EMPIRE 2013: Emotions and Personality in Personalized Services

Marko Tkalčič[1], Berardina De Carolis[2], Marco de Gemmis[2], Ante Odić[3], and Andrej Košir[3]

[1] Johannes Kepler University Department for Computational Perception, Linz, Austria
[2] University of Bari Aldo Moro, Italy
[3] University of Ljubljana Faculty of electrical engineering, Slovenia

Abstract. The EMPIRE workshop attempts to provide some answers to the growing interest of the user modeling research community on the role of human factors, especially personality and emotions, on various aspects of user modeling. This first edition of the workshop has six accepted papers and an invited talk.

Keywords: personality, emotions, user modeling, recommender systems, social signal processing

1 Introduction

The 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE 2013[4]) is taking place on 10. June 2013 in Rome at the Roma Tre University in conjunction with the 21st conference on User Modeling, Adaptation, and Personalization (UMAP 2013[5]).

While a lot of discussion has been made on filtering algorithms, and evaluation measures, few studies have stood to consider the role of emotions and personality in user models and personalized services. The workshop attempts to provide insight into these issues.

Characterizing the user model and the whole user experience with personalized service, by means of affective traits, is an important issue which merits attention from researchers and practitioners in both web technology and human factor fields.

Some questions motivate this workshop:

- Do affective traits (personality, emotions, and mood) influence and determine the acceptance of the personalized suggestions?
- How personality traits should be included in the user model?
- How the personalized services should be adapted to emotions and mood to increase user satisfaction?

[4] `http://empire2013.wordpress.com/`
[5] `http://www.umap2013.org`

2 Contributions

Personality is a recurrent theme among the accepted papers. It has been investigated in connection with users' preferences by Bologna et al. [1], Cantador et al. [2], Hu and Pu [3] and Odić et al. [5].

Bologna et al. [1] present the prototype of a recommender system for eCommerce,that exploits the users' personality in terms of their *vocational personality*, as expressed with the RIASEC model. Their system performs a classical context-aware ranking and then re-ranks the list of top-N items according to the users' personalities. The prototype is currently undergoing experimental validation.

In their work, Cantador et al. [2] present the outcomes of a study aimed at understanding the relationships between users' personalities and their preferences in different domains. Their study relies on the myPersonality dataset with over 3 million users. Of special interest is the result table with stereotypical user preferences.

A complementary view of the role of personality in users' ratings is presented by Hu and Pu [3]. The basis of their study is a dataset of a gifts retailer. The authors are interested in various aspects of a single user's rating behaviour and their relations with her/his personality type.

Odić et al. [5] present the results of a study that compares the ability of emotion induction (by movies as stimuli) in end users under different contextual situations and their personality types. The authors identify personality traits whose emotional responses are stable across different contextual values (alone vs. non-alone) and those who are not based on the COMODA dataset.

The work presented by Moore et al. [4] is focused on the validation of generally accepted representation of smileys as emotion indicators. They carried out a large survey with nearly 1000 participants. Based on their dataset, they are able to discern universal emoticons from ambiguous emoticons.

In their work, Vodlan et al. [6] present the experimental design for the evaluation of the impact of the social signal *hesitation* on users' decision making. More concretely, the authors use hesitation as an indicator of the user's preference for more diverse or less diverse items in the evaluated conversational recommender system for movies. The presented work is currently undergoing experimental validation.

3 Acknowledgement

The EMPIRE workshop chairs would like to thank all the authors for their submissions.

Furthermore, we would like to thank the UMAP workshop chairs, Shlomo Berkovsky (from NICTA, Australia) and Pasquale Lops (from the University of Bari Aldo Moro, Italy), for their guidance during the workshop organization.

Our gratitude goes also to our invited speaker, Neal Lathia (from the University of Cambridge), for sharing his insights on the recent developments in the field.

Last but not least, we want to thank the members of the programme committee who reviewed the submissions and helped to keep a high quality of the accepted papers.

3.1 Programme Committee

- Alessandro Vinciarelli, University of Glasgow
- Aleksander Valjamae, Graz University
- Elisabeth Andre, Augsburg University
- Floriana Grasso, University of Liverpool
- Francesco Ricci, Free University of Bozen-Bolzano
- Gustavo Gonzalez, MediaPro (Imagina Group)
- Ioannis Arapakis, Yahoo! Barcelona
- Judith Masthoff, University of Aberdeen
- Li Chen, Hong Kong Baptist University
- Man-Kwan Shan, National Chengchi University, Department of Computer Science
- Marius Kaminskas, Free University of Bozen-Bolzano
- Martijn Willemsen, Eindhoven University of Technology
- Markus Zanker, University of Klagenfurt
- Michal Kosinski, Microsoft
- Mohammad Soleymani, University of Geneva/Imperial college
- Neal Lathia, Cambridge University
- Rong Hu, Swiss Federal Institute of Technology in Lausanne (EPFL)

References

1. Ciro Bologna, Anna Chiara De Rosa, Alfonso De Vivo, Matteo Gaeta, Giuseppe Sansonetti and Valeria Viserta. *Personality-Based Recommendation in E-Commerce* In Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE 2013)
2. Iván Cantador, Ignacio Fernandez-Tobias and Alejandro Bellogin. *Relating Personality Types with User Preferences in Multiple Entertainment Domains* In Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE 2013)
3. Rong Hu and Pearl Pu. *Exploring Relations between Personality and User Rating Behaviors* In Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE 2013)
4. Adam Moore, Christina M. Steiner and Owen Conlan. *Design and development of an empirical smiley-based affective instrument* In Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE 2013)
5. Ante Odić, Marko Tkalčič, Jurij Franc Tasič and Andrej Košir. *Personality and Social Context: Impact on Emotion Induction from Movies* In Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE 2013)

6. Tomaž Vodlan, Marko Tkalčič and Andrej Košir. *The Role of Social Signals in Telecommunication: Experimental Design* In Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE 2013)

Building Systems to Capture, Measure, and Use Emotions and Personality

Neal Lathia

University of Cambridge
Computer Laboratory
neal.lathia@cl.cam.ac.uk

Traditionally, personalisation (e.g., that in recommender systems) has been viewed as a "black box," where machine-learning algorithms were designed and implemented to tailor content based solely on users' feedback data. Recently, a number of themes have emerged that show how researchers are unboxing this metaphor in order to build more accurate and engaging personalised systems. For example, researchers are revisiting what data can be used beyond preferences (i.e., context-awareness) and how to best measure the quality of recommendations (beyond accuracy-based metrics).

In this keynote, I aim to open a discussion about how these recent trends in personalised systems are, in fact, related to accommodating for "people" rather than "users," and how this may lead towards systems that solicit, use, and augment personalised experiences with representations of emotion and personality. In doing so, systems progress from representing 'user' data as a set of preferences towards capturing our states and traits.

Starting from an experiment I conducted that aimed to measure perceived quality of diverse recommendations [1], but also inadvertently angered some participants; I will briefly overview how emotions are starting to be used in this domain, and how they draw and build from the psychology literature. However, a number of research challenges emerge. These challenges encompass two key questions: how do we appropriately collect data about people's emotions? Moreover, how should this data be used?

Recently, we deployed a system [2] to measure people's emotions and learn how they relate to smartphone usage and sensor data. In a preliminary study [3], we found that the method we used to collect representations of emotions could influence what we inferred about people's emotional states. How can future systems avoid this bias?

In on going work, we are investigating how collaborative filtering (CF) may be augmented to use personality information. Much like context-aware CF, it is not immediately clear how to merge preference and personality or, indeed, whether doing so in any way will improve recommendations. I will discuss some progress, difficulties, and opportunities, and we can close by discussing how the research community can tackle them.

References

1. N. Lathia, S. Hailes, L. Capra, and X. Amatriain. Temporal Diversity in Recommender Systems In ACM SIGIR 2010 Annual Conference on Research Development on Information Retrieval. Geneva, Switzerland. July 19-23, 2010.
2. N. Lathia, K. Rachuri, C. Mascolo, and P. Rentfrow. Emotion Sense Android Application. Available at: http://emotionsense.org/
3. N. Lathia, K. Rachuri, C. Mascolo, and P. Rentfrow. Contextual Dissonance: Design Bias in Sensor-Based Experience Sampling Methods. In ACM International Joint Conference on Pervasive and Ubiquitous Computing. Zurich, Switzerland. September 8-12, 2013.

Personality-Based Recommendation in E-Commerce

Ciro Bologna[1], Anna Chiara De Rosa[2], Alfonso De Vivo[3], Matteo Gaeta[4], Giuseppe Sansonetti[4], and Valeria Viserta[2]

[1] Qui! Group S.p.A.
[2] Poste Italiane
[3] Consorzio CRMPA (Centro di Ricerca in Matematica Pura e Applicata) c/o Universitá degli Studi di Salerno
[4] Dipartimento di Ingegneria dell'Informazione, Ingegneria Elettrica e Matematica Applicata, Universitá degli Studi di Salerno

Abstract. In recent years there has been an exponential increase in the number of users each day adopting e-commerce as a purchasing vehicle of products and services. This has led to a growing interest from the scientific community in approaches and models that would improve the customer experience. Specifically, it has been repeatedly pointed out that the definition of a customer experience tailored to the user personality traits would likely increase the probability of purchase. In this article we illustrate a recommender system for e-commerce capable of adapting the product and service offer according to not only the user interests and preferences, and his context of use, but also his personality profile derived from information relating to his professional activities.

Keywords: Personality-based, user model, context-awareness, recommender system, e-commerce

1 Introduction

In literature there are several works that describe how the definition of a customer experience taking into account the user personality ensures the increase of the purchase probability. The study described in [11] shows the correlations among the reasons that lead a user to buy and the way in which he makes the purchase. In particular, it describes how a user driven by utilitarian reasons prefers making a "goal-oriented" research, since he has already a purchase plan, and the search has the only aim to obtain information about the product to be purchased, its cost, convenience, and availability. Another, different situation, is when the reasons that lead the user are of a hedonistic nature [1]. In such a case the user usually adopts an "exploration-oriented" search, in which he has no purchase plan yet, but he makes it by browsing and exploring different solutions. To et al. in [12] point out how the motivations that drive a user to purchase are related to personal traits distinguishing him; specifically, this study suggests a possible relationship among a theoretical personality model, such as BIG FIVE [6], and hedonistic and utilitarian motivations. Such models require the extraction of information needed for defining the user personality. For example, in literature there are several questionnaires [3], whose compilation allows us to extrapolate the user BIG FIVE profile. Unfortunately, this approach is not applicable to the e-commerce context, since the length of these questionnaires is not negligible and from the consumer point of view its purpose is not of immediate identification, so both the number of users, and the attention devoted to their compilation, will be reduced. To this end, it is appropriate to investigate different approaches to the identification of the user personality traits. The analysis of the literature has revealed the possibility of identifying the user personality from information on his profession. In particular, the theoretical personality model RIASEC [4] can be used for this purpose. Its name is an acronym of the six following personality traits: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. Hence, it is possible to associate any single profession with some personality traits of the RIASEC model. For example, a person practicing management accounting is associated with an IEC (i.e., Investigative, Enterprising, and Conventional) personality profile, which corresponds to a mainly investigative person having a good aptitude for business and repetitive activities. In this scenario, the user explicitly declares his profession, from which his personality traits are derived according to the RIASEC model.

In this article we propose a context-aware recommender system that suggests products and services in the e-commerce domain. During the recommendation process, our approach is capable of taking into account not only the user interests and preferences, but also his personality profile. For this purpose, the system makes use of a neural network whose input is the user personality profile according to the RIASEC model and output are the weights to be used in the combination of the results coming from the different modules of the system. The main goal of this process is to adapt the type of research of products and services available within the e-commerce platform to the user personality profile and, hence, to the motivations that lead him to the purchase.

2 Related Work

Recently, studies have indicated that there is a significant connection between personality and people tastes and interests [5]. Studies also show that personalities influence human decision making process and interests [9]. By drawing on the inherent inter-related patterns among users personalities and their interests/behaviors, personality-based recommenders are designed to provide personalized services. Several studies deal with the correspondence among characteristics of the personality and purchase intentions of an individual. Holland's theory (RIASEC), unlike the others, describes the strong connection between the environment and the individual personality: the latter is manifested through preferences for professional occupations and, at the same time, work environments are shaped by people working in them and what they do [4]. The user decisions are influenced emotionally, at least partially, by which content to choose, because while using applications with recommender systems he is constantly receiving various stimuli (e.g. visual, auditory) that induce emotional states. Thus it is important for the recommender system application to detect and make good use of emotional information. During the user interaction with a recommender system and the content consumption, emotions play different roles in different stages of the process. In [10] the authors subdivide the user interaction process in three stages, based on the role that emotions play: (i) the entry stage, (ii) the consumption stage, and (iii) the exit stage. Nunes and Hu [7] propose a personality-based recommender system to provide a better personalized environment for the customer. They claim that one interesting outcome of introducing a psychological dimension into the recommender system could be the possibility of products categorization based not only on their attributes (price, physical parameters, etc.), but also on the effect they may have on the consumer. Affective content profiling is still an open question, especially profiling content items that last longer than a single emotional response. Other studies put the attention on the context-aware recommender systems which help users and their desired content in a reasonable time, by exploiting the pieces of information that describe the situation in which users will consume the items [8].

3 The Proposed Approach

The proposed user model is based on the Vector Space Model technique to represent information about users and resources, namely, products and services. With this approach we define a *Concept Space* that models the knowledge base of interest with a conceptual subdivision (ontological) in R^d, with d number of ontological classes. Within this space, users and resources are represented by a *Concept Vector*, a weighted vector structure whose weights are, for resources, the level of consistency with concepts representing space and, for users, the levels of interest in the specific concepts of the knowledge domain. In the first case, a domain expert builds the vector that models the service, in the second case, the Concept Vector describes the user profile ($V_U \in R^d$) constructed and updated in function of the information related to his actions to represent the real and current interests of the consumer. Such information may be collected in explicit form, for example by completing a questionnaire, or in implicit form, through the use of implicit feedback techniques that, from user activities (e.g., purchases, queries, clicks) are able to extract information related to his interests and habits. The user action modeling and consequent user profile update occur in two distinct phases. The first, named *Concept Extraction*, builds a Concept Vector for every single user action A made by the user U ($V_{U,A} \in R^d$). The user profile update in function of his actions can be done after modeling user behavior. This phase, named *Concept Aggregation*, takes advantage of the Rocchio's algorithm to combine two vector structures ($V_{U,A}, V_U \in R^d$), thus obtaining the updated user profile ($V_U \in R^d$):

$$V_U = \alpha V_U + \beta \frac{1}{|V_{U,A}|} \sum V_{U,A}$$

The coefficients α and β represent the weights associated to the vectors $V_{U,A}$ and V_U, and they can be experimentally obtained through a preliminary testing on a small number of users. It is reasonable to expect that weights associated with explicit feedbacks are higher than those associated with implicit feedbacks. Indeed, in the latter there is likely to be a noise component due to the potential misinterpretation of the user actions. Modeling the user profile and electronic services through vector structures allows us to define a third phase, called *Concept Matching*, to propose the services of potential interest for the consumer. The Concept Matching is based on the comparison among the user profile and available resources modeled in the knowledge base. In particular, it is possible to filter the services obtained through the search activities by comparing them with the user profile characteristics. In this case, it is possible to compute the scoring value that represents the affinity level between each service and the user profile by means of the cosine similarity rule:

$$\text{Scoring} \ = \frac{V_U \times V_O}{\|V_U\|\|V_O\|}$$

Such value allows us to re-rank or, alternatively, filter, the search results. Thus it is possible to suggest personalized services or products of potential interest for a specific consumer.

3.1 Contextual Model

Current mobile devices enable users to interact with e-commerce services in multiple contexts (e.g., while traveling or in a shopping mall). Therefore, it is interesting and effective to monitor the contextual dimension that occurs during the interaction with the system in order to determine any correlations that may be useful during the product and service recommendation. In order to associate the current context with the content available on the e-commerce platform, a subset of features that can be measured on current mobile devices (e.g., smartphones, tablets) have to be initially identified. With a view to obtain a matching among the current context and the elements of the e-commerce service that may be of interest to the user, a domain expert has to identify all the features of a specific item that can be influenced by this context. The expert encodes such information in vector terms, where each dimension can take a value in a real interval (e.g., relative distance), or in a finite set of elements (e.g., "purchasable during the summer" in $\{\text{true, false}\}$). The V_C contextual vector is thus compared with the vector representing the domain elements identified by the expert. The matching process consists in the following steps:

1. the system identifies a first list of N results through the Concept Matching procedure in combination with a metric based on the user location;
2. each one of the N elements retrieved in the previous step is associated with the V_O vector and the V_C contextual vector is combined with each of them;
3. by means of a decision analysis algorithm based on Decision Trees, a value is computed for each of N elements, which expresses the relevance of the element as to the current context;

4. the list of N elements retrieved at the first step is re-ranked based on the relevance value of each element.

3.2 Search Results Combination

Now we can consider three lists of results: (1) the first list is obtained through a traditional matching process among the user query and available resources indexed by name and description. Such a matching process relies on traditional Information Retrieval techniques. The output corresponds to a R subset of resources; (2) the second list can be obtained through the Concept Matching process described above, that is based on the comparison among the user profile and available resources; (3) the third list is given as output of the contextual module. Subsequently, a combination of the three results lists can be performed. In other terms, the score associated with each element belonging to the R subset and obtained through the filtering is combined with the scores related to the matching with the user profile and the contextualization module. The three weights to be associated with the scores can be obtained as output of a neural network (see Figure 1), whose input is the user personality profile according to the RIASEC model. Specifically, a feed forward multi-layer perceptron can be employed for this purpose. The training data can be extracted from implicit and explicit feedbacks provided by users. A recent experimental prototype employed in artificial and real settings, and based on these technologies, has proven effective in the particular context of the recommendation of points of interest (e.g., restaurants) [2].

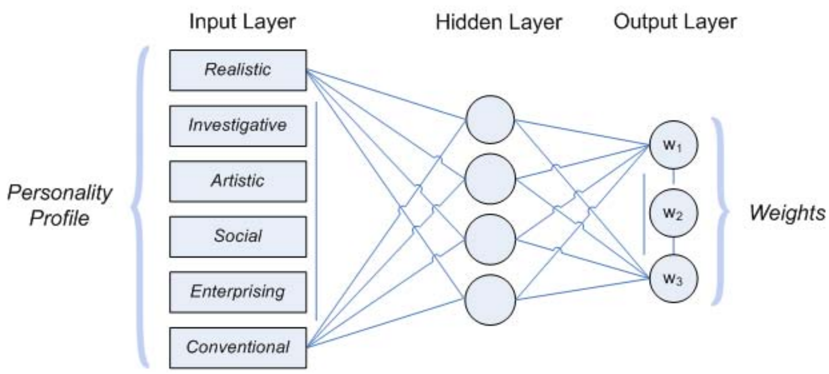


Fig. 1. Neural network that takes as input the user personality profile according to the RIASEC model, and returns as output the weights to be used in the combination of results from the different modules of the system.

4 Conclusion

In this article we have described a context-aware recommender system in the e-commerce domain capable of adapting the suggestion of products and services, not only to the user interests and preferences, but also to his personality. For this purpose, the system makes use of a neural network which takes as input the user personality profile according to the model RIASEC, and returns as output the value of the weights to be employed in the combination of the results coming from the different modules of the system. A prototype of the proposed model has been realized and is currently undergoing experimental validation. The first results we have obtained are encouraging and tend to confirm the validity and soundness of the advanced approach.

Acknowledgments. This research is partially supported by the Italian Ministry of University and Research under the Project TITAN "System for e-Money and Multi-Channel Value Added Service", PON RC, Number PON01 02136.

