.-N. et al. 2005, Zhang, M. and Hurley, N. 2008, Hijikata et al. 2009).

The quality of user experience is not necessarily improved by simply enhancing the quality of recommendation results (Olmo and Gaudioso 2008). The above article argued that service providers should consider not only the item to be recommended but also when and in what situation should the item be recommended. Thus, studies on user experience in recommender systems have become popular in recent years (Konstan and Riedl 2012, Knijnenburg et al. 2012). User experience is the overall experience of a user in the system. The prediction algorithms, interface, and overall interactions should be designed in the study of user experience (known as UX design) (Hassenzahl and Tractinsky 2006).

In the context of user experience studies, some trials have aimed to import the psychological factors of users in the design of recommender systems. Among many types of psychological factors (Nolen-Hoeksema et al. 2014), personality has been used in these trials. Personality is a psychological concept that comprehensively expresses the inner face of an individual person (Burger 2010). Personality is formed from the intrinsic temperament of a person and the environment that he/she has experienced from his/her childhood; it also determines or controls human behavior.

If there exists an apparent tendency that connects the inner feature of a user and his/her actual behavior, it can be utilized to anticipate the next action of the user, which means that it can predict items that the user will purchase. Recommender systems that exploit the personalities of users are sometimes called “personality-based recommender systems.” This approach may improve the quality of recommendation results, the degree of personalization, and the user experience in recommender systems (Hu 2010, Hu and Pu 2010).

The process of recommender systems is generally divided into three parts: (I) input of preference, (P) prediction of the preference of the user, and (O) output of the recommendation results; this process is called the “O-I-P model” (Output-Input-Process model) (Konstan and Riedl 2003). It is expected that the user experience will be improved when the relationship between the personality of the user and their behavior in each stage of the O-I-P model is clarified. Studies have focused on each stage of the O-I-P model in exploiting personality to recommend items. In the process stage (P), some studies examine whether the personality of the user can explain his/her preference for items, especially in a cold-start case when only a small number of ratings are obtained. In the input stage (I), two types of user input, a traditional rating and a personality questionnaire survey, are examined to determine which one will be preferred by the user. In the output stage (O), the system examines whether users with a specific personality type prefer diversified recommendations of topics (genres). However, there are only a limited number of studies on personality in the research field of recommender systems.

This study focuses on the relationship between the personalities of the users and their behavior in the input stage of the O-I-P model. In the input stage, users are usually required to input rating values for items selected by the system, which will subsequently be used for learning their preferences using machine-learning techniques. The user behavior in rating items may differ according to their personality. For example, improper users, in other words, users with low conscientiousness, might not give ratings to many items. Rating values of emotional users, in other words, users with high neuroticism might not be stable due to their capricious temperament. This study examines the following behavioral features of users: number of ratings, time required for ratings, and statistics (variance, bias, and stability) of rating values. These features will be assessed in regard to the personality of a user.

The big five personality traits, also known as the five-factor model, are used in this study as personality traits. This model represents personality using five factors: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (Digman 1990, Goldberg 1990, Costa and McCrae 1992). This study examines the relationship between the five features of user behavior in the input stage of the O-I-P model and the five factors of personality included in the big five personality traits.

The rest of this paper is organized as follows: Section 2 introduces works related to this study; Section 3 presents the details of the O-I-P model and personality traits; Section 4 details the procedure of the experiment conducted in this study; the experimental results and a discussion of the results are given in Section 5 and 6, respectively; finally, conclusions and future work are presented in Section 7.

2. Related Work

Some studies have exploited personality to build recommender systems. This section introduces these studies for each stage of the O-I-P model.

For the input stage (I), Hu and Pu (2009) compared two types of methods for inputting user information. One method is item rating that asks users to rate items on a Likert scale; the other method is a personality test that evaluates the personality of a user. The former is used in many recommender systems. Evaluation metrics used in this work include perceived accuracy, user effort (psychological effort and temporal cost), and user loyalty (whether the user wants to use the recommender system again). The experimental results demonstrated that there is no difference in perceived accuracy between the two types of input methods. However, the personality test outperformed the item rating in psychological effort, temporal cost, and user loyalty. The researchers also developed a music recommender system and compared these two types of input methods (Hu and Pu 2010). They conducted a user experiment with a subjective evaluation and found that domain experts (music lovers) prefer item rating while domain novices prefer a personality test.

