Using Personality to Adjust Diversity in Recommender Systems

Wen Wu
[1.] Department of Computer Science and Technology
East China Normal University
Shanghai, China
[2.] Department of Computer Science
Hong Kong Baptist University
Kowloon, Hong Kong
51111201013@ecnu.cn

Li Chen
[1.] Department of Computer Science
Hong Kong Baptist University
Kowloon, Hong Kong
[2.] Division of Science and Technology
BNU-HKBU United International College, China
lichen@comp.hkbu.edu.hk

Liang He
Department of Computer Science and Technology
East China Normal University
Shanghai, China
lhe@cs.ecnu.edu.cn

ABSTRACT
Nowadays, although some approaches have been proposed to enhance the diversity in online recommendations, they neglect the user's spontaneous needs that might be possibly influenced by her/his personality. Previously, we did a user survey that showed some personality dimensions (such as *conscientiousness* which is one of personality factors according to the big-five factor model) have significant impact not only on users' diversity preference over items' individual attributes, but also on their overall diversity needs when all attributes are combined. Motivated by the findings, in the current work, we propose a strategy that explicitly embeds *personality*, as a moderating factor, to adjust the diversity degree within multiple recommendations. Moreover, we performed a user evaluation on the developed system. The experimental results demonstrate an effective solution to generate personality-based diversity in recommender systems.

Categories and Subject Descriptors
H1.2 [User/Machine Systems]: Human Factors, Software Psychology; H5.2 [Information interfaces and presentation]: User Interfaces – evaluation/methodology, interaction style.

General Terms
Design, Experimentation, Human Factors.

Keywords
Personality-based recommender systems, diversity, user evaluation.

1. INTRODUCTION
Recommender systems have been popularly applied in the current Web environment, with the primary aim of eliminating the information overload and assisting users in efficiently locating interesting items (e.g., movies, books, music). In recent years, in addition to improving the content-based and collaborative filtering based recommender approaches [2], some attentions have been paid to study whether/how users' inherent interests are potentially affected by their psychological characteristics, such as personality. Indeed, prior studies in the area of psychology showed that personality can likely affect users' attitudes, tastes and behavior, motivated by which a few researchers have started to build the so called *personality-based recommender system* [9, 14]. Their focus has mainly on revealing the impact of personality on the user's preference over a single item or an attribute (e.g., music genre). However, given that a recommender system normally returns multiple recommendations at a time to the user, it should be meaningful to further investigate how personality affects users' perceptions of N ($N > 1$) recommendations. The answer to this question can be helpful to solve the current challenging issue of how to effectively adjust the *diversity* degree within a set of recommendations. Actually, it has been widely recognized that the recommended N items should not be too similar to each other, so as to allow users to discover various unexpected items that they may be more interested in [12]. Unfortunately, though *diversity* has emerged as an important metric, its ideal balance with *similarity* has not been well solved. Existing approaches usually adopted a fixed strategy to control the recommendations' diversity degree from the algorithm's perspective [10, 13], which however is not personalized to individual users' spontaneous needs.

Due to these limitations, previously, we conducted a user survey (with 181 participants) to identify whether people, with different personality values, would have different diversity needs [3]. For each user, we obtained her/his movie selections as well as personality values [1]. Specifically, two levels of analysis were performed: the diversity (within the user's selections) in respect of the movie's individual attributes (like genre, director, actor/actress, etc.); and her/his item selections' overall diversity when all attributes are combined. The correlation results showed that some personality factors have significant impact on users' diversity needs. For instance, it suggests that more reactive, excited and nervous person is more inclined to choose diverse directors (which is related to the *neuroticism* personality factor), and suspicious/ antagonistic users (related to a*greeableness* factor) prefers the diversity w.r.t. movie country. As for the movie's release time, its

[1] The personality was assessed via the representative big-five factor model [11] that defines personality as five factors: Openness to Experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism.

24th ACM Conference on Hypertext and Social Media
1–3 May 2013, Paris, France.

diversity is preferred by efficient/organized users (related to *conscientiousness* factor), and for the movie's actor/actress, its diversity is preferred by imaginative /creative users (*openness* factor). At the second level of analysis regarding the overall diversity, *conscientiousness* was shown significantly negatively correlated with it. That is, people with low conscientiousness value preferred high level of diversity, no matter of how the weights placed on different attributes were varied. More details of this experiment's procedure and results can be found in [3].