References

1. Arnold, M.J., Reynolds, K.E.: Hedonic shopping motivations. Journal of Retailing 79(2), 77–95 (2003)
2. Biancalana, C., Flamini, A., Gasparetti, F., Micarelli, A., Millevolte, S., Sansonetti, G.: Enhancing traditional local search recommendations with context-awareness. In: Proc. of UMAP'11. pp. 335–340 (2011)
3. Caprara, G.V., Barbaranelli, C., Borgogni, L., Perugini, M.: The "big five questionnaire": A new questionnaire to assess the five factor model. Personality and Individual Differences 15(3), 281–288 (1993)
4. Holland, J.L.: Making vocational choices: A theory of vocational personalities and work environments, vol. 3. Psychological Assessment Resources (1997)
5. Hu, R., Pu, P.: A study on user perception of personality-based recommender systems. In: Proc. of UMAP'10. pp. 291–302 (2010)
6. McCrae, R.R., Costa, P.T.: Validation of the five-factor model of personality across instruments and observers. J. of Pers. and Soc. Psych. 52(1), 81–90 (1987)
7. Nunes, M.A.S., Hu, R.: Personality-based recommender systems: an overview. In: Proc. of the 6th ACM Conf. on Rec. Sys. pp. 5–6. RecSys '12, ACM, New York, NY, USA (2012)
8. Odic, A., Tkalcic, M., Tasic, J., Kosir, A.: Relevant context in a movie recommender system: Users' opinion vs. statistical detection. In: Proc. of the 4th Workshop on Context-Aware Recommender Systems (2011)
9. Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlates of music preferences. J. of Pers. and Soc. Psych. 84(6), 1236–1256 (Jun 2003)
10. Tkalcic, M., Kosir, A., Tasic, J.: Affective recommender systems: the role of emotions in recommender systems. In: Proc. of the 5th ACM Conf. on Rec. Sys. (2011)
11. To, P.L., Liao, C., Lin, T.H.: Shopping motivations on internet: A study based on utilitarian and hedonic value. Technovation 27(12), 774–787 (2007)
12. Tsao, W.c., Chang, H.r.: Exploring the impact of personality traits on online shopping behavior. Journal of Business Management 4(9), 1800–1812 (2010)

Relating Personality Types with User Preferences in Multiple Entertainment Domains

Iván Cantador[1], Ignacio Fernández-Tobías[1], Alejandro Bellogín[1,2]

[1] Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain
[2] Centrum Wiskunde & Informatica, Amsterdam, Netherlands
{ivan.cantador, i.fernandez, alejandro.bellogin}@uam.es

Abstract. We present a preliminary study on the relations between personality types and user preferences in multiple entertainment domains, namely movies, TV shows, music, and books. We analyze a total of 53,226 Facebook user profiles composed of both personality scores (openness, conscientiousness, extraversion, agreeableness, neuroticism) from the Five Factor model, and explicit interests about 16 genres in each of the above domains. As a result of our analysis, we extract personality-based user stereotypes and association rules for some of the considered domain genres, and infer similarities of personality types related to genres in different domains.

Keywords: types of personality, user profiles, personalization, cross domains.

1 Introduction

The majority of services for personalized information access, retrieval, and filtering deals with the exploitation of user preferences – tastes, interests, goals – obtained explicitly (e.g. by means of ratings) or implicitly (e.g. by mining click-through and log data). In addition to these preferences, contextual signals – such as the current time, and the user's location and social companion – are also taken into consideration [1].

Many effective personalization approaches have been proposed in a large number of applications [3]. However, little work has been done on the characterization of user models in personalized services with regard to *affective traits*, such as moods, emotions, and types of personality. Emotions are intense feelings that are directed at someone or something [9]. Moods, in contrast, are feelings that tend to be less intense than emotions, and often – though not always – lack a contextual stimulus [8]. Personality, on the other hand, can be defined as a combination of characteristics or qualities that form an individual's style of thinking, feeling and behaving in different situations [21].

Personality influences how people make their decisions [17]. In fact, it has been proved that people with similar personality characteristics are likely to have similar preferences. In [19] Rentfrow and Gosling show that "reflective" people with *openness to experiences* usually have preferences for jazz, blues and classical music, and "energetic" people with high degree of *extraversion and agreeableness* usually appreciate rap, hip-hop, funk and electronic music. In [5] Chausson presents a preliminary study showing that people *open to experiences* are likely to prefer comedy and fantasy movies, *conscientious* individuals are more inclined to enjoy the action movie genre, and *neurotic* people tend to like romantic movies. Traits and types of personality could thus help explain why people prefer one option to other, and could be used to improve personalization services and enhance user experience [12][16].

In psychology the Five Factor model [6] – also known as *the Big Five* model – is a widely accepted theory that establishes five factors to describe the human personality: *openness* to experiences, *conscientiousness*, *extraversion*, *agreeableness* and *neuroticism*. Similarly to user preferences, the personality factors can be inferred explicitly, e.g. by means of personality questionnaires [6], or implicitly, e.g. by analyzing linguistic features of user texts [20], and by correlating personality traits with patterns of social network use – such as posting, rating, establishing friendship relations, and participating in user groups [2].

Once extracted, personality factors can be used to build personality-based user profiles that may be exploited by personalized information retrieval and filtering approaches. In [22] Tkalčič et al. apply and evaluate three user similarity metrics for the heuristic-based collaborative filtering strategy: a typical rating-based similarity, a similarity based on the Euclidean distance with the Big Five data, and a similarity based on a weighted Euclidean distance with the Big Five data. The reported results show that approaches using the Big Five data perform statistically equivalent or better than the rating-based approach. Following the findings of Rentfrow and Gosling [19], in [13] Hu and Pu present a collaborative filtering approach based on the correlations between personality types and music preferences, in which the similarity between two users is estimated by means of the Pearson's correlation coefficient on the users' Big Five personality scores. Combining this approach with a rating-based collaborative filtering, the authors show significant improvements over the baseline of considering only ratings data. Finally, in [20] Roshchina presents an approach that extracts Big Five factor-based profiles by analyzing hotel reviews written by users, and incorporates these profiles into a nearest neighbor algorithm to enhance personalized recommendations.

Before developing personality-based recommender systems, we aim to **establish and understand relations existing between personality types and user preferences in multiple entertainment domains**. Rentfrow's work [19] only addresses the music domain, and Chausson's analysis [5] considers a very limited number of movies. In this paper we present a preliminary, but extensive study on the relations between personality types and user preferences in multiple domains, namely movies, TV shows, music, and books. Specifically, we analyze a total of 53,226 Facebook[1] user profiles composed of both Big Five personality scores and explicit interests about 16 genres in each of the above domains. As a result of our analysis, we extract personality-based user stereotypes and association rules for some of the considered domain genres, and infer similarities of personality types related to genres in different domains.

[1] Facebook social network, http://www.facebook.com

The reminder of the paper is structured as follows. In Section 2 we detail the Five Factor model. In Section 3 we describe the analyzed personality- and preference-based Facebook user profiles. In Sections 4 and 5 we present our analysis on the relations between personality types and user preferences on single and crossed domains respectively. Finally, in Section 6 we conclude with some lines of future work.

2 Personality Traits: The Five Factor Model

In psychology identifying the structure and types of human personality has been a fundamental goal. Researchers have extensively studied known personality traits, and have analyzed a large variety of measures of such traits – on self-report and questionnaire data, and objective measures from experimental settings – in order to find underlying personality factors.

The Five Factor model [6] is a theory that establishes five broad domains or dimensions – called factors – to describe human personality. These factors are commonly known as the *Big Five* personality traits, and can be defined as follows:

- **Openness (OPE)**: from *cautious*/*consistent* to *curious*/*inventive*. This factor reflects a person's tendency to intellectual curiosity, creativity and preference for novelty and variety of experiences. A high score of openness entails strong degrees of imagination, artistic interest, emotionality, adventurousness, intellect and liberalism.
- **Conscientiousness (COS)**: from *careless*/*easy-going* to *organized*/*efficient*. This factor reflects a person's tendency to show self-discipline and aim for personal achievements, and to have an organized (not spontaneous) and dependable behavior. A high score of conscientiousness entails strong degrees of self-efficacy, orderliness, dutifulness, achievement-striving and cautiousness.
- **Extraversion (EXT)**: from *solitary*/*reserved* to *outgoing*/*energetic*. This factor reflects a person's tendency to seek stimulation in the company of others – showing sociability, talkativeness and assertiveness traits –, and to put energy in finding positive emotions, such as happiness, satisfaction and excitation. A high score of extraversion entails strong degrees of friendliness, gregariousness, activity level, excitement-seeking and cheerfulness.
- **Agreeableness (AGR)**: from *cold*/*unkind* to *friendly*/*compassionate*. This factor reflects a person's tendency to be kind, concerned, truthful and cooperative towards others. A high score of agreeableness entails strong degrees of morality, altruism, sympathy, modesty, trust, cooperation and conciliation.
- **Neuroticism (NEU)**: from *secure/calm* to *unconfident*/*nervous*. This factor reflects a person's tendency to experience unpleasant emotions, such as anger, anxiety, depression and vulnerability, and refers to the degree of emotional stability and impulse control. A high score of neuroticism entails strong degrees of hostility, social anxiety, depression, immoderation, vulnerability and impulsivity.

It is important to note that although these factors are statistical aggregates, exceptions may exist on individual personality profiles. In general, people who register high in *openness* are willing to new experiences, receptive to emotions, intellectually curious, and interested in art. A particular individual, however, may have a high overall *openness* score, and may be interested in learning and exploring new cultures, but may have no great interest in art.

Nonetheless, the Big Five factors have been shown to encompass most known personality traits, and are assumed to represent the basic structure behind all personality traits [18]. They thus provide a rich conceptual framework for integrating all the research findings and theories in personality psychology.

Moreover, the Five Factor model is a comprehensive, empirical, data-driven research finding investigated, discovered and defined by distinct groups of researchers. Tupes and Christal presented an initial model of personality factors [23]. Digman [7] proposed a five factor model of personality, which was extended with a high level of organization by Goldberg [10]. Independently, Cattell et al. [4] and Costa and McCrae [15] used different methods with which the five personality factors were found. Hence, each set of five factors found has had different names and definitions. All of them, however, have been proved to be highly inter-correlated and factor-analytically aligned [11].

The measurement of the Big Five factors comprises items that are self-descriptive sentences or adjectives, commonly presented in the form of short tests. In this context, the International Personality Item Pool[2] (IPIP) is a publicly available collection of items for use in psychometric tests, and the 20-100 item IPIP proxy for Costa and McCrae's NEO-PI-R test [6] is one of the most popular and widely accepted questionnaires to measure the Big Five factors in adult (≥18 years old) men and women without overt psychopathology.

3 Personality- and Preference-based User Profiles

myPersonality [14] is a Facebook application with which users take real psychometric tests. As of May 2013, the tool has let record a database that contains more than 6 million test results and more than 4 million individual Facebook profiles with a variety of personal user information, such a demographic and geo-location data, *likes*, status updates, and friendship relations, among others.

The members of myPersonality project have made publicly available[3] part of such dataset. Specifically, the public dataset contains the Big Five personality scores of 3.1 million users, collected using 20 to 336 item IPIP proxy for Costa and McCrae's questionnaires, and from which around 40% of the users took the 100 item version. The users either decided the length of questionnaire they wanted to take in advance, or took extra questions in blocks of 10 until finishing all 100 items. The users filled the questionnaires to get feedback, so they were quite well motivated, which resulted in high accuracy (reliability > 0.8, better than in most supervised pen-and-paper applications of the same measure). The public dataset also contains 46 million *likes* of 220,000 users for 5.5 million items of diverse nature: people (celebrities, politicians,

[2] International Personality Item Pool (IPIP), http://ipip.ori.org
[3] myPersonality project, http://mypersonality.org

directors, actors, musicians, writers, sportsmen, etc.), objects (movies, TV shows, songs, books, games, etc.), organizations, events, etc.

Due to the size and complexity of the dataset, in this paper we restrict our analysis to a subset of the dataset's items. Specifically, we selected all *like records* associated to items belonging to one of the following 4 categories: `Movie genre`, `TV genre`, `Musical genre`, and `Book genre`. Thus, for instance, selected items belonging to the `Movie genre` category are movie genres, such as *comedy*, *action*, *adventure*, *drama*, and *science fiction*. Note that we do not take into consideration a large number of potential valuable items, such as particular *movies* preferred by users. In our study, we discarded such items because no domain categories were available for them in the dataset, and acquiring their categories requires complex data acquisition and processing methods: 1) the items have to be found in external knowledge sources – like Wikipedia – with category information, and 2) the items in the dataset are identified by plain text names that do not follow a well-established naming convention: a certain concept (person, object, etc.) may have different names.

Next, we selected those users of the dataset that had *likes* for the considered items. Once the items and users were selected, we conducted text processing operations to consolidate morphological derivations of certain item names (e.g. *science fiction*, *science-fiction*, *sci-fi*, *sf*). Finally, we chose the items (i.e., domain genres) with the highest number of user *likes*. Specifically, we chose the top 16 genres of each of the 4 considered domains: movies, TV shows, music and books. In the end, the dataset used in our study contained **53,226 users** (60.37% female, 39.63% male) and their corresponding Big Five personality scores, 16 x 4 = **64 items**, and **58,576 *like* records** (12,420 in the movie domain, 3,705 in the TV show domain, 32,784 in the music domain, and 9,667 in the book domain).

4 Relations between Personality Types and User Preferences in Individual Domains

We have analyzed the dataset presented in Section 3 aiming to find meaningful relations between personality types and user preferences in each of the 4 considered entertainment domains: movies, TV shows, music and books. In the following we present preliminary results of our study: personality-based *user stereotypes* (Subsection 4.1) and *association rules* (Subsection 4.2) for some of the considered domain genres.

4.1 Domain-specific Personality-based User Stereotypes

Table 1 shows personality-based user stereotypes for the 16 genres selected in each domain, distinguishing female and male users. These stereotypes are vectors of 5 real values in the [1, 5] range that correspond to the average scores of the Big Five personality factors of the dataset users who had *likes* for the corresponding genres.

The colors of the cells are used to facilitate the analysis of differences in the factor scores for distinct stereotypes. In the table the colors are assigned for each column as follows. The highest scores of the column are marked in green, and the lowest ones are marked in red. The color intensities indicate how high/low the scores are. In this section we will refer as “high” scores to those colored with darkest green, and as “low” scores to those colored with darkest red. For the sake of simplicity and due to the preliminary nature of this study, we do not provide results on the statistically significance of all score differences.

It is important to note that for some genres, the number of users whose profiles were used to build the stereotypes is quite small; e.g. in the movie domain, the personality stereotype associated to the *tragedy* genre for male users was built with only 8 profiles. This is taken into consideration in the discussions and conclusions we provide in the reminder of the paper, by discarding the analysis of personality-based stereotypes built with a few (less than 40) users. Moreover, in each domain, the numbers of user profiles are quite different among the genre stereotypes. We could perform the analysis with equally sized groups of profiles. However, we decided to use all the available profiles in order to not lose information. In the following we discuss obtained relations between user personality types and preferences.

In the **movie domain**, users with high degree of *openness* (OPE) tend to like tragedy, neo-noir, independent, cult, and foreign movies, whereas a low degree of this factor corresponds with user preferences for war, romance, action, and comedy movies. For the *conscientiousness* (CON) factor, high scores correspond to independent, adventure, and science fiction movies, whereas low values correspond to cult, animation, and cartoon movies. For the *extraversion* (EXT) factor, drama, romance, comedy, and action movies are linked to high scores, whereas animation, tragedy, neo-noir, and science fiction are associated with lower scores. Adventure, romance, comedy, and drama movies tend to be liked by people with high degree of *agreeableness* (AGR), whereas a low degree of this factor is associated with parody, animation, neo-noir, cult, and horror movies. Finally, users with high degrees of *neuroticism* (NEU) prefer cult, tragedy, and animation movies, while users with low values tend to like adventure, independent, and war movies.

In the **TV show domain**, high scores of OPE implies user preferences for surviving shows, documentaries, and standup comedies, while low score imply user preferences for soap operas, and game and sports shows. Moreover, high values of CON are associated to surviving, talk, and sports shows, whereas low degrees correspond to cartoon and animation genres. People with low degrees of EXT tend to prefer animation and cartoon shows, and people with high values of this factor tend to like sports and reality shows. Game and talk shows are liked by users with high scores of AGR, while surviving and prank shows are generally liked by users with low scores of that factor. Finally, high values of NEU are associated to preferences for cartoon, music videos and soap opera, whereas people with low values prefer sports, prank, and surviving shows.

In the **music domain**, users with high degree of OPE tend to prefer blues, classical and indie music, whereas low values of this factor correspond to user preferences for r&b, rap, pop, and hip hop. For the CON factor, high values correspond to country, jazz, salsa, and r&b, while low values correspond to indie, metal, techno, and rap music. Users with high EXT scores prefer salsa music, but also tend to like hip hop and rap; low scores of that factor refer to user preferences for metal, techno and rock music. With respect to AGR, people with high degree of this factor tend to like country, oldies, dance, and jazz music, while people with lower values prefer metal, rap, and indie genres. Finally, indie, metal, and rock music genres are also usually preferred by users with high degree of NEU, and salsa, jazz, and hip hop are preferred by those users with low values of that factor.

Finally, in the **book domain**, people with a high degree of OPE tend to like poetry and science fiction, whereas those with low degree prefer drama, scary and crime books. Regarding the CON factor, high values are related to educational books, and low values to like comic, fantasy, and poetry genres. Users with a low degree of EXT tend to like fantasy and comic genres, along with science fiction and war books, whereas users with high values of that factor prefer scary and humor books. Drama and educational books are generally preferred by people with high degree of AGR, and war, crime, and comic books are liked by those with a low degree of this factor. Finally, people with a high degree of NEU prefer crime and poetry books, while those with a lower degree prefer educational, thriller, mystery, and non-fiction books.

Some differences between **genders** are worth mentioning. For instance, in the **movie domain** the adventure, animation, and parody genres present significant differences with respect to which personality factors have higher and lower scores depending on the gender. Hence, adventure movies are preferred by female users with high EXT and AGR, whereas low degrees of EXT and neutral of AGR are observed for male users who like the genre. For animation movies, the opposite situation is observed; female users with low degrees of CON, EXT, and AGR tend to like the genre, but male users with high scores of EXT and AGR and neutral of CON are predominant. Similarly, for the parody genre, female users tend to have low degrees of CON and EXT and high degree of NEU, which is completely the opposite situation for male users. In the **TV show domain**, music videos, reality shows, soap operas, standup comedies, and surviving and talk shows present interesting differences between female and male users. On the other hand, in the **music domain** there are less differences, although some distinctions could be observed for jazz, pop, r&b, and salsa genres, where male users show higher degrees of AGR and CON factors. Finally, these differences are more remarkable in **the book domain**, where crime, drama, fiction, humor, mystery, non-fiction, romance, and self-help genres show significant variability in the personality scores for each gender. For instance, male users with high degrees of CON tend to like humor, non-fiction, romance, and self-help books, whereas for female users this factor seems to be neutral.

4.2 Association Rules Relating User Personality Factors and Domain-specific Preferences

In the previous section, relations between personality types and user preferences were derived by analyzing personality-based user stereotypes, and considered individual personality factors. In order to find more complex relations that take various personality factors into account, we applied the well-known Apriori algorithm, which generates association rules based on co-occurrences of attribute values in a set of data patterns.

To apply the Apriori algorithm we processed our dataset with the Weka[4] machine learning toolkit as follows. We transformed the dataset into a set of data patterns, each of them with 5 attributes corresponding to the Big Five personality factors of a particular user, and a discrete class label corresponding to a domain genre liked by such user. Each of the attributes had discrete values associated to 10 ranges of personality factor scores based on the attribute's factor score distribution, and are automatically generated by the Apriori algorithm. Moreover, aiming to generate generic (non overfitted) association rules, for each domain genre, we applied the Apriori algorithm on the 20 patterns most similar (i.e., with smallest Euclidean distance) to the genre's personality-based user stereotype. We ran the Apriori algorithm with other numbers of similar patterns, but here we do not report the rules obtained with them. When using less than 20 similar patterns we obtained a very small number of generic rules relating wide factor score ranges, whereas for more than 20 similar patterns we obtained a very large number of specific rules relating narrow factor score ranges.

The Apriori algorithm derives two strength measures for each rule $X \to Y$: *confidence* and *support*. The *confidence* represents the conditional probability $\Pr(Y|X)$ and is computed as $(X \cup Y).count/X.count$, and the *support* metric represents the probability $\Pr(X \cup Y)$ and is computed as $(X \cup Y).count/n$, where n is the total number of patterns, and $X.count$ is the number of patterns that contain the set of attribute values X.

Tables 2, 3, 4 and 5 show the generated association rules with highest confidence values. The reported support values give an idea of the "coverage" of the rules, i.e., the percentage of patterns satisfied by the rules.

The rules generated from **movie** preferences relate the OPE and EXT personality factors with different genres, depending on other personality factors. If a user with high OPE and EXT factors has high AGR, then she is likely to prefer comedy movies, but if she has low NEU, she is likely to prefer horror movies. Horror movies are also preferred by people with more complex personalities, such as those with high OPE, EXT, and AGR but low NEU. Additionally, cult movies tend to be liked by people with moderate CON and low AGR.

Regarding **TV show** preferences, only high confidence rules for two genres were generated: news and reality shows. People with high OPE and EXT, and moderately high CON tend to like news, whereas reality shows are liked by people with different types of personality: either high CON and EXT, and low NEU; high CON, EXT and OPE; or low NEU, and high OPE and AGR.