For the process stage (P), Hu and Pu (2011) developed a recommendation algorithm that combines a predictive score calculated by collaborative filtering and a personality score obtained from a personality test. In detail, the authors proposed three types of recommendation algorithms: (i) a recommendation method based on the personality information of a user, (ii) a linear combination of both personality and rating information and (iii) a cascade mechanism to leverage both resources (first, a personality-based algorithm is used, and traditional collaborative filtering is used after the rating matrix becomes dense). The experimental results demonstrated that these algorithms outperform the traditional rating-based collaborative filtering algorithm for sparse datasets and new users, indicating that these algorithms resolve the cold-start problem in recommender systems.

For the output stage (O), Tintarev et al. (2013) examined the influence of personality on the diversity of recommendation results. The authors focused on openness to experience among the big five personality traits (the detail of the factors will be explained in Section 3) and examined the acceptance tendency for diversified recommendation results. In detail, they conducted a user experiment in which users were asked to provide a book recommendation to their friends. The results showed that the users tend to diversify the genre of books in the recommendation list to their friends with high openness to experience.

Although studies on personality-based recommender systems have become popular in recent years, research results are still limited. This study conducts a fundamental investigation on personality for the input stage of the O-I-P model.

3. Personality and the O-I-P Model

The details of the big five personality traits and user behavior in the input stage of the O-I-P model are explained in this section.

3.1 Big Five Personality Traits

The model of the big five personality traits or the five-factor model is a model of popular psychological traits that represents the personality of a person (or characteristics). These psychological traits are practical and reliable compared to other personality traits (Nunes 2008). Thus, the author uses these traits in this study. These traits consist of the following five factors (Costa and McCrae 1992):

Openness to experience: Tendency to appreciate art, new products, adventure, and unusual ideas. Having intellectual curiosity and seeking a variety of experiences.

Conscientiousness: Tendency to be organized and dependable, show self-discipline, act dutifully, aim for achievement, and prefer planned rather than spontaneous behavior.
Extraversion: Tendency to seek stimulation in the company of others. Being energetic, sociable, and talkative.
Agreeableness: Tendency to be compassionate and cooperative rather than suspicious and antagonistic toward others. Prefers collaborative works with others.
Neuroticism: Tendency to be prone to psychological stress and to easily experience unpleasant emotions such as anger, anxiety, depression, and vulnerability.

3.2 O-I-P Model

The O-I-P model represents the process of recommender systems, consisting of preference input (I), preference prediction (P), and recommendation output (O). This subsection details the input stage (I) of the model analyzed in this study.

In the input stage, users usually input rating values to items selected by the system, which presents their degree of preference or interest. Users usually input the rating values in five- or seven-point Likert scales. This input method is widely used in commercial recommender systems because it enables users to easily express their preference or interest.

However, this approach also has a drawback: fluctuations (Hill et al. 1995, Cosley et al. 2003, Amatriain et al. 2009) and bias (Weigend 2003, Kamishima and Akaho 2006) exist in the rating values. Fluctuation is a phenomenon in which the rating value given by the user at a certain point in time differs from what he/she gave at another point in time. Thus, the rating values are not coherent or stable over time. Bias is a phenomenon in which the rating values are unevenly distributed on a Likert scale. For example, the rating values given by a user all fall on the positive or negative side of the scale. These types of phenomena can be observed for any user, but the degree of fluctuation and the direction of bias might differ for each user. In particular, differences may be caused by the personality of a user.

Other behavioral features that might be influenced by personality include rating time, rating quantity (the number of ratings (checking)), and the dispersion of rating values. These features may also affect recommendation performance. In particular, the rating quantity will directly affect the recommendation performance because most recommender systems exploit machine-learning techniques, and the performance of machine-learning algorithms generally improves when a larger amount of learning data is given. This study examines the relationship between the above-mentioned five types of behavioral features and personality.