1.1 Contribution of Our Current Work

Inspired by the previous work, in this paper, we report the follow-up implementation of a personality-based strategy for adjusting recommendations' diversity, in accordance with the prior survey's results. Furthermore, we have performed system evaluation in the form of a user study. By means of comparing our system to a variant that incorporated personality in the contrary way (i.e., offering less diverse items to the user though s/he spontaneously requires higher level of diversity given her/his personality values), we have found that users perceived our system significantly more accurate and helpful. Users' overall satisfaction with the system is also higher. The findings thus not only consolidate the previous survey's results, but also suggest an effective solution in terms of taking personality into account for generating more personalized diverse recommendations.

In the following, we first introduce related works on diversity and personality studies in recommender systems (Section 2). We then present our system implementation (Section 3), followed by the experiment setup and results analysis (Section 4). At the end, we draw the conclusion (Section 5).

2. RELATED WORK

The related work can be classified into two branches: diversity studies; and personality studies, in recommender system.

As for the *diversity* study, it has been first recognized that a good recommender should offer a diverse set of items instead of too similar ones, as they may encourage users to select from a broader range [5, 12]. Some researchers have hence endeavored to achieve the ideal balance between the two conflicting objectives, similarity and diversity, from the algorithm's perspective. For example, Smyth and McClave proposed the bounded greedy selection algorithm and proved that it can improve recommendations' diversity, without significantly compromising their similarity [13]. Ziegler *et al.* developed the topic diversification, which is a heuristic algorithm based on taxonomy similarity, to increase the diversity in the recommendation list [15]. Adomavicius and Kwon further proposed a sophisticated graph-theoretic approach that models the diversity maximization problem as the network flow maximization problem, which was demonstrated to achieve good performance in terms of both accuracy and diversity [1]. Hurley and Zhang [10] regarded the tradeoff between similarity and diversity as a binary optimization problem and defined a controller to explicitly tune the two metrics for obtaining the optimal tradeoff.

In parallel, it has been found that *personality* can be leveraged into addressing the cold-start problem in user-based recommender systems. For instance, Tkalcic and Kunaver integrated personality, obtained based on the big-five personality model, to enhance the nearest neighborhood measure in the collaborative filtering systems [14]. The results showed that the personality based measure is more effective than the rating based. Hu and Pu also exerted to solve the cold-start problem by incorporating human's personality properties [9]. They found that the recommender system that considers users' personality is more effective in terms of increasing users' loyalty towards the system and decreasing their cognitive effort, relative to the non-personality based system [8]. Some commercial sites such as Whattorent [2] also use personality quiz to identify a user's interest profile, before giving him/her the movie recommendations.

2.1 Limitation of Related Work

The limitation of related work on recommendations' diversity is that they have just emphasized the algorithm development, but rarely investigated whether users would be inherently influenced by their own personality in terms of the need for recommendation diversity. On the other hand, the existing personality studies have less considered the potential effect of personality on affecting users' attitudes and behavior within multiple recommendations. Thus, we have been driven to reduce the gap between the two branches of research. As noted before, we have previously revealed the causal relationship from several personality factors to users' diversity needs [3]. Thus, in the current work, we are more interested in exploiting these results for actually enhancing the generation of recommendations in a concrete system.

3. SYSTEM IMPLEMENTATION

For the system implementation, we still use movie as the sample product domain, so as to be consistent with our previous survey. We have concretely incorporated personality, as a moderating factor, into a content-based recommender system. Its primary function is to adjust the diversity degree within a set of N recommendations that the system presents to a user at a time. The content-based recommending technique has indeed been one of successful approaches applied in commercial products such as Amazon and Pandora. Therefore, our work can be potentially beneficial to both academia and industry researchers.

3.1 Recommending Process

Our algorithm is under the assumption that people usually just focus on one or two attributes (that are most important for them) when choosing movies. The algorithm steps are listed in Figure 1. The meanings of parameters are as follows:

u — an active user;

$PS_u[5]$ — an array contains the five personality values (obtained via the big-five personality quiz [6]) of the user u;