In the **music** domain, we found that people with high scores of OPE and EXT but low scores of NEU tend to like country music, whereas people with moderately low EXT and low NEU tend to like metal music. Jazz is liked by people with high EXT and AGR, and high CON or low NEU, whereas reggae is preferred by people with high OPE and AGR, and salsa is preferred by people with high scores of CON and EXT.

Finally, regarding **book** preferences, the rules indicate that people with high OPE and CON tend to like education books, whereas if CON is lower and OPE and AGR are high, people prefer science fiction books.

[4] Weka machine learning toolkit, http://www.cs.waikato.ac.nz/ml/weka

Taken into account the users' **gender**, it is worth noting that significant differences exist in the support values of the rules generated for female and male users. In the movie, TV show, and music domains, the support of the rules generated for male users was higher than that of the rules generated for female users, which may reflect a higher variability or complexity of the female personality-preference relations. In the **movie** domain, there was an interesting difference in the generation of rules for the animation genre, which is preferred by female users with low EXT and high NEU, and by male users with low CON and (much lower than female) NEU. In the **TV show** domain, news seemed to be preferred by female and male users sharing high OPE and CON and low NEU. In the **music** domain, the rules extracted for country music are very similar for female and male users (moderately high OPE, CON, and AGR). On the other hand, the rules for r&b music show differences: female users with NEU higher than male and high OPE, and male users with a moderately low CON and high EXT. Finally, in the **book** domain, no rules where generated for common genres, since the rules for female users were associated to crime, scary, and poetry books, and rules for male users were associated to humor books.

5 Relations between Personality Types and User Preferences in Crossed Domains

Table 6 shows the similarities between types of personality related to genres across the analyzed domains. In the table the colors of the cells correspond to relative values of the Euclidean distances between the personality-based user stereotypes associated to each pair of genres (maybe in distinct domains). Green cells correspond to high relative similarities (small distances), and red cells correspond to low relative similarities (large distances).

In the **movie domain** users who like action and comedy genres present very similar personality-based profiles. Similarly, people who like romance films show personality traits close to those who like comedy or drama. Users with tastes for cult movies show low similarity with other genres, in particular with those who like adventure movies. We can also observe that people who like foreign movies and people who like TV documentaries have similar personalities. Users that prefer tragic, independent or cult movies are in general dissimilar to people who like any kind of TV shows, especially to users that watch sports shows, a pattern that also appears in the music and book domains. Specifically, salsa and r&b are the farthest music genres, whereas educational, scary and thriller are the least similar book genres.

In the **TV show domain** there is little positive correlation between personality stereotypes of different genres. We can observe for instance differences between users who like sports shows and those who prefer animation and cartoons. We also notice that news, music videos, sitcoms and cooking shows are mostly independent of the user personality. With regard to the music domain there is a high similarity between documentaries and classical music, reality shows and dance music, and news and jazz. We also observe dissimilarity between people who like cartoons and those who enjoy hip hop, jazz or salsa. Regarding books, we see that comics and fantasy are the genres with users with similar personalities who like animation TV shows. On the other hand, people who like cartoons have on average a dissimilar personality to those who prefer educational books, and users who prefer sports shows have low similarity with respect to book preferences in general, specially comic, fantasy, science fiction and poetry.

In the **music domain** we can see that stereotypical users who like indie and metal music are quite dissimilar to those who like country, hip hop, r&b and salsa music. Interestingly, people who enjoy country and dance music show similar personality traits.

Finally, in the **book domain** we see that the genres with the closest associated user personalities are thrillers and mystery, as opposed to educational and war, or science fiction and scary books. Relating domains, we observe that in general, comic, fantasy, science fiction, poetry and war books link to a few personality traits in common with most music genres. Moreover, people who enjoy humor, mystery and romance books, on the other hand, have associated personality-based stereotypes similar to most of the music genre personality-based stereotypes.

6 Future Work

In this paper, we have preliminary analyzed the relations between user personality types and preferences. We have restricted our analysis to explicit preferences for a limited number of genres in the movie, TV, music, and book domains. However, as done by other authors [5], it would also be interesting to analyze those relations for particular items, such as representative or popular movies, TV shows, songs, and books. In our analysis we have focused on the users' gender as a characteristic that may influence the underlying personality-based preferences. In addition to or in combination with it, other user characteristics could be utilized. In this context, we consider interesting to explore the users' age and educational attainment. For instance, we may hypothesize that people with highest levels of education are more open-minded, and have larger and more diverse sets of preferences. Moreover, we have focused on the relations of personality-based stereotypes with preferences for single domains and genres. We plan to extend our analysis by considering relations involving several domains and genres. For instance, we could analyze the number and diversity of user preferences for particular personality-based stereotypes.

In any case, we believe that, although preliminary, the obtained relations between personality types and user preferences can be very valuable to enhance personalization services, not only on single domains, but also in cross-domain scenarios where user preferences on a source domain are used to infer user preferences on a different target domain, which can be a useful approach in e-commerce systems, among others. Hence, for example, we could investigate if a fact like "people with personality type x enjoy items i and j" (i and j belonging to different domains), could be used to suggest i, j (and other related items) to a person from whom we only known that has personality type x, or could be used to suggest item j to a person from whom we only know that has a preference for i (of a different domain of j). This, together with the investigation of more exhaustive automatic processes to infer and model user personality-preference relations, represents the next steps of our research.

	All users						Female users						Male users					
MOVIE GENRE	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**
action	3.87	3.45	3.57	3.58	2.72	2488	3.89	3.46	3.57	3.61	2.85	1361	3.85	3.44	3.57	3.56	2.57	1127
adventure	3.91	3.56	3.54	3.68	2.61	179	3.88	3.55	3.64	3.77	2.66	104	3.96	3.57	3.39	3.55	2.52	75
animation	4.04	3.22	3.26	3.35	3.02	85	4.02	3.14	3.10	3.22	3.28	56	4.09	3.39	3.55	3.59	2.52	29
cartoon	3.95	3.33	3.49	3.57	2.81	957	3.92	3.32	3.48	3.55	2.88	655	4.03	3.34	3.51	3.61	2.66	302
comedy	3.88	3.44	3.58	3.60	2.75	3969	3.87	3.48	3.61	3.62	2.86	2392	3.89	3.39	3.54	3.56	2.60	1577
cult	4.27	3.10	3.45	3.40	3.16	38	4.21	2.97	3.54	3.27	3.34	22	4.36	3.26	3.33	3.59	2.90	16
drama	3.99	3.43	3.66	3.60	2.86	905	3.96	3.44	3.65	3.63	2.89	668	4.07	3.41	3.67	3.51	2.78	237
foreign	4.15	3.46	3.47	3.54	2.81	112	4.20	3.52	3.53	3.54	2.81	90	3.95	3.22	3.22	3.53	2.84	22
horror	3.90	3.38	3.52	3.47	2.91	2284	3.89	3.38	3.53	3.47	2.98	1614	3.94	3.37	3.49	3.45	2.75	670
independent	4.31	3.59	3.51	3.55	2.69	104	4.27	3.56	3.49	3.53	2.79	65	4.36	3.65	3.56	3.60	2.52	39
neo-noir	4.34	3.35	3.33	3.37	2.97	92	4.39	3.41	3.40	3.39	3.07	46	4.28	3.28	3.25	3.35	2.87	46
parody	4.13	3.36	3.35	3.28	2.73	25	4.18	3.08	2.84	3.22	3.07	10	4.10	3.56	3.68	3.33	2.50	15
romance	3.84	3.48	3.62	3.62	2.85	776	3.83	3.48	3.63	3.62	2.93	610	3.88	3.50	3.59	3.63	2.59	166
science fiction	3.99	3.55	3.33	3.57	2.73	215	3.95	3.59	3.36	3.58	2.84	96	4.02	3.51	3.30	3.57	2.63	119
tragedy	4.40	3.34	3.27	3.52	3.11	26	4.31	3.45	3.37	3.44	3.09	18	4.60	3.07	3.05	3.69	3.15	8
war	3.82	3.51	3.49	3.50	2.71	148	3.99	3.57	3.62	3.56	2.94	44	3.75	3.49	3.44	3.47	2.62	104
	4.05	**3.41**	**3.46**	**3.51**	**2.84**		**4.05**	**3.40**	**3.46**	**3.50**	**2.95**		**4.07**	**3.40**	**3.45**	**3.54**	**2.69**	

	All users						Female users						Male users					
TV GENRE	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**
animation	4.03	3.23	3.25	3.48	2.83	166	4.07	3.22	3.25	3.51	2.96	108	3.96	3.23	3.24	3.43	2.58	58
cartoon	3.77	3.21	3.22	3.48	3.03	64	3.84	3.18	3.27	3.47	3.10	39	3.65	3.26	3.13	3.51	2.91	25
cooking show	3.71	3.54	3.54	3.51	2.76	134	3.68	3.56	3.53	3.53	2.80	108	3.83	3.45	3.59	3.46	2.60	26
documentary	4.12	3.45	3.37	3.53	2.85	722	4.14	3.45	3.41	3.57	2.88	466	4.09	3.46	3.30	3.45	2.79	256
game/quiz show	3.58	3.45	3.30	3.68	2.90	72	3.53	3.55	3.32	3.62	3.09	44	3.65	3.29	3.27	3.78	2.61	28
lgbt show	4.28	3.28	3.54	3.51	2.95	62	4.24	3.31	3.44	3.42	2.98	39	4.35	3.21	3.72	3.66	2.90	23
music video	3.98	3.32	3.56	3.55	3.02	163	3.95	3.30	3.50	3.60	3.13	123	4.09	3.38	3.76	3.39	2.67	40
news	3.97	3.58	3.58	3.54	2.74	676	3.96	3.59	3.65	3.58	2.82	374	3.98	3.56	3.49	3.50	2.63	302
prank show	4.08	3.41	3.51	3.33	2.65	65	4.13	3.40	3.63	3.36	2.78	31	4.04	3.43	3.41	3.31	2.54	34
reality show	3.76	3.56	3.61	3.58	2.75	808	3.73	3.55	3.63	3.58	2.78	623	3.86	3.61	3.56	3.55	2.64	185
sitcom	3.93	3.37	3.22	3.44	2.78	53	3.82	3.26	3.26	3.50	2.98	29	4.06	3.49	3.16	3.36	2.53	24
soap opera	3.57	3.46	3.56	3.42	3.01	151	3.56	3.45	3.51	3.40	3.05	135	3.66	3.52	3.96	3.59	2.67	16
sports show	3.67	3.58	3.67	3.64	2.52	119	3.68	3.73	3.62	3.72	2.74	37	3.66	3.52	3.70	3.60	2.42	82
standup comedy	4.10	3.36	3.53	3.49	2.85	339	4.07	3.43	3.59	3.42	3.06	156	4.12	3.29	3.48	3.55	2.67	183
surviving show	4.11	3.62	3.48	3.26	2.66	43	4.06	3.89	3.66	3.26	2.75	16	4.14	3.46	3.37	3.26	2.60	27
talk show	3.81	3.58	3.58	3.68	2.67	61	3.79	3.71	3.67	3.62	2.50	38	3.86	3.38	3.43	3.77	2.95	23
	3.90	**3.44**	**3.47**	**3.51**	**2.81**		**3.89**	**3.47**	**3.50**	**3.51**	**2.90**		**3.94**	**3.41**	**3.47**	**3.51**	**2.67**	

	All users						Female users						Male users					
MUSIC GENRE	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**users**	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**
blues	4.08	3.48	3.55	3.57	2.71	1054	4.05	3.51	3.57	3.64	2.81	501	4.10	3.46	3.53	3.52	2.62	553
classical	4.10	3.50	3.39	3.54	2.77	2235	4.10	3.55	3.40	3.57	2.87	1231	4.11	3.44	3.37	3.50	2.66	1004
country	3.79	3.55	3.64	3.65	2.78	7475	3.79	3.54	3.65	3.67	2.86	5370	3.79	3.55	3.62	3.58	2.56	2105
dance	3.79	3.50	3.66	3.60	2.79	456	3.74	3.51	3.64	3.61	2.95	275	3.86	3.50	3.69	3.58	2.54	181
hip hop	3.82	3.49	3.75	3.52	2.66	2510	3.81	3.50	3.78	3.52	2.77	1513	3.82	3.46	3.71	3.53	2.49	997
indie	4.06	3.26	3.51	3.49	2.98	476	4.05	3.22	3.45	3.50	3.12	309	4.09	3.35	3.63	3.47	2.73	167
jazz	4.03	3.55	3.61	3.60	2.64	1735	4.02	3.61	3.65	3.63	2.73	944	4.05	3.48	3.57	3.58	2.53	791
metal	3.97	3.30	3.37	3.34	2.87	1273	4.03	3.34	3.35	3.40	3.05	474	3.93	3.27	3.38	3.30	2.76	799
oldies	3.86	3.51	3.57	3.65	2.75	932	3.83	3.53	3.59	3.67	2.81	693	3.93	3.45	3.51	3.59	2.58	239
pop	3.79	3.41	3.55	3.55	2.77	412	3.80	3.39	3.56	3.54	2.86	285	3.79	3.45	3.53	3.58	2.57	127
r&b	3.69	3.54	3.63	3.57	2.71	1367	3.66	3.55	3.63	3.56	2.78	974	3.75	3.54	3.64	3.58	2.52	393
rap	3.74	3.33	3.69	3.43	2.67	323	3.73	3.36	3.66	3.45	2.88	163	3.75	3.30	3.72	3.41	2.45	160
reggae	4.00	3.43	3.66	3.53	2.67	2802	3.98	3.46	3.68	3.52	2.77	1701	4.02	3.37	3.63	3.56	2.53	1101
rock	3.93	3.37	3.48	3.51	2.83	6196	3.95	3.38	3.46	3.50	2.99	3429	3.91	3.35	3.51	3.52	2.64	2767
salsa	3.94	3.55	3.80	3.55	2.62	467	3.93	3.53	3.82	3.56	2.69	334	3.98	3.58	3.77	3.53	2.45	133
techno	4.00	3.32	3.51	3.50	2.81	2481	4.01	3.31	3.51	3.49	2.97	1249	3.99	3.33	3.51	3.51	2.64	1232
	3.91	**3.44**	**3.59**	**3.54**	**2.75**		**3.91**	**3.46**	**3.59**	**3.55**	**2.87**		**3.93**	**3.43**	**3.58**	**3.52**	**2.58**	

	All users						Female users						Male users					
BOOK GENRE	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**	**OPE**	**CON**	**EXT**	**AGR**	**NEU**	**#users**
comic	4.06	3.28	3.38	3.47	2.86	1107	4.05	3.24	3.36	3.49	2.98	540	4.08	3.31	3.40	3.45	2.73	567
crime	3.83	3.44	3.43	3.47	2.99	191	3.82	3.46	3.38	3.47	3.06	146	3.88	3.39	3.59	3.46	2.76	45
drama	3.81	3.36	3.53	3.67	2.84	66	3.83	3.38	3.55	3.72	2.95	52	3.75	3.29	3.43	3.51	2.44	14
educational	4.02	3.66	3.57	3.66	2.74	977	3.97	3.71	3.59	3.68	2.82	656	4.12	3.55	3.53	3.62	2.57	321
fantasy	4.04	3.34	3.27	3.54	2.87	994	4.03	3.34	3.29	3.56	2.97	624	4.05	3.35	3.23	3.51	2.70	370
fiction	4.00	3.41	3.42	3.55	2.82	339	3.97	3.43	3.45	3.53	2.90	214	4.04	3.36	3.37	3.60	2.67	125
humor	3.90	3.40	3.62	3.56	2.78	743	3.88	3.36	3.61	3.56	2.94	470	3.93	3.47	3.64	3.56	2.51	273
mystery	3.91	3.53	3.51	3.61	2.76	302	3.93	3.58	3.53	3.63	2.77	219	3.83	3.39	3.43	3.55	2.71	83
non fiction	4.01	3.51	3.43	3.62	2.76	319	4.02	3.51	3.49	3.65	2.87	205	4.00	3.52	3.31	3.57	2.58	114
poetry	4.16	3.34	3.38	3.54	2.94	160	4.11	3.35	3.41	3.59	2.98	108	4.25	3.32	3.33	3.45	2.86	52
romance	3.89	3.52	3.49	3.60	2.85	1132	3.88	3.52	3.49	3.61	2.86	987	3.99	3.52	3.47	3.60	2.80	145
scary	3.81	3.41	3.68	3.55	2.83	1084	3.83	3.43	3.68	3.54	2.89	822	3.75	3.36	3.67	3.55	2.61	262
science fiction	4.13	3.42	3.25	3.51	2.81	1191	4.15	3.44	3.25	3.52	2.95	552	4.12	3.40	3.25	3.50	2.68	639
self help	4.03	3.50	3.42	3.62	2.83	196	4.05	3.49	3.45	3.71	2.98	129	3.98	3.53	3.35	3.45	2.55	67
thriller	3.85	3.54	3.51	3.59	2.76	639	3.86	3.55	3.53	3.61	2.87	410	3.84	3.50	3.47	3.55	2.56	229
war	3.87	3.44	3.33	3.23	2.80	108	4.14	3.45	3.41	3.29	3.05	15	3.83	3.44	3.32	3.22	2.76	93
	3.96	**3.44**	**3.45**	**3.55**	**2.83**		**3.97**	**3.45**	**3.47**	**3.57**	**2.93**		**3.97**	**3.42**	**3.42**	**3.51**	**2.66**	

Table 1. Personality-based user stereotypes in individual domain genres.

MOVIES	**Rule**	**Confidence**	**Support**
All users (support ≤ 20.30 %)	$con \in (3, 3.25] \wedge agr \in [2.55, 2.87] \rightarrow$ cult	67 %	1.87 %
	$ope \in (3.6, 3.80] \wedge ext \in (3.35, 3.62] \wedge agr \in (3.52, 3.85] \rightarrow$ comedy	67 %	1.87 %
	$ope \in (3.8, 4] \wedge con \in (3.25, 3.5] \wedge agr \in (3.2, 3.52] \wedge neu \in (2.85, 3.17] \rightarrow$ horror	67 %	1.87 %
	$ope \in (4.88, 5] \rightarrow$ tragedy	63 %	2.50 %
	$ope \in (4.4, 4.6] \wedge ext \in (3.62, 3.89] \rightarrow$ foreign	63 %	2.50 %
	$ope \in (3.6, 3.8] \wedge ext \in (3.62, 3.89] \wedge agr \in (3.2, 3.52] \rightarrow$ horror	57 %	2.19 %
	$ope \in (3.6, 3.8] \wedge ext \in (3.62, 3.89] \wedge agr \in (3.20, 3.52] \wedge neu \in (2.53, 2.85] \rightarrow$ horror	57 %	2.19 %
	$ope \in (3.6, 3.8] \wedge agr \in (3.2, 3.52] \wedge neu \in (2.53, 2.85] \rightarrow$ horror	50 %	2.50 %
	$ope \in (3.8, 4] \wedge ext \in (3.62, 3.89] \wedge agr \in (3.52, 3.85] \rightarrow$ comedy	50 %	2.81 %
Female users (support ≤ 6.82 %)	$ext \in (2.37, 2.75] \wedge neu \in (3.7, 4.02] \rightarrow$ animation	67 %	1.95 %
	$ope \in (3.8, 4.1] \wedge ext \in (3.5, 3.87] \wedge neu \in (2.4, 2.72] \rightarrow$ cartoon	57 %	2.27 %
	$con \in (3.67, 3.98] \wedge agr \in (3.3, 3.6] \wedge neu \in (2.72, 3.05] \rightarrow$ romance	50 %	2.60 %
Male users (support ≤ 81.61 %)	$ext \in (4.2, 4.6] \wedge neu \in (1.79, 2.1] \rightarrow$ independent	100 %	1.34 %
	$ope \in (4.6, 4.8] \wedge con \in (3.25, 3.6] \rightarrow$ neo-noir	100 %	1.00 %
	$con \in (3.25, 3.6] \wedge neu \in (1.47, 1.79] \rightarrow$ animation	100 %	1.00 %
	$ope \in (4.2, 4.4] \wedge con(3.25, 3.6] \wedge ext \in (3.8, 4.2] \rightarrow$ animation	100 %	1.00 %
	$ope \in (4.2, 4.4] \wedge con(3.6, 3.95] \wedge neu \in (2.74, 3.058] \rightarrow$ drama	80 %	1.67 %
	$con \in (1.85, 2.2] \rightarrow$ cult	75 %	1.34 %
	$ope \in (3.4, 3.6] \wedge ext \in (3.8, 4.2] \rightarrow$ war	75 %	1.34 %
	$con \in (4.65, 5] \wedge \mathrm{agr} \in (3.7, 4.05] \rightarrow$ independent	75 %	1.34 %
	$ext \in (3, 3.4] \wedge neu \in (2.10, 2.42] \rightarrow$ adventure	75 %	1.34 %
	$ope \in (3.8, 4] \wedge ext \in (3, 3.4] \wedge neu \in (2.42, 2.74] \rightarrow$ comedy	75 %	1.34 %
	$ope \in (4.2, 4.4] \wedge agr \in (3.35, 3.7] \wedge neu \in (1.786, 2.104] \rightarrow$ adventure	75 %	1.34 %
	$con \in (3.6, 3.9] \wedge ext \in (3, 3.4] \wedge neu \in (2.74, 3.06] \rightarrow$ drama	75 %	1.34 %
	$ope \in (3.6, 3.8] \wedge ext \in (3.4, 3.8] \wedge agr \in (3.7, 4.05] \wedge neu \in (2.42, 2.74] \rightarrow$ action	75 %	1.34 %

Table 2. Association rules relating user personality factors and movie preferences.

