The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation Lists

Bruce Ferwerda
Johannes Kepler University
Altenberger Str. 69
4040 Linz, AT
bruce.ferwerda@jku.at

Mark Graus
Eindhoven University of Technology
IPO 0.20, P.O. Box 513
5600 MB Eindhoven, NL
m.p.graus@tue.nl

Andreu Vall
Johannes Kepler University
Altenberger Str. 69
4040 Linz, AT
andreu.vall@jku.at

Marko Tkalcic
Free University of Bolzano
Piazza Domenicani, 3
39100 Bozen-Bolzano, IT
marko.tkalcic@unibz.it

Markus Schedl
Johannes Kepler University
Altenberger Str. 69
4040 Linz, AT
markus.schedl@jku.at

ABSTRACT

Diversifying recommendations has shown to be a good means to counteract on choice difficulties and overload, and is able to positively influence subjective evaluations, such as satisfaction and attractiveness. Personal characteristics (e.g., domain expertise, prior preference strength) have shown to influence the desired level of diversity in a recommendation list. However, only personal characteristics that are directly related to the domain have been investigated so far. In this work we take personality traits as a general user model and show that specific traits are related to a preference for different levels of diversity (in terms of recommendation satisfaction and attractiveness). Among 103 participants we show that conscientiousness is related to a preference for a higher degree of diversification, while agreeableness is related to a mid-level diversification of the recommendations. Our results have implications on how to personalize recommendation lists (i.e., the amount of diversity that should be provided) depending on users' personality.

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); User models; User studies;

Keywords

Diversity; Recommender Systems; User-Centric Evaluation; Personality

1. INTRODUCTION

Providing users with a diversified list of recommendations has shown to have positive effects on the user experience. With an abundance of choices available nowadays, providing diversity in the recommendations can counteract on the negative psychological effects that users may experience, such as choice overload and choice difficulties [26]. These negative effects are caused by recommender systems, which are originally designed to output recommendations that are closest to the user's interest. The closer to the user's interest, the higher the accuracy of the recommender system algorithm, but also results in recommendations that are often too similar to each other (e.g., same level of attractiveness to the user). This does not only increase the chance of choice overload and choice difficulties to the user, but also increases the possibility of not covering the full spectrum of the user's interest [3].

Although prior research has shown that recommendation diversity has positive effects on the user experience, differences between diversity needs of users have not been given a lot of attention. Domain expertise and prior choice preferences have shown to play a role in the amount of diversity desired by the user [2, 6, 26]. Others have shown that diversity needs can also be related to cultural dimensions [8, 14]. In this work we consider personality traits as an indicator of satisfaction and attractiveness on differently diversified music recommendation lists.

The use of personality as a general model for users has gained increased interest. Several works revealed personality-based relationships with users' behavior, preferences, and needs (e.g., [10, 15, 25]), how to implicitly acquire personality traits of users from social media trails (e.g., Facebook [1, 4, 12, 20], Twitter [16, 21], and Instagram [11, 13, 24]), and how personality traits can be implemented into a personalized system [7, 9]. With our work we contribute to the personality research by providing more insights into personality-related diversity needs. We found among 103 participants that the conscientiousness and agreeableness personality traits play a role in the desired amount of diversity in a recommendation list. While conscientious participants showed a higher degree of satisfaction and attractiveness with the more diversified recommendations, agreeable participants were more satisfied and found the list more attractive with medium amount of diversity in the recommendations.

EMPIRE 2016, September 16, 2016, Boston, MA, USA.

2. RELATED WORK

The positive effects of recommendation list diversity has been shown by several researchers. Bollen et al. [2] and Willemsen et al. [26] investigated the influence of diversity on movie recommendations and found that diversity has a positive effect on the attractiveness of the recommendation set, the difficulty to make a choice, and eventually on the choice satisfaction. Besides the positive effects of diversification, also personal characteristics play a role on the attractiveness of the diversified recommendation list (e.g., strength of prior preference or domain expertise [2, 23]). Bollen et al. [2] found that expertise in the domain showed a positive effect on the item attractiveness.