4. Experimental Method

The evaluation was performed by employing a user experiment. This section gives the details of the experimental method.

4.1 Introduction of Pinterest

Sixteen users, aged 21 to 25 years, participated in the experiment. The participants are Japanese students studying system engineering at a university. An overview of the experimental tasks of the users is described below. First, the users provide rating values to items shown by the recommender system. This task (hereinafter called the "rating experiment") is conducted again after a specific time interval (10 days) to measure fluctuations in the ratings. Subsequently, the participants complete a questionnaire to assess their personality. Finally, the participants complete a questionnaire that asks them the degree of interest in the item domain (movie). In summary, the experimental tasks are as follows:

1. First-round experiment
   - Rating experiment (first round)
2. Second-round experiment (10 days later)
   - Rating experiment (second round)
   - Personality test
   - Questionnaire on interest in movies

4.2 Rating Experiment

The users were asked to provide rating values for up to approximately 3000 items. Movies were selected as the item domain because they are popular among many people regardless of gender or age. The movie data used in this experiment were acquired from the movie rankings of TSUTAYA[1]. In detail, 200 movies were selected from the ranking list of each genre (only major genres were selected as a source ranking). A total of 2775 movies were acquired after removing any overlaps.

In the rating experiment, the users were asked to imagine a situation in which they had just started using the recommender system. The instructions presented to the users were as follows: "Recommender systems (e.g., "Your favorite" in Amazon) ask users to give rating values to several items before receiving a recommendation. These values represent the degree of interest or preference for an item. Recommender systems infer the favorite items of a user based on the given rating values and show items with a high predicted preference as a recommendation result." These instructions ensure that the users have at least a minimal knowledge of recommender systems and give them motivation to rate items.

---

1 https://movie-tsutaya.tsite.jp/netdvd/dvd/hotrankingTopTotal.do?pT=0

The users can rate as many items as they want, until they are tired of rating or feel that they have rated enough items to receive a high-quality recommendation. The rating methods are generally categorized into binary ratings (give ratings as “like” or “dislike” or simply select favorite items) and discrete ratings (give ratings with discrete values). Because this study examines the fluctuation or variance of ratings, the latter method is used in the experiment. In detail, the users provide rating values on a 7-point Likert scale (Likert 1932). The rating experiment was conducted again on each user after a 10-day interval.

The author developed an experimental recommender system, as shown in Figure 1. After the “Start” button is pushed, the experimental session will start, and items will be shown in the lower part of the window. The item information (movie title and image) is shown in the window with a 7-point star scale. The users can rate items one at a time as they scroll the window (new items are shown when scrolling down). The users can give rating values by clicking the appropriate star. They can quit rating at any time in the session. Furthermore, the experiment is stopped by clicking the “End” button.

Figure 1. Screenshot of the experimental recommender system

4.3 Behavioral Features

The measurements of the behavioral features are as follows:

Rating time: This factor indicates how much time is taken by the user in rating, and is defined as the time between pushing the “Start” button and pushing the “End” button.

Rating quantity: This parameter denotes how many items are rated by the user. The following two types of measurements are introduced: (1) the number of items the user has checked (browsed) for rating and (2) the number of items rated by the user. The former is introduced because the user does not have to give a rating value if he/she is not familiar with an item.

Dispersion of rating values: This factor indicates whether the user employs the evaluation scale for rating items, from the lower end to the higher end. The variance of rating values is used in this study.

Bias: This parameter indicates whether the user tends to give a rating value on the lower side or higher side of the scale. This value is calculated from (1) the mean of the rating values, (2) the ratio of higher ratings (7-star, 6-star, and 5-star ratings on a 7-point scale), and (3) the ratio of lower ratings (1-star, 2-star, and 3-star ratings on a 7-point scale).