TV	Rule	Confidence	Support
All users (support ≤ 5.93 %)	$ope \in (3.65, 3.87] \wedge con \in (3.18, 3.4] \wedge ext \in (3.47, 3.73] \rightarrow$ news	100 %	1.25 %
	$con \in (3.62, 3.84] \wedge ext \in (3.47, 3.73] \wedge neu \in (2.65, 2.87] \rightarrow$ reality show	100 %	1.25 %
	$ope \in (3.65, 3.87] \wedge agr \in (3.69, 3.87] \wedge neu \in (2.65, 2.87] \rightarrow$ reality show	80 %	1.56 %
	$ope \in (3.65, 3.87] \wedge con \in (3.62, 3.84] \wedge ext \in (3.47, 3.73] \rightarrow$ reality show	67 %	1.87 %
Female users (support ≤ 3.48 %)	$ope \in (4, 4.25] \wedge con \in (3.55, 3.84] \wedge neu \in (2.66, 3.00] \rightarrow$ news	100 %	1.27 %
	$con \in (4.13, 4.42] \rightarrow$ talk show	57 %	2.21 %
Male users (support ≤ 14.58 %)	$con \in (3.78, 4.08] \wedge agr \in (3.2, 3.5] \wedge neu \in (2.08, 2.43] \rightarrow$ sports	100 %	1.27 %
	$ope \in (4.33, 4.67] \wedge con \in (2.86, 3.17] \rightarrow$ lgbt show	57 %	2.22 %
	$ope \in (4.67, 5] \wedge agr \in (2.9, 3.2] \rightarrow$ surviving show	57 %	2.22 %
	$ope \in (3.67, 4] \wedge ext \in (3.47, 3.78] \wedge neu \in (2.77, 3.12] \rightarrow$ news	57 %	2.22 %
	$ope \in (4.33, 4.67] \wedge ext \in (3.47, 3.78] \wedge neu \in (2.43, 2.77] \rightarrow$ standup comedy	57 %	2.22 %
	$agr \in (3.2, 3.5] \wedge neu \in (2.08, 2.43] \rightarrow$ sports	50 %	1.90 %
	$ope \in (4, 4.33] \wedge neu \in (3.12, 3.46] \rightarrow$ music video	50 %	2.53 %

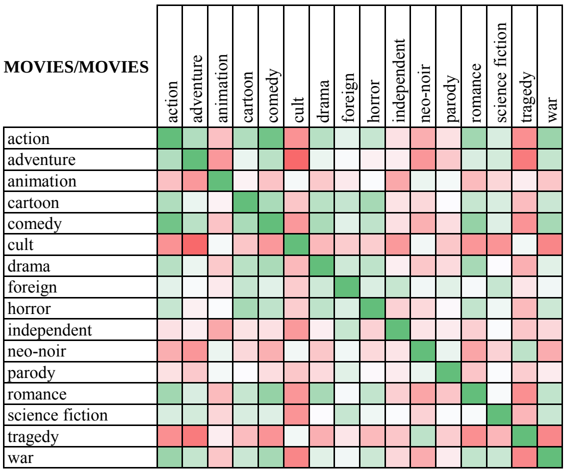
Table 3. Association rules relating user personality factors and TV preferences.

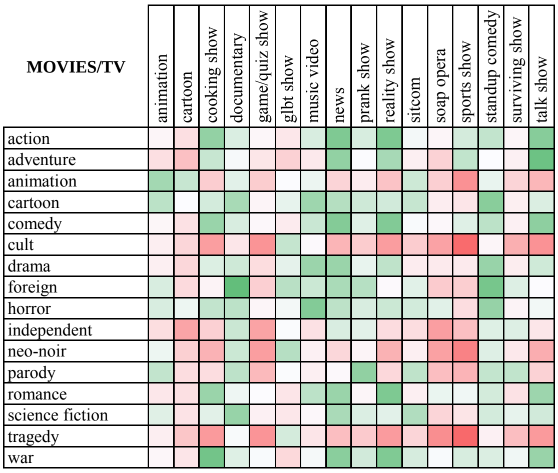
MUSIC	Rule	Confidence	Support
All users (support ≤ 14.35 %)	$con \in (3.64, 3.76] \wedge ext \in (3.635, 3.774] \wedge agr \in (3.598, 3.731] \rightarrow$ jazz	80 %	1.56 %
	$ope \in (3.75, 3.875] \wedge agr \in (3.465, 3.598] \rightarrow$ reggae	67 %	1.87 %
	$con \in (3.64, 3.76] \wedge ext \in (3.913, 4.052] \rightarrow$ salsa	67 %	1.87 %
	$ope \in (3.625, 3.75] \wedge con \in (3.4, 3.52] \wedge ext \in (3.496, 3.635] \rightarrow$ country	67 %	1.87 %
	$ext \in (3.635, 3.774] \wedge agr \in (3.598, 3.731] \wedge neu \in (2.49, 2.61] \rightarrow$ jazz	67 %	1.87 %
	$ext \in (3.218, 3.357] \wedge neu \in (2.97, 3.09] \rightarrow$ metal	57 %	2.19 %
	$ope \in (3.625, 3.75] \wedge ext \in (3.496, 3.635] \wedge neu \in (2.49, 2.61] \rightarrow$ country	50 %	3.12 %
Female users (support ≤ 7.81 %)	$ope \in (4.25, 4.375] \wedge con \in (3.42, 3.57] \rightarrow$ classic	80 %	1.56 %
	$ope \in (3.625, 3.75] \wedge con \in (3.42, 3.57] \wedge ext \in (3.4, 3.55] \rightarrow$ country	80 %	1.56 %
	$ope \in (3.625, 3.75] \wedge con \in (3.42, 3.57] \wedge agr \in (3.652, 3.789] \rightarrow$ country	57 %	2.19 %
	$ope \in (3.375, 3.5] \wedge neu \in (2.91, 3.058] \rightarrow$ r&b	50 %	2.50 %
Male users (support ≤ 12.49 %)	$agr \in [1, 2.915] \rightarrow$ rap	80 %	1.56 %
	$ope \in (3.35, 3.525] \wedge con \in (2.9, 3.05] \rightarrow$ pop	80 %	1.56 %
	$ope \in (3.7, 3.875] \wedge con \in (3.35, 3.5] \wedge agr \in (3.41, 3.575] \rightarrow$ country	80 %	1.56 %
	$con \in (3.2, 3.35] \wedge ext \in (3.47, 3.625] \wedge neu \in (2.618, 2.776] \rightarrow$ r&b	80 %	1.56 %
	$con \in (2.9, 3.05] \wedge agr \in (2.915, 3.08] \rightarrow$ metal	67 %	1.87 %
	$ope \in (3.875, 4.05] \wedge agr \in (2.915, 3.08] \rightarrow$ metal	57 %	2.19 %
	$con \in (3.35, 3.5] \wedge agr \in (3.41, 3.575] \wedge neu \in (2.144, 2.302] \rightarrow$ country	57 %	2.19 %

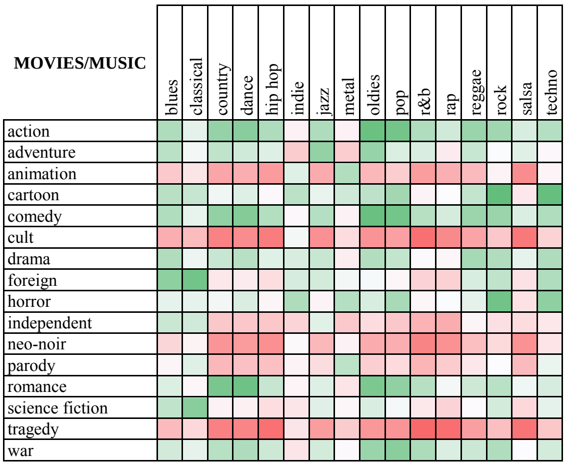
Table 4. Association rules relating user personality factors and music preferences.

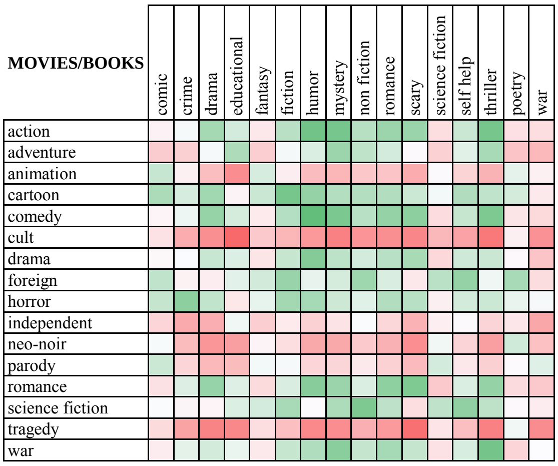
BOOKS	Rule	Confidence	Support
All users (support ≤ 8.74 %)	$ope \in (4.09, 4.28] \wedge con \in (3.9, 4.07] \rightarrow$ education	67 %	1.87 %
	$ope \in (3.91, 4.09] \wedge con \in (3.375, 3.55] \wedge agr \in (3.5, 3.65] \rightarrow$ science fiction	67 %	1.87 %
	$ope \in (1, 3.37] \rightarrow$ drama	57 %	2.19 %
	$ope \in (3.91, 4.09] \wedge agr \in (3.5, 3.65] \rightarrow$ science fiction	56 %	2.81 %
Female users (support ≤ 9.53 %)	$ope \in (3.4, 3.6] \wedge agr \in (3.95, 4.3] \rightarrow$ crime	80 %	1.59 %
	$ope \in (3.8, 4] \wedge ext \in (3.62, 3.9] \wedge neu \in (3, 3.4] \rightarrow$ scary	63 %	2.54 %
	$ext \in (2.52, 2.8] \rightarrow$ poetry	56 %	2.86 %
	$ope \in (4.6, 4.8] \rightarrow$ poetry	50 %	2.54 %
Male users (support ≤ 7.21 %)	$ope \in (3.52, 3.89] \wedge agr \in (3.25, 3.55] \wedge neu \in (2.04, 2.32] \rightarrow$ humor	80 %	1.57 %
	$ext \in (3.86, 4.16] \wedge neu \in (2.04, 2.32] \rightarrow$ humor	57 %	2.19 %
	$ope \in (3.52, 3.89] \wedge neu \in (2.04, 2.32] \rightarrow$ humor	55 %	3.45 %

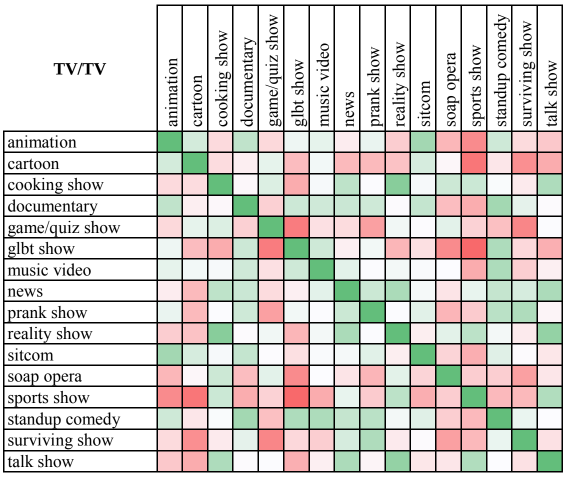
Table 5. Association rules relating user personality factors and book preferences.


MOVIES/MOVIES
action
adventure
animation
cartoon
comedy
cult
drama
foreign
horror
independent
neo-noir
parody
romance
science fiction
tragedy
war


MOVIES/TV
animation
cartoon
cooking show
documentary
game/quiz show
glbt show
music video
news
prank show
reality show
sitcom
soap opera
sports show
standup comedy
surviving show
talk show
action
adventure
animation
cartoon
comedy
cult
drama
foreign
horror
independent
neo-noir
parody
romance
science fiction
tragedy
war


MOVIES/MUSIC
blues
classical
country
dance
hip hop
indie
jazz
metal
oldies
pop
r&b
rap
reggae
rock
salsa
techno
action
adventure
animation
cartoon
comedy
cult
drama
foreign
horror
independent
neo-noir
parody
romance
science fiction
tragedy
war


MOVIES/BOOKS
comic
crime
drama
educational
fantasy
fiction
humor
mystery
non fiction
romance
scary
science fiction
self help
thriller
poetry
war
action
adventure
animation
cartoon
comedy
cult
drama
foreign
horror
independent
neo-noir
parody
romance
science fiction
tragedy
war


TV/TV
animation
cartoon
cooking show
documentary
game/quiz show
glbt show
music video
news
prank show
reality show
sitcom
soap opera
sports show
standup comedy
surviving show
talk show

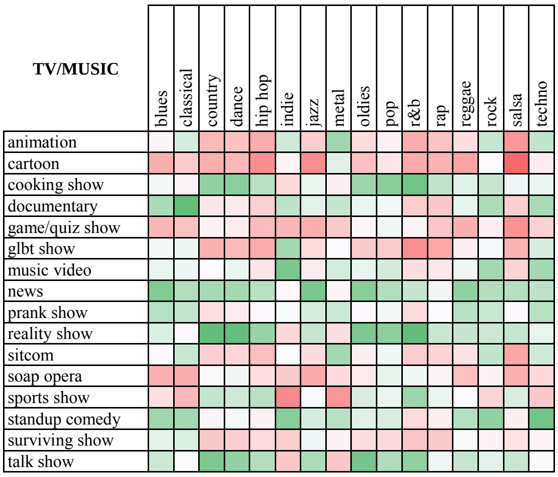

TV/MUSIC
blues
classical
country
dance
hip hop
indie
jazz
metal
oldies
pop
r&b
rap
reggae
rock
salsa
techno
animation
cartoon
cooking show
documentary
game/quiz show
glbt show
music video
news
prank show
reality show
sitcom
soap opera
sports show
standup comedy
surviving show
talk show

Table 6. Similarities between personality-based user stereotypes for genres in different domains.

Acknowledgements

This work was supported by the Spanish Government (TIN2011-28538-C02) and the Regional Government of Madrid (S2009TIC-1542). The authors sincerely thank the members of myPersonality project for their kind attention and help on downloading and processing the provided data.

References

1. Adomavicius, G., Tuzhilin, A.: Context-Aware Recommender Systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P. B. (Eds.), *Recommender Systems Handbook*, pp. 217-253 (2011)
2. Bachrach, Y., Kohli, P., Graepel, T., Stillwell, D. J., Kosinski, M.: Personality and Patterns of Facebook Usage. In: *Proceedings of the 4th Web Science Conference (WebSci'12)*, pp. 24-32 (2012)

3. Brusilovsky, P., Kobsa, A., Nejdl, W.: *The Adaptive Web, Methods and Strategies of Web Personalization*. Springer Verlag (2007)
4. Cattell, R. B., Marshall, M. B., Georgiades, S.: Personality and Motivation: Structure and Measurement. *Journal of Personality Disorders* 19(1), pp. 53-67 (1957)
5. Chausson, O.: Who Watches What? Assessing the Impact of Gender and Personality on Film Preferences. *myPersonality project*, Univeristy of Cambridge (2010)
6. Costa, P. T., McCrae, R. R.: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual. *Psychological Assessment Resources* (1992)
7. Digman, J. M.: Five Robust Trait Dimensions: Development, Stability, and Utility. *Journal of Personality* 57 (2), 195-214 (1989)
8. Ekman, P., Davidson, R. J.: *The Nature of Emotions: Fundamental Questions*. Oxford University Press (1994)
9. George, J. M.: Trait and State Affect. In: Murphy, K. R. (Ed.), *Individual Differences and Behavior in Organization*, pp. 145-171 (1996)
10. Goldberg, L. R.: The Structure of Phenotypic Personality Traits. *American Psychologist* 48(1), pp. 26-34 (1993)
11. Grucza, R. A., Goldberg, L. R.: The Comparative Validity of 11 Modern Personality Inventories: Predictions of Behavioral Acts, Informant Reports, and Clinical Indicators. *Journal of Personality Assessment* 89(2), pp. 167-187 (2007)
12. Hu, R., Pu, P.: A Study on User Perception of Personality-Based Recommender Systems. In: *Proceedings of the 18th International Conference on User Modeling, Adaptation, and Personalization (UMAP'10)*, pp. 291-302 (2010)
13. Hu, R., Pu, P.: Enhancing Collaborative Filtering Systems with Personality Information. In: *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys'11)*, pp. 197-204 (2011)
14. Kosinski, M., Stillwell, D. J.: myPersonality Research. *myPersonality project*, University of Cambridge, http://mypersonality.org (2011)
15. McCrae, R. R., Costa, P. T.: Validation of the Five-factor Model of Personality across Instruments and Observers. *Journal of Personality and Social Psychology* 52(1), pp. 81-90 (1987)
16. Nunes, M. A. S. N.: *Recommender Systems based on Personality Traits: Could Human Psychological Aspects Influence the Computer Decision-making Process?* VDM Verlag (2009)
17. Nunes, M. A. S. N., Hu, R.: Personality-based Recommender Systems: An Overview. In: *Proc. of the 6th ACM Conference on Recommender Systems (RecSys'12)*, pp. 5-6 (2012)
18. O'Connor, B.: A Quantitative Review of the Comprehensiveness of the Five-Factor Model in Relation to Popular Personality Inventories. *Assessment* 9(2), pp. 188-203 (2002)
19. Rentfrow, P.J. and 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), pp. 1236-1256 (2003)
20. Roshchina, A.: *TWIN: Personality-based Recommender System*. MSc thesis, Institute of Technology Tallaght, Ireland (2012)
21. Schacter, D. L., Gilbert, D. T., Wegner, D. M.: *Psychology*. Worth Publishers (2010)
22. Tkalčič, M., Kunaver, M., Tasič, J., Košir, A.: 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*, pp. 30-37 (2009)
23. Tupes, E. C., Christal, R. E.: Recurrent Personality Factors Based on Trait Ratings. Technical Report ASD-TR-61-97, Personnel Laboratory, Air Force Systems Command, Lackland Air Force Base, TX, USA (1961)

Exploring Relations between Personality and User Rating Behaviors

Rong Hu, Pearl Pu

Human Computer Interaction Group
Swiss Federal Institute of Technology in Lausanne (EPFL)
CH-1015, Lausanne, Switzerland
{rong.hu,pearl.pu}@epfl.ch

Abstract. In this study, we conducted an online survey and collected 86 reliable responses on both a personality assessment inventory and ratings retail products ratings, with the aim of investigating whether personality characteristics have an impact on user rating behaviors. Besides personality factors, another four independent variables (i.e., age, gender, previous experience on using recommenders and e-commerce systems) were taken into account when we examined the relationship. The correlation analysis results show that Conscientiousness is negatively correlated with the number of total ratings, category coverage and interest diversity. Individuals high on Agreeableness tend to give more positive ratings. In addition, Gender plays a significant role on all rating behavior variables except percentage of positive ratings. We further explored users' personality profiles along the long tail of the number of ratings. We found that users high on Openness tend to rate more items than required, while low Conscientiousness is a critical factor which provokes users to rate items in an explosive way. Our findings are useful for researchers interested in user modeling, preference elicitation, recommender systems and online marketing.

Keywords. Personality, User Modeling, Rating Behavior, Preference

1 Introduction

Research in psychology has suggested that behavior and preferences of individuals can be explained to a great extent by underlying psychological constructs (or so called personality traits). For example, personality traits have been found to correlate with people's music tastes [1], and impact the formation of social relations [2]. In addition, personality is useful in predicting job success [3] and marital satisfaction [4].

Likewise, in online settings, previous research has shown that certain personality traits are correlated with total Internet usage, preference for different interfaces and with the propensity of users to use social media and social networking sites [5]. More recently, studies have demonstrated that personality characteristics significantly relate to people's social network profiles [6, 7]. Knowing an individual's personality enables us to predict his behavior and preferences across contexts and environments and to enhance user experience by personalizing interfaces and presented information.