$Pref_u[5]$ — an array includes the user's initially stated preferences over the movie's major attributes: {(*attribute*, *value*)} where the *attribute* is among the set {genre, director, country, release time, actor/actress}, and *value* is the user's criterion. For example, a user's preferences are {(genre, *action*), (director, *Steven Spielberg*), (country, *None*), (release time, *1990s*), (actor, *Tom Cruise*)} ("*None"* means that no preference was stated on that attribute). For each user, the tuples are stored in the order of her/his weights placed on these attributes (i.e., the tuple with the most important attribute is positioned at the first one in the array). To obtain the weight information, we can ask the user to perform the conjoint analysis [4]. Concretely, we can generate 16 cards via the orthogonal setup, reflecting different combinations of weights (for example, one card is with the weights on attributes {genre, director, country, release time, actor} as {4,3,2,1,5}, and another card is {5,3,4,3,3}; every weight is in the range 1 "least

[2] http://www.whattorent.com/

important" to 5 " most important"). We can then ask the user to rank those cards according to her/his preferences. From the ranking, we can infer her/his weights on these attributes.

As it can be seen from Figure 1, the set of N recommendations is composed of two subsets which are generated from the algorithm (N = 10 in our experiment). One subset contains m movies that best match the user's stated preferences on important attributes (with emphasis on the *similarity*). Another subset includes n movies that take into account the causal relation from the user's personality values to her/his *diversity* need in respect of the most important attribute. The two parameters m and n ($m + n = N$) are determined via the diversity adjusting strategy (see the next section).

Algorithm Steps

Input:

An active user u; u's personality information $PS_u[5]$;

u's preference information $Pref_u[5]:\{(attribute, value)\}$;

Output:

A recommendation list with N movies

Recommend (u, $PS_u[5]$, $Pref_u[5]$)

1. begin
2. /* generate m similar recommendations for user u according to $\{(attribute_1, value_1), (attribute_2, value_2)\}$ */
   $R_1 \leftarrow$ select m movies from MovieSet
   where ($attribute_1$=$value_1$ or $attribute_2$=$value_2$)
3. /* generate n diverse recommendations for user u according to u's needs for diversity and $Pref_u[5]$ */
   $R_2 \leftarrow$ select n movies from MovieSet
   via diversity adjusting strategy (section 3.2)
4. return ($R_1 \cup R_2$)
5. end

Figure 1. Algorithm steps.

3.2 Diversity Adjusting Strategy

Here we explain how the diversity is adjusted within the set of N movies by taking into account users' personality values. Firstly, for the current user, we convert each of her personality factors' values into one of three levels: *High Level*, *Middle Level*, and *Low Level* [6]. The system will then map it to the user's diversity need in respect of her/his most important attribute, through checking the correlation results as reported in [3]. For example, *High Level* of *Openness* is linked to high need for diversity w.r.t. "actor/actress". Thus, in the case that the "actor/actress" is a user's most important attribute and s/he possesses high *Openness* value, the system will return movies with diverse actors/actresses to the user. In addition, if the user has low *Conscientiousness* value, the system will further increase all recommendations' overall diversity degree, since *Low Level* of *Conscientiousness* is correlated with high need for the overall diversity. The value of n (i.e., the number of diverse movies as defined above) is then accordingly adjusted in reference to Table 1 (which lists our proposed numbers in various conditions). Consequently, the m recommendations ($N - n = m$) that best satisfy the user's stated preferences on p attributes (that are with highest weights) are retrieved (p is set as 2 in our experiment). Regarding the detailed steps of computing diversity, due to the space limit, they can be referred to [3].

Table 1. Adjustment of *n* value for embodying diversity (when *N* = 10)

Overall Diversity	**Attribute's Diversity** (w.r.t. the most important attribute)	**n** (the number of diverse movies)
High need	High need	**7**
High need	Middle need	**6**
Middle need	High need	**6**
High need	Low need	**5**
Low need	High need	**5**
Middle need	Low need	**4**
Low need	Middle need	**4**
Low need	Low need	**3**

With a real user's data as the example, Table 2 lists her initially stated preferences on the movie's attributes, plus her personality values (note that the attributes in the first column are ordered by the user's weights on them). Figure 2 shows the list of 10 recommended movies presented to this user. Specifically, since the user's *Low Level* on *Conscientiousness* is significantly correlated to high need for both the diversity of "genre" (which is the user's most important attribute) and the overall diversity, n is set as 7 (see Table 1). These n movies are diverse in terms of all attributes, for which the weight of "genre" in the computation of diversity degree is higher than others. The other three movies (m = 3, which are displayed at the top of the list) include two that exactly match the user's preferences on both genre and director (i.e., {(genre, *suspense*), (director, *David Fincher*)}), and one movie that satisfies the genre preference.