The personal characteristics that have been identified so far are domain specific to the kind of recommendations. However, a more general personal characteristic may be present that influences the subjective evaluations with the diversified recommendations. Personality has shown to be an enduring factor, which can relate to one's taste, preference, and interest (e.g., [5, 10, 25]). Chen et al. [5] and Wu et al. [27] showed relationships with personality and preference for diversification based on different movie characteristics (e.g., genre, artist, director). Ferwerda et al. [10] showed that music preferences can be related to the personality of the listener, whereas Tkalcic et al. [25] found relationships between personality traits and the preference of being exposed to certain amounts of multimedia meta-information.

In this work we investigate whether personality traits can be considered a personal characteristic that influences the subjective evaluations of diversified recommendation lists. To this end, we rely on the widely used five-factor model (FFM), which categorizes personality into five general dimensions: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [19].

3. DATA PREPARATION & PROCEDURES

We created differently diversified music recommendation lists in order to investigate the influence of personality traits on the subjective evaluation of the recommendation lists. Since we created the recommendation lists off-line, we separated the study in two parts. In the first part participants were recruited and their *complete* Last.fm listening history was crawled in order to create the recommendation lists. After the lists were created, participants from the first part were invited for the second part where they were asked to assess the diversified recommendation lists.

We recruited 254 participants through Amazon Mechanical Turk for the first part of the study. Participation was restricted to those located in the United States with a very good reputation ($\geq$95% HIT approval rate and $\geq$1000 HITs approved) and a Last.fm account with at least 25 listening events. Furthermore, they were asked to fill in the 44-item Big Five Inventory personality questionnaire [19] to measure the FFM. Control questions were asked to filter out fake and careless contributions. A compensation of $1 was provided. We crawled the complete listening history of each participant and aggregated the listening events to represent artist and playcount (i.e., number of times listened to an artist).

In order to prepare the music recommendation lists for each participant, we complemented our data with the LFM-1b dataset [22]. [1] This dataset consists of the complete listening histories of 120,322 Last.fm users from different countries. Since our participants were all located in the United States, we only used the United State users of the LFM-1b dataset to complement our dataset. This resulted in 10,255 additional users, which we also aggregated into artist and playcount for each user. The final dataset consists of user, artist, and artist playcount triplets with a total of 387,037 unique artists for the creation of the recommendation lists.

We used the weighted matrix factorization algorithm of [18] on our final dataset to calculate the recommended items. This algorithm is specifically designed to deal with datasets consisting of implicit feedback (e.g., artist playcounts). We optimized the factorization hyper-parameters by conducting grid-search and picking the setting that yielded the best 5-fold cross-validated mean percentile rank. Specifically, using 20 factors, confidence scaling factor α=40, regularization weight λ=1000 and 10 iterations of alternating least squares, we achieved the best 5-fold cross-validated mean percentile rank of 1.78%. [2] Afterwards we factorized the whole user-artist triplets using this set of hyper-parameters.

The recommended items were diversified as was done in [26] by using the method of [28]. By using the latent features as the basis of diversification instead of additional metadata like genre information (as is done in content-based recommender systems) guarantees that diversity is manipulated in line with user preferences. Previous research demonstrated that this way of diversifying recommendations is perceived accordingly by users [26].

A greedy selection to optimize the intra-list similarity [3] was run on the top 200 recommended artists (i.e., the 200 artists with highest predicted relevance) to maximize the distances between item vectors in the matrix factorization space. This algorithm starts with a recommendation set consisting of the artist with highest predicted relevance. In an iterative fashion items are added to the recommendation set until it contains 10 items.

In each step of the iteration, for each candidate item i the sum of all distances from its item vector to each item vector in the recommendation set is calculated: $c_i = \sum_{j=1}^{z} d(i,j)$, where z is the number of items in the recommendation set and $d(i,j)$ is the Euclidean distance between two item vectors i and j). All candidate items are ranked based on decreasing value of c_i (P_{c_i}) and on predicted relevance (P_{r_i}). A weighting factor β is introduced to balance the trade-off between predicted relevance and diversity. For each candidate item the combined rank is calculated following $w_i^* = \beta * P_{c_i} + (1-\beta) * P_{r_i}$. The item with the highest combined rank is added to the recommendation set and the next step is taken until 10 items are selected.