In this paper, we are trying to investigate the relations between personality characteristics and user rating behaviors. Modeling users' preferences is one critical step in intelligent systems to tailor personalized services. For example, recommender systems (RS) seek to suggest (or recommend) unseen contents that a user would find to be of interest. A common approach in RS to build user preference models is asking users to explicitly rate items in order to infer their preferences. Therefore, investigating users' rating behaviors could benefit effectiveness and accuracy of user preference modeling [8]. However, to the best of our knowledge, little attempt has been made to relate psychological profiles to user rating behaviors yet.

We conducted an online survey and collected 86 validated responses. The results demonstrate that personality characteristics really have an influence on the way user gave ratings. Besides, gender variable plays a significant role on rating behavior variables. The main contributions of this paper include:

1. Investigate how user's personality characteristics would affect user rating behaviors, comprising of the number of ratings, the number of positive ratings, the categorical coverage of user ratings, and their interest diversity, considering age, gender and previous experience with user rating behaviors.
2. Explore the personality distinction along the long tail of user ratings.

Our results not only provide insights on the effect of user personality characteristics on user modeling, but also suggest practical applications in a variety of areas, including social media websites, e-commerce retailers and recommender systems.

The remainder of this paper is organized as follows. We begin by presenting Big Five Personality model in Section 2, and background and related work in Section 3. We then present our experiment methodology including materials, procedure and participants in Section 4. In Section 5, we describe our dataset by defining the rating behavior variables and independent variables. We provide detailed result analysis in Section 6 and a depth discussion of potential theoretical and practical implications in Section 7 followed by a conclusion.

2 Personality Model

We decided to use the Five Factor Model (FFM, or the Big Five Model) in this study, since it is currently the most widespread and generally accepted model of personality and its ability to predict human behavior has been well studied [9, 10]. This model has been shown to subsume the most known personality traits and provides a nomenclature and a conceptual framework that unifies much of the research findings in psychology of individual differences and personality.

The Five Factor Model divides personality into five dimensional traits: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). Each dimension has its representative characteristics.

- **Openness to experience** measures a person's imagination, curiosity, seeking of new experiences and interest in culture, ideas, and aesthetics.

- **Conscientiousness** reflects the degree to which an individual is organized, diligent and scrupulous.
- **Extraversion** measures a person's tendency to seek stimulation in the external world, company of others, and express positive emotions.
- **Agreeableness** measures the extent to which a person is focused on maintaining positive social relations, reflecting a tendency to be trustful, sympathetic and cooperative.
- **Neuroticism** often referred to as emotional instability, is a tendency to experience mood swings and negative emotions such as guilt, anger, anxiety, and depression.

The five traits have been observed to be genetically heritable, stable over time and consistent across genders, cultures, and races [11]. **Table 1** summarizes the big five personality traits along with their representative descriptive terms for both low and high scorers.

Table 1. Big five personality dimensions and representataive descriptive terms.

Trait	Description	Low scorer	High Scorer
Openness	A willingness to consider alternative approaches, be intellectually curious and enjoy artistic pursuits	Close-minded, Conventional	Imaginative, Curious
Conscien-tiousness	The degree to which an individual is organized, diligent and scrupulous.	Spontaneous, Creative	Organized, Reliable
Extraver-sion	A tendency to be sociable and able to experience positive emotions	Solitary, Reserved	Sociable, Energetic
Agreeable-ness	A tendency to be trusting, sympathetic and cooperative.	Competitive, Assertive	Cooperative, Trusting
Neuroticism	A tendency to experience psychological distress.	Emotionally stable, Self-confident	Prone to negative emotions

3 Background and Related Work

Prior research has shown that personality can efficiently explain a substantial amount of variability in human preferences and behavior across different domains, for example media and cultural preferences [1, 12], and social networking websites usage [6].

According to information processing theory, the satisfaction people derive from outside stimulation, depends on their optimal or preferred arousal levels. One's preference over one item is thought to be affected by the corresponding information processing capacity and affective orientations [13]. Personality is therefore found to be relevant for understanding individuals' appreciation of the arts, for example, paintings and music [1, 14]. Recent research suggested that personality characteristics could be considered as important mediators of media content preferences. Kraaykamp and Eijck [12] examined the impact of the Big Five personality factors on media preferences (TV programs) and cultural participation (book reading and attending museums and concerts). They found that openness clearly encourages an interest in complex and exciting recreational practices. Conscientiousness and friendliness (agreeableness) tend to have negative effects on activities that are either difficult or unconventional, whereas emotional stability negatively influences more predictable means of escape from everyday life. The work in [15] showed that website preferences are influenced by personality characteristics, like those for objects in real world. The authors found that website audiences often have distinct personality profiles, and the relationship between personality and preferences related to website and website categories is psychologically meaningful.

Recently, social media websites (e.g., Facebook, Twitter) have emerged as a major media people communicate with each other and express their personal opinions. Researchers have become interested in how personality impacts user interactions on those social media websites. The work in [16] showed that Extroverts tend to find social media site easy to use and useful. Users are likely to select contacts with similar personality characteristics, and they generally tend to prefer people high in Agreeableness [17]. Current study interests have been more focused on the relations between personality and users' usage behaviors (e.g., the number of posts, likes) and profiles (e.g., the number of friends/followings/followers, age, gender) in social websites [6, 7]. Moreover, increasing attention has been paid on the prediction of personality traits scores based on those publically available behavior and profile information [7, 18].

Golbeck et al. [18] shown that users with different personality tend to use disparate words in their posts and descriptions. Quercia et al. [7] studied Twitter users and found that both popular users and influentials are extroverts and emotionally stable. They further discovered that popular users are 'imaginative' (high in Openness), while influentials tend to be 'organized' (high in Conscientiousness). In [6], Quercia et al. examined the relationship between sociometric popularity (number of Facebook contacts) and personality traits on a different social networking platform, Facebook. They concluded that popular Facebook users tend to have the same personality as people popular in the real world. Similarly, [19] demonstrated a significant connection between personality traits and various features of Facebook profiles.

To the best of our knowledge, few studies have been done on the effects of personality on users' behavior in user preference modeling. In this paper, we are trying to answer this central research question: to what extent does personality factors affect rating behaviors?

4 Methodology

4.1 Materials

We crawled detailed information of totally 18,793 retail products from gifts.com, a gift finder recommender system, covering 44 primary categories (e.g., accessories, alcohol & tobacco, arts & crafts, etc.). The category ontology given by gifts.com is a structure of three levels. For example, under primary category accessories, there contain categories: cufflinks, handbags & briefcases, shoes, ties & suit accessories, wallets & small goods, hats, gloves & scarves and other accessories (so-called subcategory). The shoes sub-category is further divided into casual shoes, dressy shoes and slippers (so-called subsub-category). Thanks to gifts.com, all products have a label for gender. That is, it is known whether one product suits women or men. Using this information, we constraint a user in the product space which contains products match his/her gender. For example, a female user cannot see and rate the products labeled with male. By doing so, we could reduce users' effort on browsing and selecting products to rate. Then, we randomly selected 8 unique products (half for female and half for male if applicable) from each subsub-category to comprise our experimental dataset, which finally includes 871 products.

The Big Five Inventory (BFI, 44 items) [10] was used to assess users' personalities (Big Five Personality Traits) on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The Big Five Inventory (BFI) is a self-report inventory consisting of short phrases with relatively accessible vocabulary. Among the 44 items, BFI possesses 16 pairs of items with opposite implications for personality (e.g., "is talkative" and "tend to be quiet"). The responses' consistency on each pair of items was adopted to measure their reliability. The acquisition process takes about 5 minutes on average.

4.2 Procedure

To assess users' personality and collect their ratings, we implemented an online experiment platform. Therefore, participants could easily participated in this study in any place and any time they feel comfortable. In this platform, an online procedure containing instructions, personality assessment questionnaire and rating systems was implemented so that participants could easily follow the task steps. Participants were first debriefed on the objective of the experiment and the upcoming tasks, and then fulfilled the required tasks by following the step-by-step instructions. Participants could exit the experiment anytime they want. The main user tasks contains three steps:

1. Fill in a background questionnaire, including gender, profession, age etc.
2. Accomplish the 44-item BFI personality assessment questionnaire.
3. Select and rate at least 30 items on a binary scale (like or dislike).

4.3 Participants

We recruited participants on the campus (e.g., in library, laboratories, cafeterias and metro station, or via mailing lists) or by announcing our advertisement on Facebook. All participants were also invited to provide an email address to be entered into a raffle for one gift voucher valued at 100 CHF. A total of 122 participants were recruited in our study. We examined their responses' reliability by checking their consistency on the 16 pairs of opposite items possessed by BFI. We filtered out those whose responses have more than 4 inconsistencies among these 16 pairs of items and we ended up having 86 users with reliable responses. The set of those participants is composed of 23 women (26.7%) and 63 men (73.3%). These participants are from 22 different countries (China, Korea, Switzerland, French, etc.), have different professions (student, research assistant, software engineer, company employee, administrative staff, entrepreneur, and so on.). Most of them (74 out of 86) are in the age group ranging from 21-30, 6 users are from age group 0-20, and 6 users are from 31- 40. 23 users have college education background, 58 users have a graduate school education background, and only 4 just graduated from high school and 1 is others. 56% of users (48) have used recommender systems before and among them, 18 users used recommender systems more than 3 times per week. 76% of users (65) have used e-commerce websites to purchase online and 14 users used them more than 3 times per week.

5 Dataset

In this study, we consider the following rating behavior variables.

1. *Number of rated items* (*NRI*). It measures how many items a user have rated, which sometimes closely deal with the accuracy of user preference modeling. For example, the number of items a user has rated directly affect the prediction accuracy of collaborative filtering recommender systems [20]. That is, as the number of ratings number increases, recommendation prediction accuracy can be improved. However, the effect is not monotone. After some point, the accuracy will tend towards stable. In this study, we are wondering which kind of users are following the introductions to only rate 30 items, and who will rate more.
2. *Percentage of positive ratings* (*PerPR*). To build users' preference models, we need to know not only their positive ratings ("like") so as to promote relevant items, but also their negative ratings ("dislike") to avoid irrelevant items. Therefore, it is interesting to investigate how many items will be rated to as "like" out of the whole set of rated items. It is related to how accurate and complete we could know about a user's preference. In this study, we further are interested in how personality would relate to such rating behavior.
3. *Category coverage* (*CatCoverage*). In our rating experiment platform, users are able to select items from one specific category by choosing it from a dropdown list including all of the first level (primary) categories. If a user selects "any category" (default value), shown items are randomly selected from all categories. We are interested in whether users with different personality characteristics will rate items covers a board range of categories, or a narrowed/focused list of categories. Therefore, we utilize the number of categories of rated items as a measure of category coverage. If one item belongs to more than one category, we count it once for each category. There are three levels of categories in our dataset, as described before. We calculate the category coverage for each level. They are indicated as CatCoverage-1 (for primary categories), CatCoverage-2 (for sub-categories) and CatCoverage-3 (for subsub-categories).
4. *Interest diversity* (*IntDiversity*). Different from category coverage, this variable measures the distribution of users' interests in each category. We are interested in whether a user has evenly distributed (diverse) interest in all covered categories, or he has a stronger interest on some specific categories compared to other covered

categories. To answer this question, we adopt Shannon index from information theory as a measure of interest diversity:

$$s = -\sum_{i \in C} f_i \mathrm{In} f_i \quad (1)$$

where C is the above set of categories and f_i is the fraction of items (out of the total number of rated items) that belong to ith category. Similar to the variable category coverage, we consider the interest diversity at three levels, IntDiversity-1, IntDiversity-2 and IntDiversity-3.

Together with the five personality traits, in our study, we take age, gender, and related experiences into account. Previous studies have shown that all of them have an effect on users' behaviors and preferences [12, 21]. Age is measured in three categories, ranging from 0 (0-20 years old), 1 (21-30) and 2 (31-40). Gender is classified into 0 (female) and 1 (male). Frequency of using a recommender system and frequency of doing online shopping are measured at four levels (0: Never, 1: 1-2 times, 2: 3-4 times, 3: over 5 times).

6 Results Analysis

6.1 Correlation with rating behavior variables

We first study the relationship between personality traits and user rating behavior variables, including the number of rated items (NRI), the percentage of positive ratings (PerPR), the category coverage of rated items (CatCoverage-1, CatCoverage-2, CatCoverage-3), and the interest diversity (IntDiversity-1, IntDiversity-2, IntDiversity-3). We calculate the Pearson product-moment correlation between rating behavior variables and personality traits, plus four additional independent attributes, namely age, gender, frequency of using recommender, and frequency of online shopping. The results are reported in **Table 2**.

Conscientiousness is negatively related to the number of rated items (β = -0.177, $p < 0.1$). That is reasonable since people with high Conscientiousness scores are more responsible for their required tasks. They would carefully select and rate products, and obey requirements strictly. Gender is negatively correlated with the number of rated items as well (β = -0.261, $p < 0.05$). It means that female participants rated more items than male participants did.

Those who are willing to give positive ratings tend to be high in Agreeableness (β = 0.179, $p < 0.1$). Agreeableness reflects a tendency to be sympathetic and cooperative. High Agreeableness people tend to be friendly and compassionate to maintain positive social relations, while those low on Agreeableness are less compromise and gullible. Agreeable individuals thus tend to give positive responses to behave friendly.

Personality trait Conscientiousness is found to negatively correlate with the category coverage (CatCoverage-2, b = -0.188, p < 0.1; CatCoverage-3, b = -0.201, p < 0.1).

Conscientiousness reflects the degree to which an individual is organized and scrupulous. Therefore, the covered categories are limited. Moreover, such negative correlation is stronger when the inner category level is considered. We don't find such correlation for the primary categories. In addition, it has been found that gender plays an important role in categorical coverage on all three levels (CatCoverage-1, b = -0.289, $p < 0.01$; CatCoverage-2, b = -0.247, $p < 0.05$; CatCoverage-3, b = -0.270, $p < 0.05$). Negative coefficients mean that female participants rate items within more categories.

Table 2. Correlation coefficients between big five personality traits and rating behavior variables. Statistically significant correlations are in bold and their p-values are expressed with *'s: $p < 0.01$(***), $p < 0.05$(**) and $p < 0.1$(*).

Personality Trait	NRI	PerPR	CatCoverage			IntDiversity		
			1	2	3	1	2	3
Openness	-0.028	0.135	-0.076	-0.021	-0.021	-0.140	-0.061	-0.046
Conscientiousness	**-0.177***	0.107	-0.138	**-0.188***	**-0.201***	-0.044	-0.146	**-0.187***
Extraversion	-0.141	0.059	-0.151	-0.145	-0.151	-0.083	-0.110	-0.122
Agreeableness	0.071	**0.179***	0.042	0.056	0.070	0.016	-0.001	0.042
Neuroticism	0.089	0.030	0.050	0.067	0.078	-0.055	0.034	0.065
Age	0.025	0.049	-0.041	-0.013	-0.029	-0.120	-0.040	-0.076
Gender	**-0.261****	-0.092	**-0.289*****	**-0.247****	**-0.270****	**-0.204***	**-0.192***	**-0.241****
Freq. of using recommender	0.192	0.100	0.057	0.079	0.117	-0.104	-0.034	0.018
Freq. of online shopping	0.136	-0.036	0.029	0.118	0.147	-0.081	0.033	0.102

Conscientiousness is moderately negatively correlated with interest diversity (IntDiversity-3, β= -0.187, $p < 0.1$). That is, high Conscientiousness individuals tend to have low interest diversity. That means most of their ratings focus on a narrowed range of categories. On the other hand, low Conscientiousness individuals tend to have a broad range of interested categories. Likewise, gender plays an important role in interest diversity on all three levels (IntDiversity-1, β= -0.204, $p < 0.1$; IntDiversity-2, β= -0.192, $p < 0.1$; IntDiversity-3, β= -0.241, $p < 0.05$). Negative coefficients mean that female participants rated items covering more diverse categories (interests evenly distributed) than male participants did.

No statistical significant relationships were found between the other independent variables, age and frequency of online shopping, and all the rating behavior variables.

6.2 Personality in different behavior groups

In this section, we look deeper inside at the long tail of the number of ratings. **Fig. 1** plots the distribution of the number of rated items. The x-axis represents the number of rated items, while the y-axis is the number of participants. As we could see from the distribution, most (28 out of 86) of participants only rated the required 30 items. We define this group as "obligation group", since users in area just accomplished the task they asked. Almost equivalent number of participants rated slightly (one or two) more items than the required amount, i.e., 31 items or 32 items. This group is defined as "inertia group", which is potentially influenced by the required number of ratings.

After that, few users rate more. We divide this long tail into two parts with equal number of participants, so that two groups are able to have enough participants to conduct meaningful statistical analysis. We get a reasonable cutting point, 50 ratings. The two groups are called "dispersion group" and "explosion group" respectively, based on the amount of ratings they gave. We are curious whether people from the four groups, representing different rating behavior patterns, vary in personality.

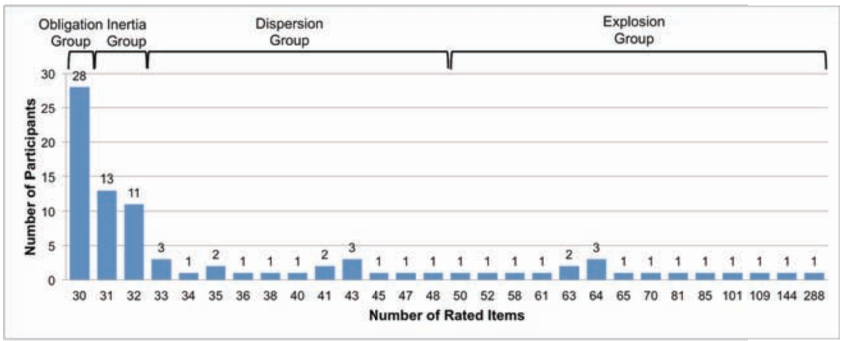


Fig. 1. Distributions of the number of ratings and the division of four behavior pattern groups.

We conducted one-way ANOVA with rating groups as IVs and personality trait scores as DVs, followed by post-hoc pairwise comparisons (Bonferroni) to identify how the four groups of personality characteristics varied from one another. The average scores of each personality traits in the four rating groups are shown in **Fig. 2**.

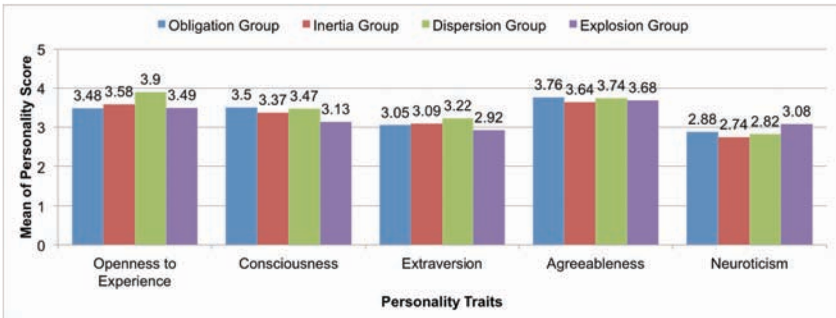


Fig. 2. Mean of personality scores in the four groups.

The ANOVA results indicate significant differences in the personality trait Openness to Experience ($F(3, 82) = 3.171$, $p = 0.029$). Pairwise comparison results show that users in dispersion group scored significantly higher on Openness to Experience (mean: 3.90, SD: 0.33) than those in other three groups, obligation group (mean: 3.48, SD: 0.46; $t = 3.541$, $p = 0.001$), inertia group (mean: 3.58, SD: 0.48; $t = 2.510$, $p = 0.016$), and explosion group (mean: 3.49, SD: 0.58; $t = 2.514$, $p = 0.019$).

Even though there is no statistically significant difference on personality Consciousness among the four groups, we found that users in explosion group have significant higher score than those in obligation group (mean: 3.13, SD: 0.53 vs. mean: 3.50, SD: 0.65 respectively; t = 2.109, p = 0.041).

With regard to other three personality traits, we didn't find statistically significant differences among groups and between pairs.

7 Discussion

Rating is a major way for users to explicitly express their preferences and opinions. It is critical to understand the nature of rating behaviors and which factors will influence these behaviors. In this study, we investigated the relations between personality and user rating behaviors. Our results have both theoretical and practical implications.

Theoretical Implications. Our results show that low Conscientiousness individuals tend to rate more items, while those with high Conscientiousness scores tend to only rate the required number of items. Similarly, Openness to Experience also affects the number of items a user will rate. However, they are more likely to rate more items in a certain range, probably in order to satisfy their curiosity. Above that boundary, they will stop rating, while low Consciousness individuals will keep rating more. This finding lets us to rethink the validity of ratings a user gives and how to find out those valid ratings. It might be an interesting research in the field of user modeling.