Table 2. An example with a real user's preferences on movie attributes and her personality values

Attribute	**Value**	**Personality**	**Value**
Genre	*Suspense*	Openness	*High Level*
Director	*David Fincher*	Conscientiousness	*Low Level*
Actor	*Edison Chen*	Extraversion	*High Level*
Country	*America*	Agreeableness	*High Level*
Release Time	*Latest*	Neuroticism	*High Level*

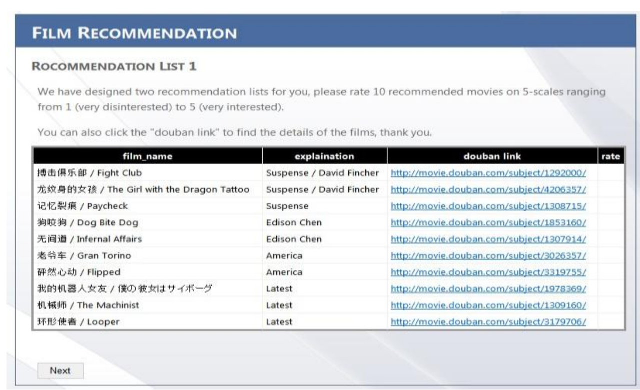

Figure 2. The A version of recommendation list for the user (described in Table 2).

4. EXPERIMENT

4.1 Controlled Experiment

In the experiment, for the comparison purpose, we implemented a variant (called version B) for which the impact of personality on the recommendations' diversity is reversed (that is, the personality is integrated into taking confounding, negative effect). With the same example given above, in the version B (see Figure 3), most of movies match the user's preference on "genre" (which however is diversified in the version A). Therefore, it can be seen that version A is targeted to show recommendations that positively incorporate the personality into adjusting the diversity, while B is mainly used to justify whether people would have negative opinions if the diversity does not match to their personality norms.

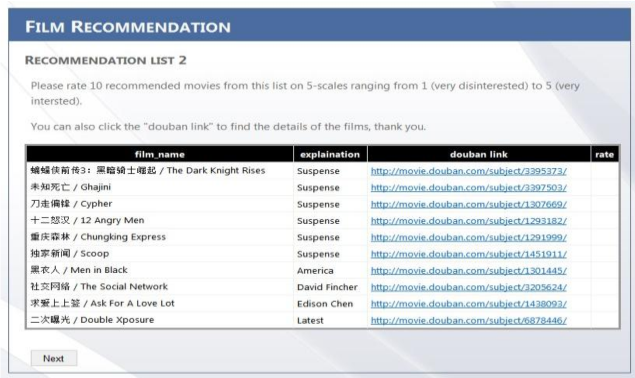

Figure 3. The B version of recommendation list for the user (described in Table 2).

4.2 Materials and Participants

The dataset used to implement our system contains 16,777 movies, which were crawled from a movie database system, Douban movie (http://movie.douban.com/). Each movie is associated with five major attributes: genre, director, country, release time, and actor/actress. More details, such as storyline, user ratings/ reviews, and Douban link, are also stored and available to the user.

The experiment was set up as a within-subject user study. Each participant was asked to rate the recommended movies in both versions (A & B). To minimize any carryover effects, the order of showing the two lists is randomly changed, that results in two user groups: the 1st group evaluated the version A first, then B; the 2nd group evaluated the version B at first. 52 participants (23 females) were voluntary to join this experiment. They are aged from 20-40 with different education levels (such as PhD, master, bachelor). Table 3 shows their demographic profiles.

The user's personality values were assessed with the big-five factor personality quiz [6]. Each factor is concretely measured via five sub-factor questions [6, 11][3], so the factor's score is the average of scores on these five questions. For instance, the questions used to measure *Openness to Experience* (O) include: *imagination* (rated from 1 "no-nonsense" to 5 "a dreamer"), *artistic interests* (1 "practical" to 5 "theoretical"), *liberalism* (1 "follow authority" to 5 "follow imagination"), *adventurousness* (1 "seek routine" to 5 "seek novelty"), and *intellect* (1 "prefer things clear-cut" to 5 "comfortable with ambiguity").