Fluctuation: This term represents the coherence or credibility of the rating values given by the user. The fluctuation is calculated from (1) the Pearson’s correlation coefficient (Freedman et al. 2007) for two corresponding rating values and (2) the root mean square of the deviation (RMSD) between the two rating values. The two rating values are the rating value given in the first round and that in the second round. The first fluctuation term is determined from the following equation:

$$r = \frac{\sum_{i=1}^{n}\left(x_{1,i} - \bar{x}_1\right)\left(x_{2,i} - \bar{x}_2\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{1,i} - \bar{x}_1\right)^2}\sqrt{\sum_{i=1}^{n}\left(x_{2,i} - \bar{x}_2\right)^2}}$$

The second term is represented by the following equation:

$$r = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{1,i} - x_{2,i}\right)^2}$$

Note that $x_{1,i}$ and $x_{2,i}$ represent the rating value from the first round and that from the second round, respectively.

Table 1. Examples of questions in the NEO-FFI

Factor	Question example
Openness to experience	Q1: I have original ideas. Q2: I am interested in new things. Q3: I do not repeat the same thing.
Conscientiousness	Q1: I am rather lazy. Q2: I am the type of person who decides to do things carelessly. Q3: I work and study energetically.
Extraversion	Q1: I am rather cheerful. Q2: I am not good at talking in front of people. Q3: I do not insist on my opinion.
Agreeableness	Q1: I am the type of person who has compassion for others. Q2: I am the type of person who thinks from the viewpoints of others. Q3: I will do anything to help others.
Neuroticism	Q1: I am always worried about something. Q2: I am the type of person who cares about things that do not matter. Q3: I am the type of person who always feels nervous.

4.4 Personality

The NEO-FFI (NEO Five-Factor Inventory) (Costa and McCrae 2008) was used as personality traits in this study to quantitatively assess the personalities of the users. The Japanese NEO-FFI (translated from the original NEO-FFI (English version) to Japanese by Saccess Bell Corp.[2]) was used in the experiment because the target test subjects are Japanese. Table 1 presents part of the NEO-FFI questionnaire. The user answers each question on five scales. The NEO-FFI consists of 60 questions (12 questions for each factor) regarding personality. The respondents (participants) of the NEO-FFI answer each question on a five-point Likert scale. Each response to a question has a score ranging from 0 to 4. The primary score of each factor can be calculated by summing the scores of the 12 questions. The primary scores are converted into T-scores (deviation values) using the mean and standard deviation for each gender. This score is calculated as the mean reaches 50 and was used as the intensity of each personality factor in this study.