β was manipulated to achieve different levels of diversification. In the described implementation β=1 corresponds to maximum diversity, β=0 corresponds to maximum predicted relevance. We compared recommendation lists for different values of β in terms of the sum of distances between the latent features scores of items in the recommendation set and their average range. The list for β=0.4 showed to fall halfway between maximum relevance and maximum diversity. Thus, the final β levels for diversification were set at β=0 (low), β=0.4 (medium), and β=1 (high).

After the recommendation lists were created, emails were

[1] Available at http://www.cp.jku.at/datasets/LFM-1b/

[2] See [18] for details on the hyper-parameters and the definition of the mean percentile rank metric.

sent out to all participants to invite them for the second part of the study. We created a login screen so that we could retrieve the personalized recommendation lists for each participant. After the log in, the participant was sequentially presented with a recommendation list for three times, with each time a different level of diversity (i.e., low, medium, or high). The order of presentation was randomized. Each recommended artist was enriched with metadata from Last.fm (i.e., picture, genre, Top-10 songs with the number of listeners and playcounts), which was shown when hovered over the name in the list. Additionally, example songs were provided by clicking on the artist name (new browser screen linked to the artist's YouTube page). Participants were asked to answer questions about perceived diversity, recommendation satisfaction, and recommendation attractiveness [3] before moving on to the next list. These questions needed to be answered for each of the three lists.

After the participant assessed all three recommendation lists, we performed a manipulation check by placing the three lists next to each other (randomly ordered) and asked the participant to rank order the lists by diversity.

There were 103 participants who returned for the second part of the study. We included several control questions to filter out careless contributions, which left us with 100 participants for the analyses. Age: 18-65 (median 28), gender: 54 male, 46 female, and were compensated with $2.

4. RESULTS

4.1 Manipulation Check

A Wilcoxon signed-rank test was used to test the perceived diversity levels of the recommendation lists. Results show an increase of perceived diversity by comparing the low diversity (M=1.28) against the medium (M=2.05, r=.60, Z=10.370, p<.001) and high condition (M=2.65, r=.80, Z=13.784, p<.001). A significant diversity increase was also found between medium and high (r=.45, Z=7.711, p<.001).

4.2 Measures

Items in the questionnaire were assessed using a confirmatory factor analysis (CFA) with repeated ordinal dependent variables and a weighted least squares estimator to determine whether the questions convey the predicted constructs. After deleting questions with high cross-loadings and low commonalities, the model consisting of three constructs showed a good fit: $\chi^2(32)$=108.6, p<.001, CFI=.99, TLI=.98, $RMSEA$=.06. [4] The constructs with their items are shown below (5-point Likert scale; Disagree strongly-Agree strongly). The Cronbach's alpha (α) and the average variance extracted (AVE) of each construct showed good values (i.e., α>.8, AVE>.5), indicating convergent validity. Also, the square root of the AVE for each construct is higher than any of the factor loadings (FL) of the respective construct, which indicates good discriminant validity.

Perceived Diversity (AVE=.723, α=.887):
- The list of artists was varied. (FL=.858)
- Many of the artists in the lists differed from other artists in the list. (FL=.837)
- The artists differed a lot from each other on different aspects. (FL=.855)

Recommendation Satisfaction (AVE=.821, α=.932):
- I am satisfied with the list of recommended artists. (FL=.927)
- In most ways the recommended artists were close to ideal. (FL=.905)
- The list of artist recommendations meet my exact needs. (FL=.885)

Recommendation Attractiveness (AVE=.771, α=.931):
- I would give the recommended artists a high rating. (FL=.874)
- The list of artists showed too many bad items. (FL=-.830)
- The list of artists was attractive. (FL=.914)
- The list of recommendations matched my preferences. (FL=.893)

4.3 Analysis

We used a repeated measures ANOVA in order to investigate the influence of personality traits on the subjective evaluations of the diversified music recommendation lists. Below the results of personality traits on the different subjective evaluations are provided. The effects between diversity levels are all compared against the low diversity condition.