Previous research shows that Agreeableness is positively correlated with the number of friends, groups and "likes" [19]. Consistently, our results show that individuals high on Agreeableness tend to give more positive ratings. It implies that it will be difficult for us to know the actual preferences or opinions of users with high Agreeableness. Consequently, there exists a risk to employ ratings to infer their interests due to the compromised ratings.

Conscientiousness has a negative influence on category coverage and interest diversity. However, those with high Conscientiousness tend to rate items in a limited number of categories and their ratings are likely to focus on certain categories. Diversity is a research top of concern in the realm of information retrieval and recommender systems. It seems that it is much easier to build a diverse profile for an individual with low scores on Conscientiousness compared to those with high scores. Considering the validity issue of ratings, whether such diverse profile will benefit the personalization process is still unknown. Another research question is whether users with high Conscientiousness scores really like a narrowed range of items and how to assist them to rate more diversely.

Gender is another factor that shows a statistically significant correlation with the number of ratings, category coverage and interest diversity variables. However, we didn't find significant correlations between other independent variables (i.e., age, previous experiences on recommender and e-commerce) in our current experiment setting. More exploratory and in-depth experiments are needed.

Theoretical Implications. It is valuable to realize that personality makes an effect on user rating behaviors. It suggests that when intelligent systems, such as social media websites, recommender systems and e-commerce retailers, employ rating data to model users, it is critical to take personality's influences into account. Furthermore, since ratings directly affect the accuracy of inferred user preferences, practitioners and designers can consider designing personalized interfaces to get more useful rating information. For example, Agreeable people are likely to give positive ratings. The interface shown to them could try to motivate them to give true opinions. On the other hand, when practitioners are evaluating their systems, they should avoid those evaluators with high scores on Agreeableness. Gender seems a mediator with strong influence and it is easy to obtain. Therefore, it is necessary to consider this factor in building personalized intelligent systems.

8 Conclusion

We investigated how personality influences users' rating behaviors by an online survey. The correlation analysis results show that Conscientiousness is negatively correlated with the number of total ratings, category coverage and interest diversity. Individuals high on Agreeableness tend to give more positive ratings. Gender plays a crucial role on all rating behavior variables except percentage of positive ratings. In addition, we found that users high on Openness tend to rate more items than required, while low Conscientiousness is a critical factor which provoke users to rate items in an explosive way. The current study was conducted in a small sample size and most participants were students and in the age range of 21-30. In the follow-up study, we plan to continue this study in a platform with more diverse subjects, such as Amazon Mechanical Turk, to obtain more participants with high diversity, with the goal of validating our current findings.

Acknowledgements. We thank the EPFL for sponsoring the reported research work. We are grateful to all participants for their patience and time.

References

1. Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlates of music preferences. J Pers Soc Psychol 84, 1236-1256 (2003)
2. Selfhout, M., Burk, W., Branje, S., Denissen, J., Van Aken, M., Meeus, W.: Emerging Late Adolescent Friendship Networks and Big Five Personality Traits: A Social Network Approach. Journal of Personality 78, 509-538 (2010)
3. Barrick, M.R., Mount, M.K.: The Big Five Pesonality Dimensions and Job Performance: A Meta-Analysis. Personnel Psychology 44, 1-26 (1991)
4. Kelly, E.L., Conley, J.J.: Personality and compatibility: a prospective analysis of marital stability and marital satisfaction. Journal of Personality and Social Psychology 52, 27 (1987)
5. Zhong, B., Hardin, M., Sun, T.: Less effortful thinking leads to more social networking? The associations between the use of social network sites and personality traits. Computers in Human Behavior 27, 1265-1271 (2011)

6. Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., Crowcroft, J.: The personality of popular facebook users. Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, pp. 955-964. ACM, Seattle, Washington, USA (2012)
7. Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom), pp. 180-185 (2011)
8. Amatriain, X., Pujol, J.M., Oliver, N.: I Like It... I Like It Not: Evaluating User Ratings Noise in Recommender Systems. Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization: formerly UM and AH, pp. 247-258. Springer-Verlag, Trento, Italy (2009)
9. Costa, P.T., Jr., McCrae, R.R.: NEO PI-R Professional Manual. Psychological Assessment Resources, Odessa, FL (1992)
10. John, O., Srivastava, S.: The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin, L., John, O. (eds.) Handbook of Personality: Theory and Research, pp. 102-138. Guilford Press (1999)
11. John, O.P., Robins, R.W., Pervin, L.A.: Handbook of Personality: Theory and Research (3rd Edition). Guilford Press, New York (2008)
12. Kraaykamp, G., Eijck, K.: Personality, media preferences, and cultural participation. Personality and Individual Differences 38, 1675-1688 (2005)
13. Ganzeboom, H.B.G.: Explaining differential participation in high-cultural activities: A confrontation of information-processing and status-seeking theories. In: Raub, W. (ed.) Theoretical models and empirical analyses: Contributions to the explanations of individual actions and collective phenomina. E.S. Publications, Utrecht (1982)
14. Zuckerman, M., Ulrich, R.S., McLaughlin, J.: Sensation seeking and reactions to nature paintings. Personality and Individual Differences 15, 563-576 (1993)
15. Kosinski, M., Stillwell, D., Kohli, P., Bachrach, Y., Graepel, T.: Personality and Website Choice. ACM Conference on Web Sciences, (2012)
16. Rosen, P., Kluemper, D.H.: The impact of the big five personality traits on the acceptance of social networking website. AMCIS 2008 Proceedings 274 (2008)
17. Schrammel, J., Köffel, C., Tscheligi, M.: Personality traits, usage patterns and information disclosure in online communities. Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology, pp. 169-174. British Computer Society, Cambridge, United Kingdom (2009)
18. Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. CHI '11 Extended Abstracts on Human Factors in Computing Systems, pp. 253-262. ACM, Vancouver, BC, Canada (2011)
19. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., Stillwell, D.: Personality and patterns of facebook usage. ACM Conference on Web Sciences, (2012)
20. Drenner, S., Sen, S., Terveen, L.: Crafting the initial user experience to achieve community goals. Proceedings of the 2008 ACM conference on Recommender systems, pp. 187-194. ACM, Lausanne, Switzerland (2008)
21. Kidwell, B., Jewell, R.D.: The influence of past behavior on behavioral intent: An information-processing explanation. Psychology and Marketing 25, 1151-1166 (2008)

Design and development of an empirical smiley-based affective instrument

Adam Moore[1], Christina M. Steiner[2] and Owen Conlan[1],

[1] KDEG, School of Computer Science & Statistics, Trinity College, Dublin
{mooread;oconlan}@sccs.tcd.ie
[2] Knowledge Technologies Institute, Graz University of Technology, Austria
christina.steiner@tugraz.at

Abstract. Smileys (also known as Emoticons or Emoji) are regularly used to convey emotion within internet communications, especially in text-centric media such as email and instant messaging. Herein we describe an approach to utilise these well known affect emblems to create an interactive affect indicator. The indicator displays a set of nine smileys and users are asked to select the smiley that most closely reflects their current emotional state. In order to validate the design, a small survey was conducted. Responses were analysed and the indicator updated to reflect the results. A larger survey was then performed. An initial analysis of this online survey was conducted to try and understand the emotional content embedded within users' understanding of the smileys they use. Nearly 1000 responses to the survey were collected. This report focuses on general results and then briefly examines if there are any differences based on gender, both at a quite coarse scale.

Keywords: empirical model; affective indicator; web services; affect; smileys; emoticons; emoji.

1. Introduction

Emotions play an important role in learning, exerting effects on information processing and performance [1, 2]. The Smiley Based Affect Indicator (SBAI) was developed [3] in order to create a tool that allowed learners to express their emotional states while using an online experiential training simulation [4, 5]. Uniquely, the underlying model that is used to report affective states is derived from a large online survey. It is a RESTful web service [6] that can relatively easily be integrated with or called from any internet-connected system. It was initially created to fulfil two goals:

1. To allow affect to be indicated without the need to explain a lot of theory behind the instrument. That is, it is designed as a way of conveying affect through a readily understood method widely used throughout computer-literate society
2. To allow affect indicators to be recorded by a system that could be technically separate from a system without either appearing separate or being challenging to incorporate

These goals are linked to an overarching aim to allow continuous monitoring of learners' affect while interacting with a learning technology. The gathered affect reports may be fed back to the learning system for adaptation and personalization purposes. In addition, the affect reports inform evaluation of the learning technology by analyzing learners' reactions and activities during the learning episode

The work described herein reports the initial validation of this instrument and then goes on to outline changes made as a result of that work. It then covers the subsequent survey to collect empirical data in order to derive the underlying model. Finally, it reports on some findings from the initial model derivation.

1.1 Theory

The technical development and theoretical underpinnings of the SBAI are described elsewhere [2]. Briefly, in 1980, Russell formalised the theory that healthy humans do not suddenly experience emotional states but are in flux, moving through different emotions and different magnitudes of emotions [7]. He created the circumplex – a set of fundamental emotions arranged in a circular order. He concluded that human emotions move along the circumference of his circumplex – so that, for example, a human experiencing pleasure would never immediately experience depression, but would (no matter how fleetingly) also experience contentment and sleepiness (taking the shorter path). He then simplified the circumplex to consist of two dimensions: the valence of the affect (i.e., variation along a positive-negative or pleasure-displeasure dimension), and the intensity or arousal level of the affect (i.e., low vs. high intensity or activation).

Reductive facial expressions are a well established way to elicit self reporting, especially from young children [8]. For example, the FACES pain scale [9] uses a set of 7 pictures. Patients (typically 4-8 years old) are asked to point to the picture that represents the kind of pain they are feeling, starting with no pain on the extreme left, to the most pain ever on the extreme right.

Figure 1. The FACES pain scale: Reprinted from [9], with permission from Elsevier Science;

The idea behind the SBAI is similar, to use a reductive representation to prompt the user for a self-report. Bartneck [10], for example, has shown that human beings tend to find the emotional expression from a simple caricature such as smiley more distinct and easily recognizable, in comparison to a real human face. Product design also uses pictorial scales, such as the PrEmo system by Desmet *et al* [e.g. 11].

There are currently two other emoticon-based affective indicators. One, which is embedded within Crystal Island [12] is a vertical list of seven emoticons with single word sense labels is presented to the learner at 9 minute intervals. However, this instrument is tightly integrated into the larger learning system and it's triggering is not within the control of the instructional designer. The second is the Mirror project[1] MoodMap widget (also described in [3]). This prompts separately for Valence and Activity values using two series of icons. It requires a more nuanced approach to affect self-awareness and reporting, with some explanation and training required to explain what is sought for each icon set.

Affective agents attempt to leverage on the inverse of this, providing affect cues to the learner, for example, Okonowo's emotional pedagogical agent called "Smiley" [13] that adapts his expression in accordance to user behaviour or the embodied agent with MAUI [14]. A good overview of the field of embodied affective agents can be found in Beale's overview [15].

1.2 Understanding Emoticons

Emoticons have been in use on the internet for over 30 years, with the first recorded usage of them within an email recorded in 1982 (James.Morris at CMU-10A. "Notes – Communications Breakthrough" [16]). They are primarily used to add emotional notation to textual communication [17]. As computer-based text has become richer, emoticons have evolved from simple combinations of punctuation symbols to pictorial representations, often referred to as Emoji. These Emoji first emerged in Instant Messaging clients (such as MSN Messenger) but have become widely adopted, with many services now providing a wide range of Emoji libraries of varying styles to choose from (*cf*. Adium emoticon sets[2]).

2. Methods

Two surveys were conducted, a first, closed survey to provide a pilot validation of the SBAI, and then a subsequent, open survey to collect data with which to build a model of affect values embodied by the SBAI.

2.1 Survey Design

Both surveys were deployed on LimeSurvey [18]. Each survey consisted of three parts:

1. An introduction, instructions and demographic information collection
2. A page presenting a smiley, then two nine point Likert scales to record valence (-1 through 0 to +1 for positive or negative embodiment) and activity (0 to 1 for strength of affect signal). A text box to have a single word that the respondent felt best described the smiley followed this. This was repeated nine times, one for each smiley used in the SBAI.
3. A conclusion page, with free text area for comments

[1] http://www.mirror-project.eu/
[2] http://www.adiumxtras.com/index.php?a=search&cat_id=2

The smileys were arranged in an order the authors perceived as most negative to most positive. This is the same arrangement (reading left to right and top to bottom) as on the default SBAI grid, as shown in Figure 2 (other arrangements are available from the web service).

Whilst this constructive ordering of the smileys in the survey may provide some first order skewing of the results, it was felt that was an appropriate sacrifice to make rather than have a random sequence of presentation, as, at the time of the survey design, it was not possible to anticipate the number of those taking the survey. If the sample size had been a small number, it was felt that the ordering effect would be a major confounder in any analysis.

Figure 2 The pilot rendering of the Smiley Based Affect Indicator – a 3x3 grid ordered from most negative (top left) to most positive (bottom right). Note this is the original version (*cf.* **Figure 4**)

2.2 Validation Cohort

The validation cohort was recruited by an email to the researcher's colleagues and a single tweet, both containing a link to the survey. 22 replies were received in 10 days. Survey results were analysed using MS Excel 2011 on a Macintosh. 3 responses were excluded from the analysis, as they were incomplete. Of the 19 responses analysed, 5 were from females, 14 from males. Ages ranged from 22 to 48 with the average being 33.8 (SD 7.5). All had experience of further education.

2.3 Model Cohort

The model cohort was recruited by email (including one sent to every member of Trinity College, Dublin) and several tweets, both methods containing a short outline of the goals of the survey and then a link to the survey itself. The link was via a URL shortening service (http://bitly.com) that provided some analytics for the responses, in addition to that provided by the survey itself.

996 complete replies were received in 2 months. The cohort was composed of 285 women, 700 men and 11 respondents preferred not to say. Reported ages ranged from 15 to 103, with an average of 26.7 (SD 10.4). Analytics point to a large number of

responses to have been made in answer to the mailing to Trinity College; so cultural referents are skewed as a result. For example, nearly 70% of respondents give their country of birth as Ireland, and over 90% give Ireland as their country of current residence.

3. Results (Validation)

3.1 Valence vs Arousal

The average and standard deviation for each of the nine smileys were calculated. A graph of the Smileys plotting Valence vs Arousal is show in Figure 3 below:

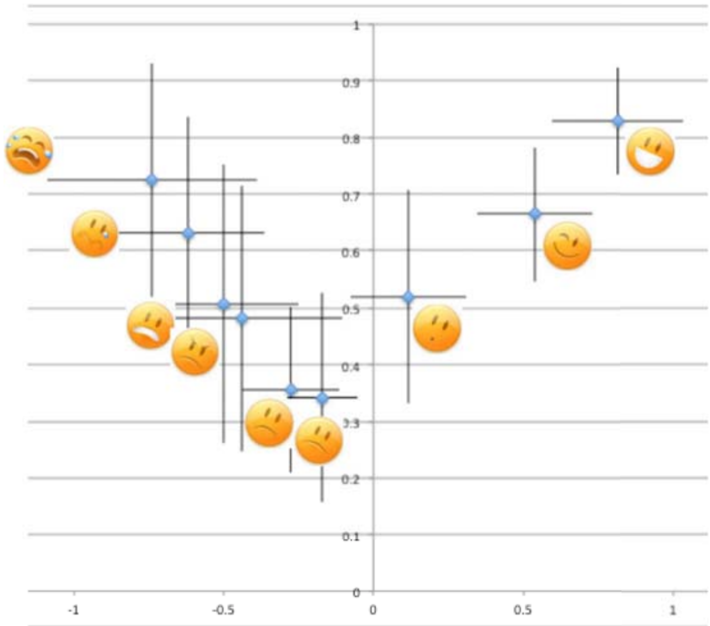


Figure 3. Smileys plotted on Valence vs Arousal with standard deviations displayed.

3.2 Smiley Sense Words

A table of each of the unique words entered for each smiley is shown below in **Table 1** (that is, redundancy only has been removed). It was soon discovered that the instruction to use a single word had not been obeyed in all cases – these compound sense phrases were recorded anyway. Sense stem words, produced by reducing the previous list to the base-level emotive word (for example, sadness, sorrow, tears, misery are all stemmed to "sad") are displayed in **Table 2**.

Table 1. Unique sense word / phrases from the validation survey per smiley.

Smiley	Unique sense words / phrases
1	*deep sadness, crying, sorrow, tears, sadness, hopelessness, Distraught, Upset, Happy but worked hard, wah, tantrum, sad, Despair, misery, desperation, distraught*
2	*undecided, sad, thrill, sorrow, unhappiness, Upset, a bit sad, sob, tearing up, hurt, ridiculus, sniff, sadness*
3	*disappointed, accident, embarrassment, angry, sadness, Anxious, stressed, grimace, worry, tense, Awkward, riduculus, afraid, disgusted, annoyed*
4	*unhappy, annoyed, anger, angry - upset, Angry, more angry, frown*
5	*wondering, unimpressed, query, depressed, disappointed, disappointment, Unhappy, being quiet, dismay, concerned, uncertain, perturbed, Bemused, perplexed, indifferent, sad, disappointed, down*
6	*neutral, unhappy, query, sad, disappointed, sadness, slightly not good, dismay, uncertain, perturbed, Peeved, Indifferent*
7	*Interested, surprise, jolt, surprised, amazement, concern*
8	*encouragement, smug, agreement, funny, happy, complicity, Understanding, joking, "OK, sounds good", wink, cheeky, clever, conviviality*
9	*very happy, beaming, happiness, happy, jubilant, cheery, smile, Delighted, agreement, laugh, excited*

Table 2. Stemmed sense words from the validation survey per smiley.

Smiley	Stemmed Unique sense words / phrases
1	*sadness, upset, anger, despair, exhausted*[3]
2	*undecided, sad, thrill, upset, hurt, ridiculous*
3	*disappointed, accident, embarrassment, angry, sad, anxious, worry, awkward, ridiculous, afraid, disgusted, annoyed*
4	*unhappy, annoyed, anger, confused*
5	*wondering, unimpressed, query, disappointed, unhappy, being quiet, dismay, concerned, indifferent, sad*
6	*neutral, unhappy, query, sad, disappointed, slightly not good, dismay, uncertain, perturbed, peeved, indifferent*
7	*interested, surprised, amazement, concern*
8	*encouragement, smug, agreement, funny, happy, complicity, understanding, joking, "OK, sounds good", wink, cheeky, clever, coniviality*

[3] Stemmed from "happy but worked hard"

9 	*happy, delighted, agreement, laugh, excited*

4. Discussion of Validation

The value of the word reports shows up when comparing smiley pairs 3 & 4 with 5 & 6. Although these two pairs show marked overlap in averages (especially when standard deviation is taken into account), there is a marked difference in the word senses attached to 3 & 4. However, there is little difference in the word senses attached to 5 & 6, stressing the similarity of the images, as perceived by the validation cohort.

In order to represent the utility to which the SBAI is put, the emoticons were deliberately displayed consistently in a certain order for the survey. Perhaps if the smileys had been presented in a random order to each participant, there may have been different sense words or valence/arousal values entered. However, that independent appreciation of the emoticons would not reflect the purpose of the instrument, and may, itself, lead to confounding factors if the random order particularly skewed to one presentation order in a small validation sample.

Due to the similarity in both Valence/Arousal values and word reports, Smiley 6 was replaced with a more neutral image, as shown in Figure 4. This updated version of SBAI was used for the survey aiming at building a model of affect values (cf. section 5 below).

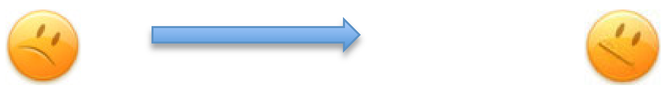


Figure 4. Smiley 6 was replaced with a more neutral emoticon**.**

5. Results (Model)

5.1 Valence vs Arousal

The average and standard deviation for each of the nine smileys were calculated. A graph of the Smileys plotting Valence vs Arousal is show in Figure 5.