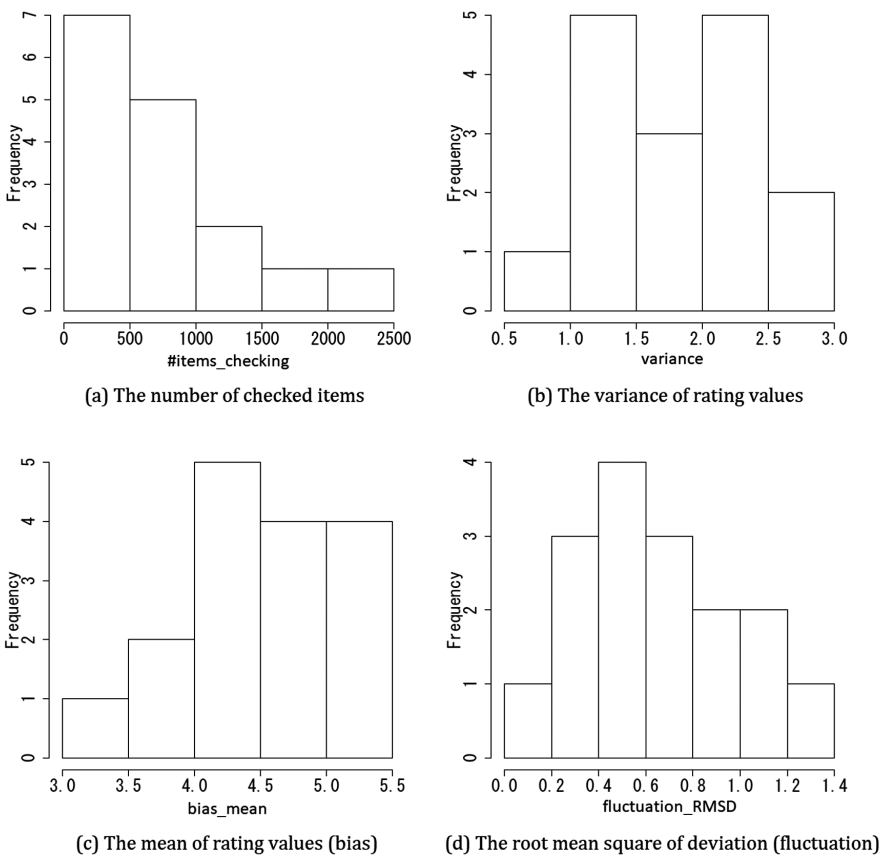
2 http://www.saccess55.co.jp/

5. Results

This section presents descriptive statistics of the behavioral features and personality and their corresponding correlations.

5.1 Descriptive Statistics of Behavioral Features

Among the nine measurements of behavioral features, only the primary measurements are shown here. Figure 2 shows the number of items browsed by the user for rating, the variance of rating values, the mean of the rating values, and the RMSD. Interestingly, the number of browsed items appears to follow a power distribution. Although the graph is slightly irregular, the variances of the rating values for most users range from 1.0 to 2.5. A bias is observed, as the users tend to give higher values. There exists some fluctuation for most of the users, ranging from 0 to 1.4.



(a) The number of checked items

(b) The variance of rating values

(c) The mean of rating values (bias)

(d) The root mean square of deviation (fluctuation)

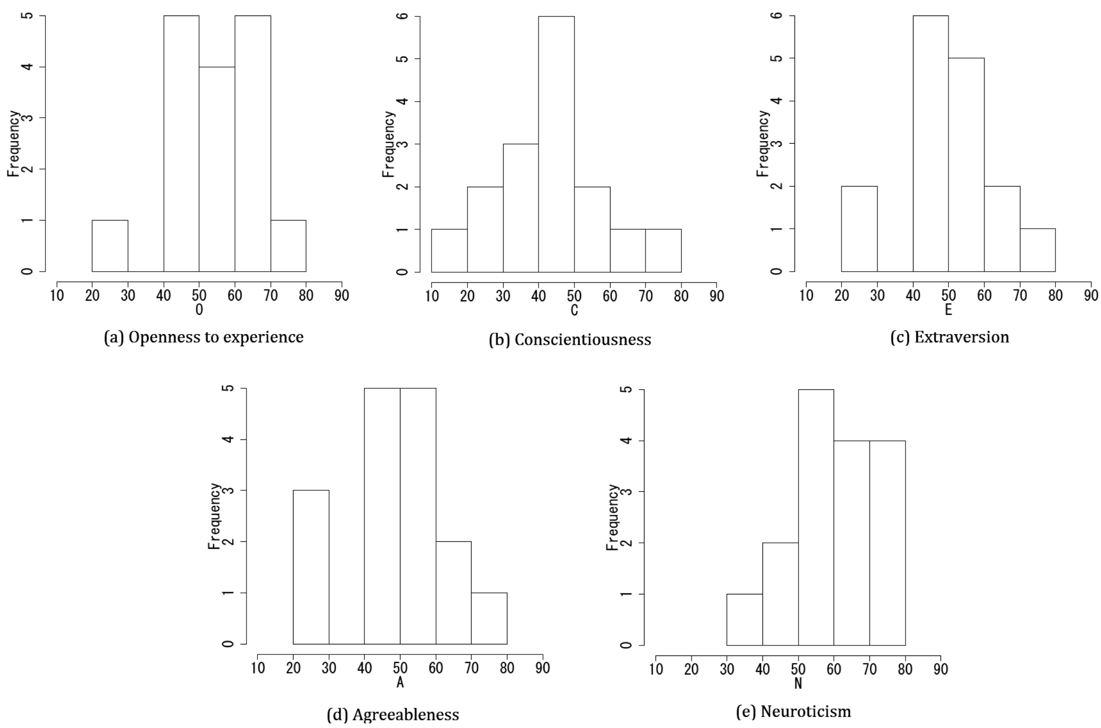
Figure 2. Descriptive statistics of behavioral features