4.3.1 Personality on Perceived Diversity

Results show that Mauchly's test is not violated ($\chi^2(2)$= .115, p=.944), so sphericity can be assumed, and therefore, no correction is needed. The results show that there are no significant main effects of the different personality traits on perceived diversity. However, a general difference in perceived diversity can be assumed ($F(2, 22)$=51.029, p<.001). Exploring the differences between the levels of diversified recommendation lists show that there is an increase in perceived diversity when comparing the low diversified list against the medium ($F(1, 11)$=11.596, p<.001) and the high diversified lists ($F(1, 11)$=31.191, p< .001). This confirms once more that our diversification was effective and was perceived as such by the participants.

4.3.2 Personality on Recommendation Satisfaction

Mauchly's test shows that sphericity is not violated ($\chi^2(2)$= 1.830, p=.401), and therefore no correction is needed. Assessing the effect of the different personality traits on the recommendation satisfaction, the following personality traits show a main effect: conscientiousness ($F(4, 22)$=2.454, p<.05) and agreeableness ($F(4, 22)$=3.886, p<.05). Additional analyses by looking at the levels between the diversity levels (i.e., low, medium, and high diversification) show that conscientious participants are increasingly satisfied when provided a higher degree of diversity: medium diversity ($F(2, 11)$=3.994, p<.05) and high diversity ($F(2, 11)$=4.036, p<.05). However, the satisfaction differences for agreeable participants show a higher satisfaction for the medium diversification ($F(2, 11)$=9.660, p<.05) than for the high diversification ($F(2, 11)$=4.036, p<.05).

[3] Questions measuring perceived diversity and recommendation attractiveness were adapted from [26].

[4] Cutoff values for a good model fit are proposed to be: CFI>.96, TLI>.95, and $RSMEA$<.05 [17].

ACM Conference on Recommender systems*, pages 161–168. ACM, 2014.

4.3.3 Personality on Recommendation Attractiveness

Assessing Mauchly's test shows that there is no violating of sphericity ($\chi^2(2)$= 1.860 p=.395). Also here, results show main effects for the conscientiousness ($F(4, 22)$=3.157, p<.05) and agreeableness ($F(4, 22)$=3.469, p<.05) personality traits. By looking at the differences between the levels of diversification, we found similar patterns as with satisfaction. Results show that conscientious participants were increasingly more attracted to more diversified recommendation lists: medium ($F(2, 11)$=2.955, p<.05), high ($F(2, 11)$=7.866, p<.05). Participants scoring high on the agreeableness personality traits show to be more attracted to the medium ($F(2, 11)$=5.933, p<.05) diversified list than to the high ($F(2, 11)$=5.314, p<.05) diversified list.

5. CONCLUSION & DISCUSSION

Our results show that certain personality traits (i.e., conscientiousness and agreeableness) are related to the subjective evaluations of diversified recommendation lists. We found that conscientious people judged a higher degree of diversity more attractive and were more satisfied with it, whereas agreeable people showed to have more interest (i.e., list attractiveness and satisfaction) in a medium degree of diversity.

The relationships that we found can be used in personality-based systems as proposed in [7]. With the increased connectedness of applications, such as recommender systems, with social networking sites, users' personality can be acquired without the need of behavioral data in the application (e.g., via Facebook [1, 4, 12, 20], Twitter [16, 21], or Instagram [11, 13, 24]). By identifying relationships with users' personality traits, such as in this work, cross-domain inferences about users' preferences and needs can be made and implemented to provide a personalized experience to users.

6. ACKNOWLEDGMENTS

This research is supported by the Austrian Science Fund (FWF): P25655.

7. REFERENCES

[1] M. D. Back, J. M. Stopfer, S. Vazire, S. Gaddis, S. C. Schmukle, B. Egloff, and S. D. Gosling. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21:372–374, 2010.

[2] D. Bollen, B. P. Knijnenburg, M. C. Willemsen, and M. Graus. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on RecSys*, pages 63–70. ACM, 2010.

[3] P. Castells, N. J. Hurley, and S. Vargas. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*, pages 881–918. Springer, 2015.