To start the evaluation, each user was required to first specify her /his preferences on the movie's attributes (including the weights and value preferences). Then, when either version was shown to the user, the user was asked to rate each recommended movie from 1 "very disinterested" to 5 "very interested". Afterwards, s/he filled in a post-task questionnaire, to express her/his overall opinions in term of the following three aspects (each was responded on a 5-point Likert scale from 1 "strongly disagree" to 5 "strongly agree"):

- Recommendation accuracy: *"the movies recommended for me matched my interests";*
- System competence: *"the website helped me to discover movies for myself";*
- Overall satisfaction: *"overall, I am satisfied with the recommended movies".*

Table 3. Demographic profiles of participants (the number of users is in the bracket)

Gender	Female (18); Male (34)
Age	<20 (1); 20-30 (42); 30-40 (8); >40 (1)
Education	Bachelor (12); Master (36); PhD (2); Others (2)
Job domain	Student (40); Enterprise (5); Institution (4); Others (3)
Frequency of watching movies (from 1 "never" to 5 "a few times per month")	Mean: 3.48 (st.d.: 1.16) Details: "Never" (1); "A few times totally" (13); "A few times one year" (10); "A few times every 3 months" (16); "A few times per month" (12)
Frequency of visiting movie sites (from 1 "never" to 5 "a few times per month")	Mean: 3.65 (st.d.: 1.03) Details: "Never" (0); "A few times totally" (8); "A few times one year" (15); "A few times every 3 months" (16); "A few times per month" (13)

4.3 Results Analysis

The comparison results are illustrated in Figure 4, from where we can see that version A obtains significantly higher scores in terms of all the three aspects. Specifically, most of users agreed that the recommendations they viewed in A match their interests (mean = 3.95, st.d. = 0.48; vs. mean = 3.55, st.d. = 0.59 in version B; t = 4.45, $p < 0.01$). They also perceived the system more competent in helping them to discover interesting movies (mean = 3.87, st.d. = 0.71, against mean = 3.15, st.d. = 1.02 in B; t = 1.81, $p < 0.01$). Overall, users were more satisfied with the version A (A: mean = 4.04, st.d. = 0.52; B: mean = 3.40, st.d. = 0.85; t = 5.01, p <0.01).

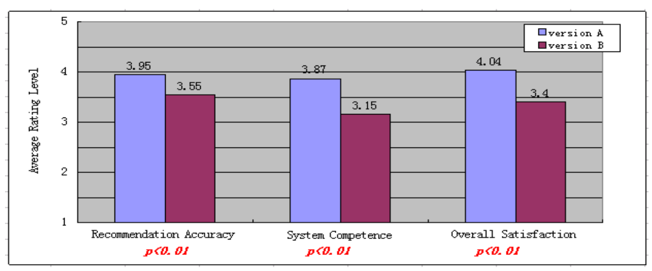

Figure 4. Comparison in respect of users' subjective perceptions (*p* was computed via Student t-Test).

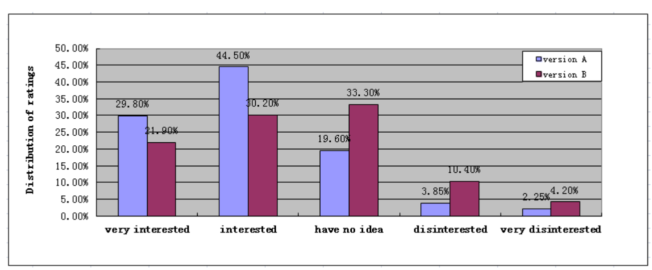

Figure 5. Distribution of users' ratings on movies.

[3] The Chinese translations are referred to [7].

The users' ratings on individual items further demonstrate the higher accuracy achieved by version A against B. Figure 5 shows the distribution of five rating scales among all movies. It can be seen that more movies in A were rated as "very interested" or "interested" (totally 74.3%) by users, while there are relatively less movies (52.1%) with such ratings in B. In fact, nearly half of the movies in B were rated below or equal to "no idea" (i.e., 47.9% movies, including 10.4% on the scale "disinterested" and 4.2% on "very disinterested").

4.4 Discussion

One limitation of our study is that the two compared systems are on extreme cases, without the comparison to a non-personality based system. Thus, in the future, we will conduct additional user evaluation in order to further verify the value of our method against system that does not fuse personality into adjusting diversity. On the other hand, in addition to personality, it will be interesting to know whether other factors such as demographic characteristics will take similar effect on users' diversity needs. According to our previous survey [3], some demographical properties (such as age, gender, and education level) are actually also significantly correlated with certain diversity variables. For example, people who are younger and/or with lower education level are more likely to prefer diverse movies. Our future studies will hence be targeted to take into account these properties for improving our diversity adjusting approach. Besides, because the current experiment mainly involved Chinese users, the results' applicability in other cultural contexts should be verified. We will also justify whether taking the personality quiz is acceptable by real system users, before they can get the recommendations.