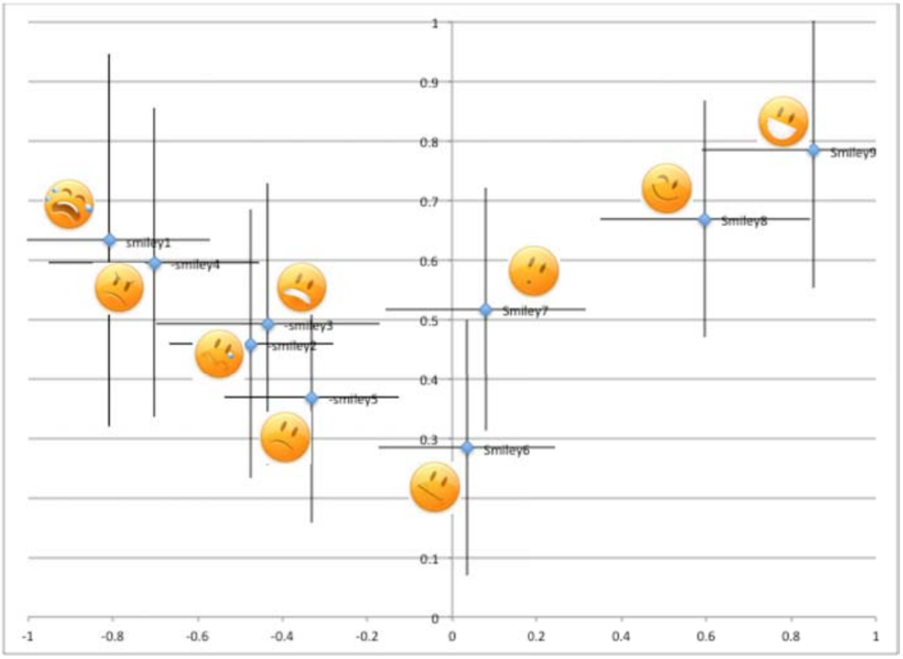
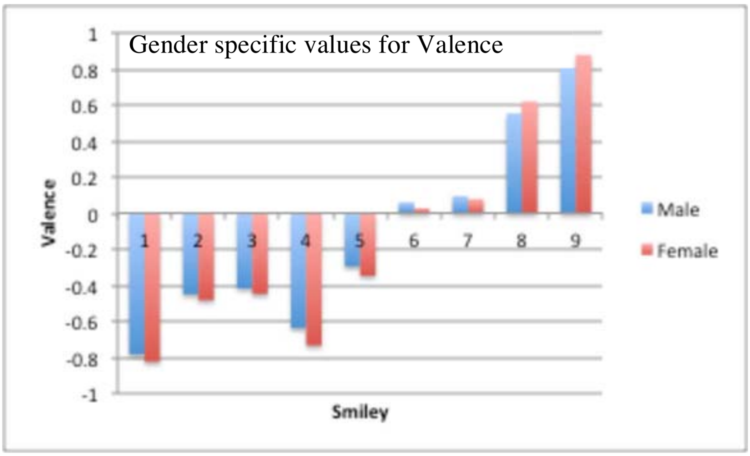


Figure 5 Plot of average valence against activity for the 9 smileys – bars are standard deviation.

The results were then categorized by gender and averages again calculated. Graphs for valence and activity are displayed in below:

a)


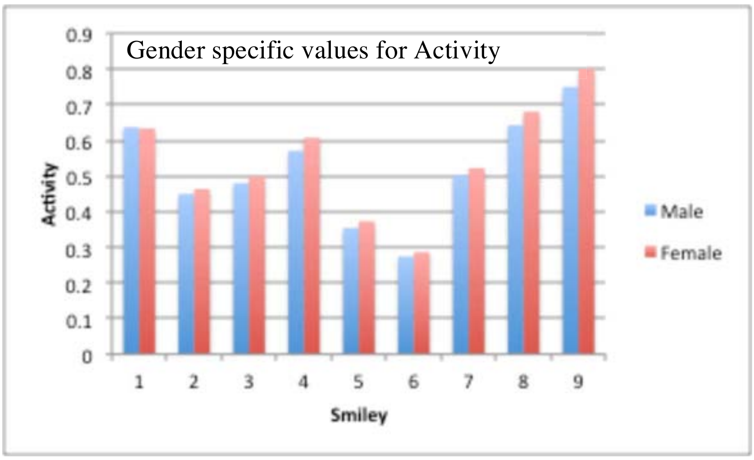
b)


Figure 6 Plot of average valence (a) and activity (b) for the 9 smileys – by gender.

Again, respondents were asked to enter one word that they felt best expressed what the smiley was trying to represent. A first interesting result is that many felt that this was not a simple task, and indeed, some comments specifically addressed this issue, either to note that the task was difficult, or that they had ignored the instruction in order to better express themselves. For each smiley the number of unique words or phrases entered was analyzed.

Along with the top occurring word, and these can be seen in **Table 3** below:

Table 3. Number Of Unique Sense Words/Phrases and top occurring one for each smiley

Smiley	*Number of Unique Sense Word/phrases*	*Top Occurring Sense Word*
1 	*210*	*Anguish*
2 	*204*	*Sadness*
3 	*258*	*Anxious*
4 	*128*	*Anger*
5 	*241*	*Unsure*
6 	*309*	*Boredom*
7 	*178*	*Surprise*
8 	*271*	*Happy*
9 	*170*	*Ecstatic*

6. Discussion of Model Results

An initial analysis of the survey responses showed that there were real differences in perception of the underlying embodiment for each of the smileys, such that they could usefully be used as a self-reporting affect instrument with confidence.

There were a variety of sense words associated with each smiley, but the variance in the number of unique words indicates that some are more singular in what they embody than others. Conspicuously, smiley 4 () seemed almost universally understood as 'Anger'.

There has been much discussion about the 'basic' emotions – for example, Ekman created a list of 15, based on examinations of their cross-cultural physical embodiment [19]. It is notable that all of his list occur somewhere in the reported sense words attached to the emoticons.

Whilst the emoticons selected represent a good range of emotions, there is currently no referent for a high valence, but low activity affective state, or their inverse (i.e. extremes of valence with low activity). Of course, this is to be expected, as a highly emotionally aroused state would generally be strongly correlated to a strong affective response (valence). However, it is important to note this when the instrument is being utilized.

As a first step to a deeper examination of the data, a further analysis was performed to examine the values assigned based on gender. There were no significant differences between genders in expectations for the representation for smileys, but empirically males seem to assign slightly lower values for both valence and activity than females.

7. Conclusions and Future Work

The SBAI is a novel, RESTful web service to provide model-based affect reporting. Uniquely, that model is currently based on a survey of nearly 1000 respondents, allowing for a degree of adaptation of the base responses, depending on the profile of the user. This allows for a powerful link to be made between technologies using the SBAI and the underlying model, driving affect reporting that reflects core characteristics of the personalization. With more data collected, the number of characteristics to which the model can respond increases.

A deeper analysis of the data to investigate culture and co-factors is ongoing, however, more data will need to be collected in order to create a more heterogeneous cohort.

Several times, context was mentioned as a contributing factor to how the respondent might understand the underlying meaning of a smiley. A new experimental design in order to investigate this aspect is currently being developed.

Recently, the SBAI was deployed as part of a large cohort study on providing Affective Metacognitive Scaffolding using the ETU RolePlay Simulator – its embedding can be seen in Figure 7:

Figure 7 SBAI inserted into ETU RolePlay Simulator

It is expected that the deployment of the SBAI will allow for the development of affect-related aspects of personalized services, for example, allowing those authoring TEL material to create material that reflects the mood of the learner, perhaps delivering additional encouraging motivation and engagement during periods of negative valence and increasing the prevalence of material that generates a positive response. Monitoring affect may also provide information on the quality and tone of the materials within personalised services, allowing the curation of a corpus to ensure that material that reports negative affect states has corresponding supportive material.

Requests for access to the anonymized data for analysis or to base affect signal instruments on should be made to the corresponding author. You can take the survey yourself at: http://bit.ly/smILEY.

Acknowledgments. The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement no 257831 (ImREAL project). The smileys are provided free as creditware by YellowIcon and were downloaded from IconEasy http://www.iconeasy.com/iconset/emotion-orange-icons/

References

1. Schutz, P.A., Pekrun, R.:: Emotion in Education. Academic Press, (2007)
2. Pekrun, R., Goetz, T., Titz, W., Perry, R.P.:: Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. Educational Psychologist 37, 91–105 (2002)
3. Moore, A., Ascolese, A.:: Explicit Tools for Affective Indicators. In: EC-TEL 2011 Workshop on Augmenting the Learning Experience with Collaborative Reflection, Palermo, Sicily, Italy (2011)
4. Moore, A. et al.:: Affective Metacognitive Scaffolding for the Enhancement of Experiential Training for Adult Learners. In: EC-TEL 2011 1st Workshop on Awareness and Reflection in Learning Networks (ARNets11), Palermo, Sicily, Italy (2011)
5. Moore, A., Macarthur, V., Conlan, O.:: Core Aspects of Affective Metacognitive User Models. In: L. Ardissono, T. Kuflik (eds.) Advances in User Modeling: selected papers from UMAP 2011 workshops. Lecture Notes in Computer Science 7138. pp. 47–59. Springer-Verlag GmbH, (2012)
6. Fielding, R.T.:: Architectural Styles and the Design of Network-based Software Architectures, PhD-thesis, Citeseer (2000)
7. Russell, J.A.:: A circumplex model of affect. Journal of Personality and Social Psychology 39, 1161–1178 (1980)
8. Chambers, C.T. et al.:: A comparison of faces scales for the measurement of pediatric pain: children's and parents' ratings. Pain 83, 25–35 (1999)
9. Bieri, D. et al.:: The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties. Pain 41, 139–150 (1990)
10. Bartneck, C.:: Affective expressions of machines. CHI 01 extended abstracts on Human factors in computing systems CHI 01 189 (2001)
11. Desmet, P.M.A., Hekkert, P.:: The Basis of Product Emotions. Technology withproduc, 60–68 (2002)
12. Rowe, J. et al.:: CRYSTAL ISLAND : A Narrative-Centered Learning Environment for Eighth Grade Microbiology. In: Education, pp. 11–20. (2009)
13. Okonkwo, C., Vassileva, J.:: Affective pedagogical agents and user persuasion. In: Proceedings of the 9th International Conference on Human- Computer Interaction, pp. 397–401. New Orleans (2001)
14. Lisetti, C.L., Nasoz, F.:: MAUI: a multimodal affective user interface. In: Proceedings of the tenth ACM international conference on Multimedia, pp. 161–170. ACM, (2002)
15. Beale, R., Creed, C.:: Affective interaction: How emotional agents affect users. International Journal of Human-Computer Studies 67, 755–776 (2009)
16. :-) Turns 25. CNN (2007)http://web.archive.org/web/20071012051803/http://www.cnn.com/2007/TECH/09/18/emoticon.anniversary.ap/index.html
17. Neviarouskaya, A., Prendinger, H., Ishizuka, M.:: Analysis of affect expressed through the evolving language of online communication. Proceedings of IUI 07 07pp, 278–281 (2007)
18. Carsten Schmitz.:: LimeSurvey: An Open Source survey tool. (2012)
19. Ekman, P.:: Basic emotions. In: T. Dalgleish, M. Power (eds.) Handbook of cognition and emotion. vol. 98, pp. 45–60. John Wiley & Sons, (1999)

Personality and Social Context: Impact on Emotion Induction from Movies

Ante Odić[1], Marko Tkalčič[2], Jurij F. Tasič[1], and Andrej Košir[1]

[1] University of Ljubljana, Faculty of Electrical Engineering, Tržaška cesta 25, Ljubljana, Slovenia,
ante.odic@ldos.fe.uni-lj.si,
[2] Johannes Kepler University Department for Computational Perception,Linz,Austria, Altenberger Str. 69, Linz Austria

Abstract. In this paper we describe our preliminary work on understanding the impact of personality on the emotion induction in different social circumstances during the consumption of movies, for the purpose of the context-aware recommender system for movies. The purpose of this study is to answer two research questions: is there a difference in emotion induction when users are alone as opposed to when they are with company during watching the movie, and do different personality profiles influence the emotion induction when users are alone as opposed to when they are with company during watching the movie? We have used the (LDOS-CoMoDa) dataset which contains ratings and associated contextual information for the consumed movies, as well as Big Five personality profiles of the users. The results showed that there is an influence of social context on emotion induction, and that personality factors have to be taken into consideration since for the different groups of users, based on the personality factors, the emotion induction was influenced differently.

Keywords: context-aware, recommender systems, user modeling, personality

1 Introduction

Employing contextual information in personalized services, such as recommender systems (RS), has been a popular research topic over the past decade. Contextual information is defined as information that can be used to describe the situation and the environment of the entities involved in such systems [3], and was proved to improve the recommendation procedure in context-aware recommender systems (CARS) [1, 2, 9], as well as other personalized services [12]. In our previous work [8] we showed that emotional context is relevant and by employing it we were able to significantly improve the quality of rating prediction in RS. In addition, the authors in [10, 11], have successfully used personality and emotions in RS for images.

In this paper we describe our preliminary work on understanding the impact of personality on emotion induction in different social circumstances, during the consumption of movies, for the purpose of CARS for movies.

1.1 Motivation and Goal

According to [13], personality refers to the enduring patterns of thought, feeling, motivation and behavior that are expressed in different circumstances. The authors in [7] state that the Big Five factor model of personality is a hierarchical organization of personality traits in terms of five basic dimensions: extraversion, agreeableness, conscientiousness, neuroticism and openness to experiences. The description of the five factors and their sub factors was provided in [5]. According to [14] all five factors influence feelings and emotional behavior.

Since we have shown the importance of emotion context in our previous work [8], we were interested in inspecting the impact of different users' personality profiles on the emotion induction. In addition, in the same study we have observed that the social context was not relevant and did not improve the rating prediction by our models. Nevertheless, we were still interested in the impact of the social context on the emotion induction. If the social context does impact the emotion induction and consequently the emotion context, such information could still be valuable for modeling users' behavior.

Therefore, the purpose of this study is to answer the following research questions: (i) Is there a difference in emotion induction when users are alone as opposed to when they are with company while watching movies? (ii) Do different personality profiles influence the emotion induction when users are alone as opposed to when they are with company while watching movies?

2 Materials and Methods

In this section we describe the dataset and the methods used in this study.

2.1 Dataset

For the purposes of this work we have used the *Context Movie Dataset* (LDOS-CoMoDa), that we have acquired in our previous work [8].

We have created an online application for rating movies which users are using in order to track the movies they watched and obtain the recommendations (www.ldos.si/recommender.html). Users are instructed to log into the system after watching a movie, enter a rating for a movie and fill in a simple questionnaire created to explicitly acquire the contextual information describing the situation during the consumption. In addition, we have asked our users to complete the standardized Big-Five questionnaire to acquire their personality profiles for research purposes. The Big Five questionnaire used consisted of 50 questions answered by selecting an answer form the five point Likert scale. As a result, for each user that had completed a questionnaire, we have obtained scores for five personality factors *extraversion*, *agreeableness*, *conscientiousness*, *neuroticism* and *openness to experiences*.

We have collected 2296 ratings from 121 users to 1232 items with associated contextual variables. Big Five profiles were collected from 78 users that were willing to participate. Additional information about our *Context Movies Database* (LDOS-CoMoDa) can be found in [6] and [8].

2.2 Preparing Data

In order to use the LDOS-CoMoDa dataset for this study we had to process the acquired rating data with associated contextual information.

First of all, we filtered out all the entries in the dataset from those users from which we have not acquired personality profile. As a result there were 1708 entries from 78 users left in the dataset.

Social context in LDOS-CoMoDa dataset is described by a categorical variable with categories: *alone*, *partner*, *friends*, *colleagues*, *parents*, *family* and *general public*. In this study we were interested in observing the differences in emotion induction when the user is alone from when the user is with company. Therefore, we regrouped the categories of the social contextual variable to only two categories: *alone* and *not alone*.

Emotional state context in the dataset is described by two categorical variables: *dominant emotional state experienced the most during watching a movie*, and *emotional state at the end of the movie*. Both variables have following categories: *sad*, *happy*, *fear*, *disgust*, *surprise*, *angry* and *neutral*. We have decided to observe the dominant emotional state experienced the most during watching the movie. Furthermore, since we were interested in observing the induction of emotions we have regrouped the categories in the following way: in the case of *sad*, *happy*, *fear*, *disgust*, *surprise* or *angry* we assume that there was an induced emotion, in the case of *neutral* we assume there was no (or at least much less) emotion induction during the consumption of a movie. Therefore we regroup the categories of the variable into two categories: *emotion* and *no emotion*.

Each personality factor holds a score from zero to 100. For example, for the *extraversion* factor a user with a score of 100 would be considered highly extroverted, while a user with a score of zero would be considered highly introverted. For each personality factor we have set a threshold at a score of 50 and have grouped users into two groups *high* (if the user's score is higher than or equal to 50) and *low* (if the user's score is lower than 50). For example, for extraversion we thus have two groups *high extraversion* and *low extraversion*. Consequently we have compared emotion induction between users that have scored high and low in each personality factor which resulted in ten "personality groups": *high extraversion*, *low extraversion*, *high agreeableness*, *low agreeableness*, etc. Note that we have observed the induction of emotions for each personality factor separately and leave the combinations of factors for future work.

2.3 Observing and Testing the Difference in Emotion Induction

For each personality group separately we have compiled contingency tables which show the interrelation between contextual variable describing the social state and contextual variable describing the dominant emotion during the movie consumption. Such table contains the numbers of occurrences of emotion induction in two different circumstances: user was alone and user was not alone. Table 1 shows an example of contingency table for the *agreeableness* groups.

Table 1. Contingency table example for *low agreeableness* and *high agreeableness* personality groups.

	Low agreeableness		High agreeableness	
	Alone	Company	Alone	Company
No emotion	30	89	244	230
Emotion	77	117	421	500

From the contingency tables, for each personality group, we have calculated the proportion of emotion induction in different social circumstances: when users were alone and when users were not alone. To test if the difference in the proportions is statistically significant we have used the z-test for proportions with significance level of 0.05.

3 Results

Results showed us that for *conscientiousness* and *neuroticism* there were no significant differences in any case, therefore, we leave the detailed results for those personality factors out of this paper.

First, from the contingency tables, we observed the difference in average proportion of emotion induction between *low* and *high* groups for personality groups. These are the average proportions of emotion occurrences, regardless of the social circumstances. Figure 1 shows the difference in proportion of emotion induction between *low* and *high* groups.

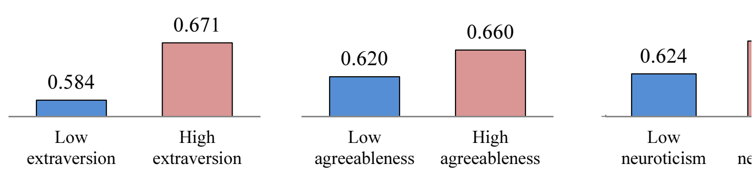


Fig. 1. Average proportion of emotion for personality groups.

Next we observed the difference in emotion induction proportion in different social circumstances. Figure 2 shows the proportions for low- and high- extraversion personality groups. Figure 3 shows the proportions for low- and high- agreeableness personality groups. Figure 4 shows the proportions for low- and high- neuroticism personality groups. Note that the low and high charts do not share the x-axis. Axes are scaled to best show the pattern of opposite impact of high and low groups on the emotion induction. Values of proportions are correctly stated at the top of the bars on the charts.

Results of the z-test for the statistical significance of the difference in proportions between social circumstances are shown in Table 2. In the personal-

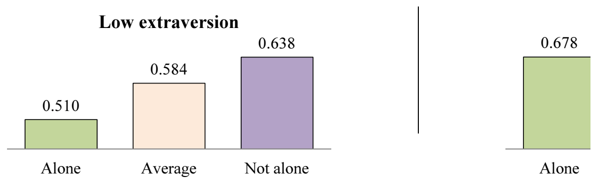


Fig. 2. Proportions of induced emotion in users when alone, on average and when not alone for the low and high extraversion groups.

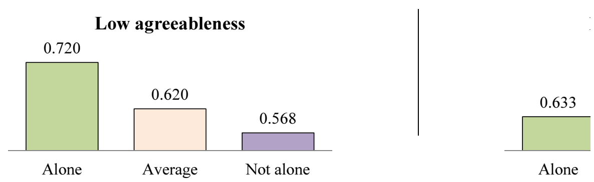


Fig. 3. Proportions of induced emotion in users when alone, on average and when not alone for the low and high agreeableness groups.

ity group column, groups for which the difference is statistically significant are marked with bold characters.

Table 2. Statistical significance results from the z-test for proportions for different personality groups (alone vs. not alone).

personality group	p-value (alone vs. not alone)
Low extroversion	0.015
High extroversion	0.623
Low agreeableness	0.009
High agreeableness	0.041
Low neuroticism	0.031
High neuroticism	0.366

4 Discussion

As it can be seen on Figure 1, the highest 8.9% difference in the average proportions between high and low groups are for the extraversion factor. For the agreeableness factor the difference is 4%, and for the neuroticism 5%. For each factor the *high* group had higher proportion of emotion induction than the *low* group.