[4] F. Celli, E. Bruni, and B. Lepri. Automatic personality and interaction style recognition from facebook profile pictures. In *Proceedings of the ACM MM*, 2014.

[5] L. Chen, W. Wu, and L. He. How personality influences users' needs for recommendation diversity? In *Proceeding of CHI'13 EA*. ACM, 2013.

[6] M. D. Ekstrand, F. M. Harper, M. C. Willemsen, and J. A. Konstan. User perception of differences in recommender algorithms. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 161–168. ACM, 2014.

[7] B. Ferwerda and M. Schedl. Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal. In *Proceedings of the 2nd Workshop on EMPIRE*, 2014.

[8] B. Ferwerda and M. Schedl. Investigating the relationship between diversity in music consumption behavior and cultural dimensions: A cross-country analysis. In *Proc. of the 1st Workshop on SOAP*, 2016.

[9] B. Ferwerda and M. Schedl. Personality-Based User Modeling for Music Recommender Systems. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2016)*, Riva del Garda, Italy, 2016.

[10] B. Ferwerda, M. Schedl, and M. Tkalcic. Personality & emotional states: Understanding users' music listening needs. *UMAP 2015 Extended Proceedings*.

[11] B. Ferwerda, M. Schedl, and M. Tkalčič. Predicting Personality Traits with Instagram Pictures. In *Proceedings of the 3rd Workshop on EMPIRE*, 2015.

[12] B. Ferwerda, M. Schedl, and M. Tkalčič. Personality Traits and the Relationship with (Non-)Disclosure Behavior on Facebook. In *Companion of the 25th International WWW Conference*, 2016.

[13] B. Ferwerda, M. Schedl, and M. Tkalčič. Using Instagram Picture Features to Predict Users' Personality. In *Proceedings of the 22nd International Conference on MMM*, Miami, USA, January 2016.

[14] B. Ferwerda, A. Vall, M. Tkalčič, and M. Schedl. Exploring Music Diversity Needs Across Countries. In *Proceedings of the 24th International Conference on UMAP*, Halifax, Canada, July 2016.

[15] B. Ferwerda, E. Yang, M. Schedl, and M. Tkalčič. Personality Traits Predict Music Taxonomy Preferences. In *ACM CHI '15 EA*, 2015.

[16] J. Golbeck, C. Robles, M. Edmondson, and K. Turner. Predicting Personality from Twitter. In *Proceedings of the 3rd International Conference on SocialCom*, 2011.

[17] L.-t. Hu and P. M. Bentler. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1):1–55, 1999.

[18] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *ICDM*, 2008.

[19] O. P. John, E. M. Donahue, and R. L. Kentle. The big five inventory: Versions 4a and 54, institute of personality and social research. *UC Berkeley*, 1991.

[20] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. Seligman. Automatic Personality Assessment Through Social Media Language. *Journal of Personality and Social Psychology*, 108, November 2014.

[21] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. Our twitter profiles, our selves: Predicting personality with twitter. In *Proceedings of the 3rd International Conference on SocialCom*, 2011.

[22] M. Schedl. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proceedings on ICMR*, 2016.

[23] B. Scheibehenne, R. Greifeneder, and P. M. Todd.

What moderates the too-much-choice effect? *Psychology & Marketing*, 26(3):229–253, 2009.

[24] M. Skowron, B. Ferwerda, M. Tkalčič, and M. Schedl. Fusing Social Media Cues: Personality Prediction from Twitter and Instagram. In *Companion Proceedings of the 25th International WWW Conference*, 2016.

[25] M. Tkalcic, B. Ferwerda, D. Hauger, and M. Schedl. Personality correlates for digital concert program notes. *UMAP 2015, Springer LNCS 9146*, 2015.

[26] M. C. Willemsen, B. P. Knijnenburg, M. P. Graus, L. C. Velter-Bremmers, and K. Fu. Using latent features diversification to reduce choice difficulty in recommendation lists. *RecSys*, 11:14–20, 2011.

[27] W. Wu, L. Chen, and L. He. Using personality to adjust diversity in recommender systems. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 225–229. ACM, 2013.

[28] C.-N. Ziegler, S. M. McNee, J. a. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. *WWW*, page 22, 2005.