5.2 Descriptive Statistics of Personality

The mean and standard deviation for each personality factor and a corresponding histogram are shown in Table 2 and Figure 3, respectively. The mean for neuroticism is generally over 50 (the average of the general population), and the mean of conscientiousness is generally under 50. According to a survey conducted by Saccess Bell Corp., the average for neuroticism in university students is fairly high, and that of conscientiousness is relatively low compared to the general population. Thus, the result of this experiment shows that the participants comprise a general set of university students. The averages for extraversion and agreeableness are similar to those for the general population. The average for openness to experience is considerably high compared to the average for the general

Table 2. Descriptive statistics of personality scores

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Mean	53.5	43.6	50.0	48.9	60.4
S.D.	11.7	14.8	14.1	13.7	12.5



(a) Openness to experience (b) Conscientiousness (c) Extraversion

(d) Agreeableness (e) Neuroticism

Figure 3. Histograms of personality traits

population, perhaps because the participants of this experiment are university students majoring in system engineering.

5.3 Relationship between Behavioral Features and Personality

A correlation analysis was performed to clarify the relationship between the nine measurements of five behavioral features and the five factors of personality, with the results shown in Table 3. A test of no correlation was conducted to determine the presence of correlation. Statistically significant differences were observed among some pairs of behavioral features and personality ($p < .05$, shown in bold font and marked by*). The neuroticism score has a positive correlation with the mean rating values (bias) and a negative correlation with the ratio of lower ratings. The extraversion score has a negative correlation with the number of browsed items. However, there is no statistically significant difference for openness to experience, conscientiousness, or agreeableness.

Table 3. Results of the correlation analysis between rating behavior and personality
(a) Rating time, rating quantity and dispersion of rating values

	Rating time	Rating quantity		Variance
		#checking	#rating	
Openness	-.272	-.037	-.154	.114
Conscientiousness	.102	-.030	-.039	.263
Extraversion	-.319	**-.595***	.006	-.391
Agreeableness	-.349	-.150	-.257	-.065
Neuroticism	.124	.213	-.037	-.163

(b) Rating time, rating quantity and dispersion of rating values

	Bias			Fluctuation	
	Mean	%high-rate	%low-rate	Pearson's r	RMSD
Openness	-.358	-.274	.272	.062	.112
Conscientiousness	-.164	-.258	.250	.426	-.165
Extraversion	-.321	-.171	.161	.055	-.118
Agreeableness	-.149	-.164	.160	.073	-.075
Neuroticism	**.549***	.469	**-.577***	-.291	.069

* $p < .05$

6. Discussions

It was observed that people with a higher level of neuroticism do not tend to give low rating values. Although it is difficult to discern the reason for this trend, a possible explanation is given here. As described in Section 3.1, people with higher neuroticism are not psychologically stable or good at controlling their emotions. It is thought that such people do not think logically and are controlled by temporal emotions. It is possible that these participants did not give rating values according to a fixed standard in their mind. However, it remains unclear why these participants tended to prefer higher values. People with high neuroticism appear to prefer movies, from the results, but this cannot be a strong explanation because there is a weak negative correlation between neuroticism and the degree of interest in movies. Although these participants might be easily moved by the stories in movies because they are psychologically unstable and emotional, there is no clear evidence for this inference.

It was found that people with high extraversion do not tend to check many items for rating. As described in Section 3.1, people with high extraversion like to keep busy and prefer communication with people or activities in the real world. The task of inputting rating values for movies may be tiresome for them, which would prevent them from continuing to rate movies.