5. CONCLUSION

In this paper, we first summarized our previous survey's findings that identified the significantly causal relation between personality and users' needs for recommendation diversity. Inspired by these findings, we have developed a recommender system that explicitly adopts personality for adjusting diversity degree within the set of *N* recommendations. We further conducted a user evaluation to compare our system to a variant that used personality in the contrary way. The experiment demonstrated that our method can significantly increase users' perceptions of system competence and recommendation accuracy (i.e., helping them to discover movies that match their interests). Users were also more satisfied with such personality-based recommendations. In terms of individual recommended movies, it was found higher ratings were assigned by users to ones that were recommended in our system. These results hence not only consolidate our previous survey's observations, but also suggest an effective approach to adapt the recommendations' diversity degree to individual users' personality values. Our work is thus helpful to reduce the gap between the two separate branches of research: diversity studies; and personality studies, in recommender systems. As discussed before, the future work will be along the direction to perform more experiments among universal user groups as well as in broader product domains.

6. ACKNOWLEDGMENT

We thank all the participants for their efforts and time. This work was supported by the Shanghai Science and Technology Commission Foundation (No. 11511502203 and No. 11511504002) and the National Natural Science Foundation of China (No. 61272365).

7. REFERENCES

[1] Adomavicius, G. and Kwon, Y. 2011. Maximizing aggregate recommendation diversity: a graph-theoretic approach. In *Proc of RecSys'11 Intl. Workshop on Novelty and Diversity in Recommender Systems (DiveRS'11)*, ACM, 3-10.

[2] Adomavicius, G and Tuzhilin, A. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowledge and Data Eng.*, 17, 6, 734-749.

[3] Chen, L., Wu, W., and He, L. 2013. How personality influences users' needs for recommendation diversity. In *Proc. CHI 2013 Extended Abstracts on Human Factors in Computing Systems*, ACM.

[4] Green, P.E., Krieger, A.M, and Wind, Y. 2001. Thirty years of conjoint analysis: reflections and prospects. *Interfaces* 31 (3), Part 2 of 2, May June 2001, S56-S73.

[5] Gunawardana, A. and Shani, G. 2010. Evaluating recommendation systems. In *Recommender Systems Handbook*. Ricci, Rokach, Shpira, Kantor (Eds.), 1-41.

[6] Hellriegel, D and Wlldman, R. 2000. *Organizational Behavior* (Eleventh Edition), Cincinnati, Ohio：South-Western College pub.

[7] Hellriegel, D and Wlldman, R. 2000. *Organizational Behavior* (Ninth Edition), Authorized Translation/Adaptation of the edition by Thomson Learning and East China Normal University Press.

[8] Hu, R. and Pu, P. 2009. Acceptance issues of personality-based recommender systems. In *Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys 2009)*, ACM, 221-224.

[9] Hu, R. and Pu, P. 2011. Enhancing collaborative filtering systems with personality information. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys 2011)*, ACM, 197-204.

[10] Hurley, N. and Zhang, M. 2011. Novelty and diversity in top-n recommendation – analysis and evaluation. *ACM Trans. Internet Technol*. 10 (4), Article 14, 1-30.

[11] John A. Johnson.2009.Descriptions used in ipip-neo narrative report.http://www.personal.psu.edu/faculty/j/5/j5j/IPIPNEOdescriptions.html, June 2009.

[12] McNee, S.M., Riedl, J., and Konstan, J.A. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *Proc. CHI 2006 Extended Abstracts on Human Factors in Computing Systems*, ACM, 1097-1101.

[13] Smyth, B. and McClave, P. 2001. Similarity vs. diversity. In *Proceedings of the 4th International Conference on Case-Based Reasoning (ICCBR 2001)*, D.W. Aha and I. Watson Eds., Lecture Notes in Computer Science, Springer, 347–361.

[14] Tkalcic, J., Kunaver, M., and Tasic, J. 2009. Personality based user similarity measure for a collaborative recommender system. In *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction Real World Challenges*, 30–37.

[15] Ziegler, C., Mcnee, S. M., Konstan J. A., and Lausen, G. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web (WWW 2005)*, ACM, 22–32.