The author initially expected that people with low conscientiousness would not rate many items, while people with high neuroticism would not give stable values. However, these hypotheses were not supported by the experiment. Although the underlying reason for these results remains unclear, the users may not consider that the task is for another person. They may perceive that the task is performed for the user himself/herself (actually, rating is an action for the user himself/herself to receive high-quality recommendations). Thus, there is no difference among people with different levels of conscientiousness. Although people with high neuroticism gave stable rating values, their rating values tended to be high. This result differs from the original expectation of the author but may be caused by the emotional tendencies of people with high neuroticism.

There are two primary implications to be derived from the above results. One, the rating values for people with high neuroticism should be normalized. Because they do not use lower values, it is better to widen the dispersion of rating values to obtain an accurate interest or preference of these users. Two, some tricks or environmental aspects should be introduced to the item ratings for people with high extraversion. For example, the system could require these users to rate a minimum number of items.

As a limitation to this study, the examination was performed only for a 7-point scale rating system. In recommender systems, it is also popular to ask users to indicate their favorites among items presented by the system. Because this method is less burdensome to users, people with higher extraversion might check more items. Another limitation is that the experiment did not assume collaborative filtering. The users were simply told that the rating values would be used to infer their interests or preferences. If it is explained that the rating values will be used by collaborative filtering algorithms (where the rating values of one user are used to infer the degree of interest or preference of other users), people with high agreeableness might check and rate more items. In future research, other types of rating methods or experimental situations could be examined.

7. Conclusions

This study investigated the relationship between user behavior and personality in the input stage of the O-I-P model for recommender systems. Nine kinds of rating-behavior measurements and five factors of personality (big five personality traits) were examined in a user experiment. In the experimental results, statistically significant differences were found in the correlation between the rating bias (the mean of rating values) and neuroticism and that between the rating quantity (the number of checked items) and extraversion. However, no significant difference was found for openness to experience, conscientiousness, or agreeableness. In future work, the same experiment can be conducted for other settings of recommender systems.

Acknowledgment: This work was supported by JSPS KAKENHI Grant Number 15K12150.

References

Amatriain, X., Pujol, J. M. and Oliver, N. (2009) I like it... i like it not: Evaluating User Ratings Noise in Recommender Systems, *User Modeling, Adaptation, and Personalization*, UMAP 2009, Lecture Notes in Computer Science, 5535, Springer, pp. 247-258.

Burger, J. (2010) PSY 235 Theories of Personality Series, Cengage Learning.

Cosley, D., Lam, S. K., Albert, I., Konstan, J. A. and Riedl, J. T. (2003) Is Seeing Believing?: How Recommender System Interfaces Affect Users' Opinions, *in Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'03)*, pp. 585-592.

Costa, P. and McCrae, R. R. (1992) The NEO-PI-R Professional Manual: Revised NEO Five-Factor Inventory (NEO-FFI), *Psychological Assessment Resources*, 1992.

Costa, P. and McCrae, R. R. (2008) The Revised Neo Personality Inventory (NEO-PI-R), *The SAGE Handbook of Personality Theory and Assessment*, 2, pp. 179-198.

Digman, J. M. (1990) Personality Structure: Emergence of the Five-factor Model, *Annual Review of Psychology*, 41, pp. 417-440.

Freedman, D., Pisani, R. and Purves, R. (2007) Statistics, W. W. Norton & Company.

Goldberg, L. R. (1990) An Alternative "Description of Personality": The Big-Five Factor Structure, *Journal of Personality and Social Psychology*, 59(6), pp. 1216-1229.

Goldberg, D., Nichols, D., Oki, B. M. and Terry, D. (1992) Using Collaborative Filtering to Weave an Information Tapestry, *Communications of the ACM*, 35(12), pp. 61-70.

Hassenzahl, M. and Tractinsky, N. (2006) User Experience - A Research Agenda, *Behaviour & Information Technology*, 25(2), pp. 91-97.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004) Evaluating Collaborative Filtering Recommender Systems, *ACM Transactions on Information Systems (TOIS)*, 22(1), pp. 5-53.

Hijikata, Y., Shimizu, T., and Nishida, S. (2009) Discovery-oriented Collaborative Filtering for Improving User Satisfaction, *in Proceedings of the 14th ACM International Conference on Intelligent User Interfaces (IUI'09)*, pp. 67-76.

Hill, W., Stead, L., Rosenstein, M. and Furnas, G. (1995) Recommending and Evaluating Choices in a Virtual Community of Use, *in Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'95)*, pp. 194-201.

Hu, R. and Pu, P. (2009) A Comparative User Study on Rating vs. Personality Quiz based Preference Elicitation Methods, *in Proceedings of the 14th ACM International Conference on Intelligent User Interfaces (IUI'09)*, pp. 367-372.

Hu, R. (2010) Design and User Issues in Personality-based Recommender Systems*, in Proceedings of the fourth ACM Conference on Recommender Systems (RecSys'10)*, pp. 357-360.

Hu, R. and Pu, P. (2010) A Study on User Perception of Personality-based Recommender Systems, *User Modeling, Adaptation, and Personalization*, UMAP 2010, Lecture Notes in Computer Science, 6075. Springer, pp. 291-302.

Hu, R. and Pu, P. (2011) Enhancing Collaborative Filtering Systems with Personality Information, *in Proceedings of the fifth ACM Conference on Recommender Systems (RecSys'11)*, pp. 197-204.

Kamishima T. and Akaho S. (2006) Nantonac Collaborative Filtering - Recommendation based on Multiple Order Responses, *in Proceeding of the 1st International Workshop on DataMining and Statistical Science*, pp. 117-124.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H. and Newell, C. (2012) Explaining the User Experience of Recommender Systems, *User Modeling and User-Adapted Interaction*, 22(4-5), pp. 441-504.

Konstan, J. A. and Riedl, J. T. (2012) Recommender Systems: from Algorithms to User Experience, *User Modeling and User-Adapted Interaction*, 22(1-2), pp. 101-123.

Likert, R. (1932) A Technique for the Measurement of Attitudes, *Archives of Psychology*.

Loeb, S. and Terry, D. (1992) Information Filtering, *Communications of the ACM*, 35(12), pp. 26-81.

Nolen-Hoeksema, S., et al. (2014) Atkinson & Hilgard's Introduction to Psychology, Wadsworth Publishing Company.

Nunes, M. (2008) Recommender Systems based on Personality Traits, Ph. D. Thesis University Montpellier.

Olmo, del F. H. and Gaudioso, E. (2008) Evaluation of Recommender Systems: A New Approach, *Expert Systems with Applications*, 35, pp. 790-804.

Peppers, D., Rogers, M. and Dorf, B. (1999). Is Your Company Ready for One-to-one Marketing?, *Harvard Business Review*, January-February 1999 Issue, pp. 151-160.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. T. (1994) GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *in Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'94)*, pp. 175-186.

Resnick, P. and Varian, H. (1997) Recommender Systems, *Communications of the ACM*, 40(3), pp. 56-89.

Shardanand, U. and Maes, P. (1995) Social Information Filtering: Algorithm for Automating Word of Mouth, *in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'95)*, pp. 210-217.

Tintarev, N., Dennis, M. and Masthoff, J. (2013) Adapting Recommendation Diversity to Openness to Experience: A Study of Human Behaviour, *User Modeling, Adaptation, and Personalization*, UMAP 2013, Lecture Notes in Computer Science, 7899, Springer, pp. 190-202.

Weigend, A. S. (2003) Analyzing Customer Behavior at Amazon.com, *in Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*, Keynote, p. 5.

Zhang, M. and Hurley, N. (2008) Avoiding Monotony: Improving the Diversity of Recommendation Lists, *in Proceedings of the Second ACM Conference on Recommender Systems (RecSys'08)*, pp. 123-130, 2008.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., and Lausen, G. (2005) Improving Recommendation Lists through Topic Diversification, *in Proceedings of the 14th ACM International Conference on World Wide Web (WWW'05)*, pp. 22-32.