Figures 2, 3 and 4 show an interesting pattern for extraversion, agreeableness and neuroticism factors. For each factor, users from the *low* group experienced emotion induction exactly in the opposite way that the users from the *high* group. For example, for the agreeableness factor, users with low agreeableness that consumed movies alone experienced emotions on more occasions than

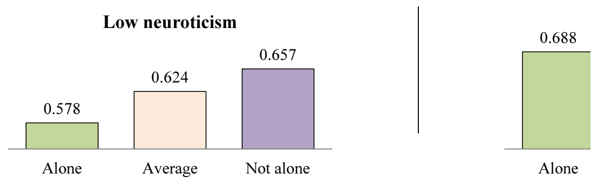


Fig. 4. Proportions of induced emotion in users when alone, on average and when not alone for the low and high neuroticism groups.

group's average, but on fewer occasions than group's average when not alone. The opposite effect was observed for the users from the high agreeableness group, who experienced emotions less when alone than when with the company of others.

Table 2 shows that in the case of the low extroversion, low agreeableness, high agreeableness and low neuroticism the difference in the proportions of the emotion induction in different social circumstances was statistically significant.

These results show that there is an influence of social context on emotion induction, however personality factors have to be taken into consideration since for the different personality groups the emotion induction was influenced differently.

5 Conclusion and Further Work

In this preliminary study we inspected the influence of social context and personality factors on emotion induction from movies. We have used the LDOS-CoMoDa dataset which contains rating data and the associated context from movie consumption. The dataset also contains Big Five personality profiles of 78 users. The results have shown that there is in fact a difference in proportion of emotion induction in different social circumstances for several personality factors. These differences were statistically significant in the cases of low extroversion, low agreeableness, high agreeableness and low neuroticism. It was also observed that users with low scores (< 50) experienced emotion induction exactly in the opposite way that the users with high scores (≥ 50) for the extroversion, agreeableness and neuroticism factors. We believe this to be an interesting results since it could lead to better understanding of the influences on the emotion induction.

For the future work we plan to inspect the detected effect further, and try to incorporate the observed behavior in CARS models. At this point we do not know why personality has the observed influence on emotion induction. For future work we would like to explain this fact for an additional insight on incorporating the observed behavior in CARS. In addition, we will try to use the observed effect in order to implicitly assess the users' personality traits from their behaviour in social and emotional contexts. Also, we believe that personality profiles and social context could be used to predict the emotional context, that was proved to improve the rating prediction [8].

6 Acknowledgement

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7 / 2007- 2013 through PHENICX project under grant agreement n° 601166. Additional funding was received by the Slovenian Research Agency ARRS by the grant number R-819.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (Jun 2005), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1423975
2. Baltrunas, L., Ludwig, B., Peer, S., Ricci, F.: Context relevance assessment and exploitation in mobile recommender systems. Personal and Ubiquitous Computing pp. 1–20 (Jun 2011), http://www.springerlink.com/index/10.1007/s00779-011-0417-x
3. Dey, A., Abowd, G.: Towards a better understanding of context and context-awareness. Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing pp. 304–307 (1999), http://portal.acm.org/citation.cfm?id=743843
4. Herlocker, J., Konstan, J., Terveen, L., Riedl, J.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 22(1), 5–53 (2004), http://portal.acm.org/citation.cfm?id=963772
5. Johnson, J.A.: Descriptions used in ipip-neo narrative report (2005)
6. Košir, A., Odić, A., Kunaver, M., Tkalcic, M., Tasic, J.: Database for contextual personalization. ELEKTROTEHNIŠKI VESTNIK 78(5), 270–274 (2011)
7. McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality 60(2), 175–215 (1992)
8. Odić, A., Tkalčič, M., Tasič, J.F., Košir, A.: Predicting and detecting the relevant contextual information in a movie-recommender system. Interacting with Computers 25(1), 74–90 (2013)
9. Odić, A., Tkalčič, M., Tasič, J.F., Košir, A.: Impact of the context relevancy on ratings prediction in a movie-recommender system. Automatika–Journal for Control, Measurement, Electronics, Computing and Communications 55(1) (2013)
10. Tkalčič, M., Burnik, U., Košir, A.: Using affective parameters in a content-based recommender system for images. User Modeling and User-Adapted Interaction 20(4), 279–311 (2010)
11. Tkalcic, M., Odic, A., Kosir, A., Tasic, J.: Affective labeling in a content-based recommender system for images (2013)
12. Toutain, F., Bouabdallah, A., Zemek, R., Daloz, C.: Interpersonal Context-Aware Communication Services. IEEE Communications Magazine (January), 68–74 (2011)
13. Westen, D.: Psychology, mind, brain &. Culture, New York: John Wiley & Sons, Inc (1996)
14. Yik, M.S., Russell, J.A., Ahn, C.k., Dols, J.M.F., Suzuki, N.: Relating the five-factor model of personality to a circumplex model of affect. In: The five-factor model of personality across cultures, pp. 79–104. Springer (2002)

The Role of Social Signals in Telecommunication: Experimental Design

Tomaz Vodlan[1], Marko Tkalcic[2], Andrej Kosir[2]

[1] Agila d.o.o., Ljubljana, Slovenia
tomaz.vodlan@gmail.com
[2] Faculty of Electrical Engineering, University of Ljubljana, Slovenia
{marko.tkalcic, andrej.kosir}@ldos.fe.uni-lj.si

Abstract. In this paper, we present the experimental design for the evaluation of the impact of social signal application on a user's decision making in the area of telecommunications. The aim of the design is to show that user's social signals are applicable feedbacks in conversational recommender systems. We use user satisfaction (with the system and content) evaluation criteria. During social interaction humans express social signals which provide quick feedbacks required by conversational recommender system. The experimental scenario is hands driven video-on-demand service with a conversational recommender system where the user selects among videos on screen. We limited our experimental scenario to the social signal of hesitation only. User is hesitating, when is faced with a variety of choices to make decisions (he is uncertain). The system adjusts the list of items to be recommended according to the extracted social signal {hesitation, no hesitation}.

Keywords: Human-Computer Interaction, Social Signals, Recommender System, Experimental Design

1 Introduction

The social signals (SS) have received much attention in recent years due to their additional natural information about human behavior which offers important benefits in human-computer interaction (HCI) [1]. Social signals similar to emotions are expressed with nonverbal behavioral cues (gestures, postures, etc.) and present human reactions to current social situations. From here on, the word system will be used as a synonym for video-on-demand system with recommender system. However, it is not clear how to utilize SS in telecommunication applications and that is the major reason why the most of the systems are socially ignorant. Based on our preliminary testing, social signal of hesitation is a frequently expressed signal when interacting with and selecting among multimedia items. It might provide additional information about how user selects one video on screen among others and not just information about which video is selected. Based on that, the system might recommend to user most suitable new videos. Recommendations of videos, provided by our system, are based on conversational recommender system.

We distinguished between two approaches in social and cognitive psychology perspectives of emotions: (i) emotion as individual experience and (ii) emotion as SS [2]. The theories of the first approach emphasize the relative significance of physiological changes, cognitive process, and the sensation and reaction where emotion emerges in an individual [2]. Instead of reflecting a person's inner feelings in emotion as SS approach, facial expressions are reflections of either real or imaginary interaction – no inner sensations need to play [2]. In our user to system interaction, emotion as SS, displays provide information about the user disposition and the situation as such.

The goals of this position paper are (i) to introduce an experimental design for the evaluation of the impact of SS in video-on-demand service, (ii) to show that social signals are applicable feedbacks in conversational recommender systems, and (iii) to list and discuss the identified potential flaws of the experiment in terms of a fair estimation of impact of SS (fair comparison of control and test group of users). However, the purpose of this paper is to describe experimental design without test results.

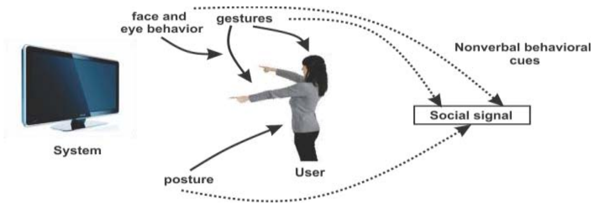


Fig. 1. Use of social signals in human-computer interaction

1.1 Motivation

The proposed methodology of our study is based on the use of social signals in human-computer interaction (Fig. 1). The system recognizes the user's social signals and uses them in interaction. Processing of social signals can be utilized in HCI in order to support a user's decision while passing through the user's interaction procedure. For humans, it is natural to produce social signals in several verbal and nonverbal ways. Consequently, the whole procedure is based on utilization of human social intelligence. We assume that social signals as additional information can improve the user experience and increase the efficiency level of a communication service, and that is why we should use these naturally produced signs by user.

1.2 The Role of Social Signals in Telecommunication

As was previously discussed, we will merge three domains in the context of our solution. We will present domains of human-computer interaction (HCI), social signal processing (SSP) and recommender systems (RS) below.

Human-Computer Interaction (HCI). Human-computer interaction (HCI) in its basic form involves the study, planning, and design of interactions between people (users) and computers [3]. We can divide HCI into two groups, simple and intelligent

HCI [4]. We are interested in intelligent interaction, where the computer understands the meaning of the message of the user, which is typically performed using speech and body gestures. Human – Centered Intelligent (HCI^2) [1] is one of the foremost challenges of computer science [4]. The domain of HCI^2 is bridging the gap between computer science and cognitive science. In the context of HCI^2, computers must have the ability to understand the meaning of the information expressed by a user and also the context of this information [4]. There are only few studies that use social signals in HCI. In [5], hesitant hand motion used by people is proposed as natural modality for a robot to communicate uncertainty in human-robot interaction.

Social Signal Processing (SSP). Social signal processing (SSP) [6] is the research domain that aims to understand social interactions through machine analysis of non-verbal behavior [7]. Social signals are initiated by the human body and present reactions to current social situations. They are expressed with nonverbal behavioral cues (gestures and postures, face and eye behavior, etc.). One of the most distinct social signals in this interaction is hesitation, which can be expressed with a facial expression, head movement, shoulder movement, etc. [8]. A review of the social signal processing research domain is given in [7], [9], and [10]. The goal of our research is the application of social signals that are inherent in our gestures, postures, facial expressions, and gaze behavior. There are not many applications that include social context. In [11] the spontaneous agreement and disagreement recognition approach is presented. The impact of mimicry on social interaction is shown in [12].

Social Signal of Hesitation. The social signal of hesitation belongs to a type of micro movement called microslip - nonverbal stutters during execution of low level action primitives [13]. Another psychological definition describes hesitation as elapsing time between the external or internal stimulation of an organism and his, her or its internal or external response [14]. Hesitation can be expressed through a facial expression, head movement, shoulder movement, prosody and special verbal markers like *eh* or *hm* [8].

Recommender System (RS). Recommender systems (RSs) are software tools and techniques that predict user preferences for the purpose of suggesting items to be of use to a user [15]. There are plenty of reasons for using the RS, but for our purposes the most important reason is increasing user satisfaction when using the system. In our case, conversational RS is used, where recommendations are generated based on natural language dialog between the user and system. However, in our video-on-demand service, RS is used for recommendations of various multimedia contents. Two functions of RS are implemented in a way to reflect the user's social signal of hesitation. If the user is hesitating, the function of get diverse multimedia items (videos) is used. If not, the function of get similar multimedia items is used. Further details are given in Subsection 4.4. RSs are directed towards users who do not have enough personal experience or priori knowledge about recommended items to make an autonomous decision [16]. Conversational RSs use natural language support, where the user and the system may query or provide information to the other partner [15]. The biggest challenges of this domain of RSs are how to design the effective dialogue strategy between user and system and what actions must be performed in the interaction between them [15].

This paper is organized as follows. Section 2 describes the problem statement of research domain with included hypothesis. Section 3 includes the description of the experiment with experimental scenario and description of application interfaces for recording. In Section 4 are described selected details of the test procedure and evaluation plan. Discussion about experiment is presented in Section 5. Finally, Section 6 presents our conclusions.

2 Problem Statement and Hypotheses

The problems addressed in this article are how to evaluate the impact of social signal on user's satisfaction in application in the area of telecommunications and to show that social signals are applicable feedbacks in conversational recommender systems. Based on study described in [5], we assumed that social signal of hesitation is distinct enough that can be extracted in human-computer interaction. It can be described with different types of cues used in [11] and [12]. The system presented here is used conversational RS on LDOS-CoMoDa, a contextual personalization dataset [17] and [18].

In the most basic form hesitation can be considered as a kind of uncertainty, when a user is faced with a variety of choices to make decisions. Nonverbal signs of the social signal of hesitation, which can be recognized from video, will be used for our purposes. Based on results of our prior test we extracted the most often applied signs of social signal of hesitation. The most often is SS expressed as facial expression and arm moving. We can describe the facial expressions with facial action coding system (FACS) [19]. We can use actions below; outer brow raiser (2), upper lip raiser (10), dimper (14), chin raiser (17), lip suck (28), blink (45), head tilt left (55), head tilt right (56), eyes up (63) and eyes down (64). Social signal of hesitation can be presented in various combinations (we use action unit number whose identify the action): 2+10, 2+17, 28, 45, 17+55+56, 14+55+56 and 14+63+64. Social signal is also presented with shoulder movements up and down, whole torso moving, arm moving up and down on side of the body (minor moving) and hand rotating. However, hesitation can also be measured by unusual delays in response time. The 'significant absence' of non-verbal communications is also considered in the context of our scenario.

Two hypotheses will be tested in the context of proposed experimental design. The statement "*Social signals improve the quality of experience (QoE)*" presents the first hypothesis. QoE, subjective measure of user experience with the system, in our case depends on various factors. We can merge them into the following equation

$$\mathrm{QoE(u)} = \psi_{\mathrm{SS}}(\mathrm{SS(u,system)})+\psi_{\mathrm{P}}(\mathrm{personality(u)})+\psi_{\mathrm{M}}(\mathrm{mood(u)})+ \psi_{\mathrm{C}}(\mathrm{content})+ \psi_{\mathrm{O}}(\mathrm{other}) \quad (1)$$

where the factors ψ present the different impacts on user's QoE. Factors were selected according to preliminary case study. Theoretical background is based on statistical theory on explained and unexplained variance [20]. Factor ψ_{SS} presents the impact of social signals expressed by user during his interaction with the system, ψ_P presents the impact of user's personality, ψ_M presents the impact of user's current mood (Subsection 4.1), ψ_C presents the impact of current contents on screen (Subsection 4.4), and ψ_O that presents the impact of the unknown factors in our design. The contribution of each factor in equation (1) will be estimated from user's answer on questions in two (pre and post) questionnaires that we will use. We identify the user's personality through pre-questionnaire (control of ψ_P). Questions in a questionnaire based on description of personality with five dimensions of personality (Big Five personality traits) [20]. These five factors are; openness, conscientiousness, extraversion, agreeableness, and neuroticism. Results of personality test will be analyzed in a standard way using statistical testing.

"*The use of social signal reduces the content selection time*" presents the second hypothesis. If social signals of the user are taken into account, the time of selection of video that user wants to watch is shorter. This is the possible assumption for a user who uses video-on-demand service. However, to test this hypothesis we must ensure the same conditions for test and control group of users.

3 Experimental Design

Experimental design must allow the control of all factors in equation (1) in order to reliably estimate the contribution of ψ_{SS} to QoE. As was previously mentioned, we will extract social signals and recognize gestures in user to system interaction, when the user selects among various video contents (video-on-demand service) in order to estimate the impact of social signals in the following specific scenario. The user selects video contents with hand gestures, while the social signals can be extracted from facial expression, head movements, shoulder movement, etc. The human operator substitutes the automatic gesture recognition and social signal extraction in real time. User is not aware of human operator. The aim of this experimental design is the design of a fair experiment in terms of fair comparison between test and control group. The test group will be represented by a group of users whose social cues during the interaction with the system will be taken into account. The control group will be represented by a comparable group (in size and other selected parameters) of users whose social cues during the interaction with the system will not be taken into account. In proposed work we apply independent-measures experimental design from the aspect of feasibility of an experiment and variables control. In this section we describe the experimental scenario and technical realization of experiment.

3.1 Experimental User Scenario

Experimental user scenario can be divided in three steps, where the first one includes activities before the interaction, the second presents the interaction between user and system, and the third includes the activities after the interaction is done. All descriptions of scenario below refer to the test group of users.

Figure 2 presents the experimental environment that consists of three rooms. In room 1 there is only a monitor where the user watches emotionally neutral induction video and fills in the questionnaire before and after interaction with the system. In the room 2 there is a system that is used in a process of interaction between user and system. In room 3 there is a human operator and a monitor. Human operator watches the interaction between user and system through a camera. He makes notes of recognized actions and social signals through a human operator interface and provides video recommendations based on recommender system.

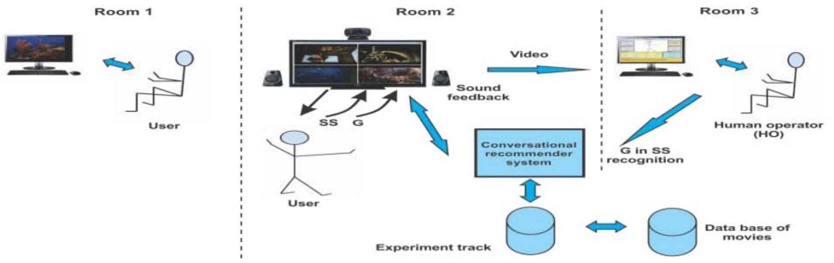


Fig. 2. Experimental environment

The first step takes place in room 1. The whole scenario is explained to the user. Then he goes to the monitor and watches the emotionally neutral video. After that, the user fills in a pre-questionnaire. At the second step the user enters room 2 and with special gesture indicates that he wishes to use the video-on-demand service. The system switches on and the interaction with the system starts. The recommender system provides four video contents – movie trailers (see Fig. 3b). These four videos are then in parallel projected on the screen. The user with a gesture (G) indicates which of four movies he is mostly interested in. The system recognizes how confident he is about his/her decision based on the social signal of hesitation (SS). If the user is not hesitating, then the system provides three additional similar items, otherwise the system provides four new diverse items and projects items on the screen together with sound feedback. User is repeating video selection process until he finds the video he wants to watch. When the user with a gesture indicates that the final decision has been made (selects the video he wants to watch), the system extends the selected video to the whole screen and turn on the sound. Then the user watches the selected movie for about 20 seconds. After this, step two is completed. The third step also takes place in room 1, where the user fills in a post-questionnaire.

The scenario for the control group of users is almost the same in all three steps. As we previously mentioned, in the control group the user's social signals are not taken into account by the system. Based on that, in the second step the system provides next three similar items. For this group, we can assume that all user decisions are made without hesitation. The decision of the system in that case is based only on gestures for video selection without social signals.

3.2 Technical Realization of the Experiment

Unfortunately, gesture recognition algorithms do not always guarantee correct results and, consequently errors in gesture recognition could provide a new uncontrolled parameter of already very complex design of our experiment. That is the main reason why human operator takes the role of automatic gesture and social signal recognition algorithms. Human operator decisions are made in real time. Technical realization of the experiment includes human operator interface and video-on-demand interface (showing videos in parallel). The human operator interface (Fig. 3a) consists of various buttons through which the human operator reports his decisions about recognized gestures and social signals. In the middle of the interface is a panel where live video from a camera recording the user takes place. User interface (Fig. 3b) represents the applied version of the video-on-demand service. It consists of four panels where videos are playing. Each decision made by the human operator is playing to user through the user interface with sound feedback.

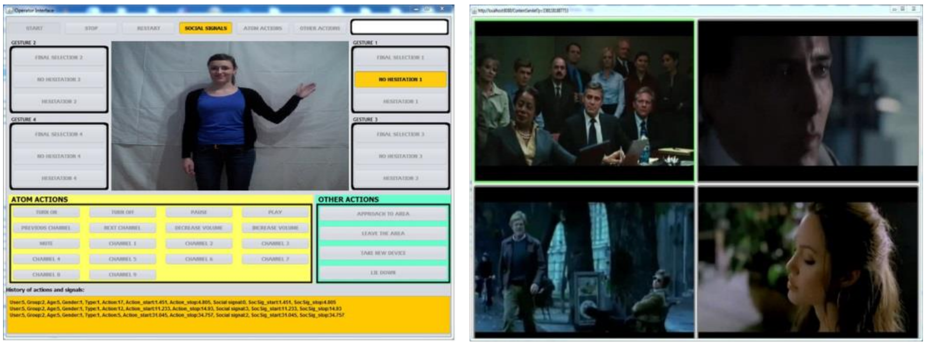


Fig. 3. a) Human operator Interface and **b)** Video-on-demand user interface

3.3 Test User Selection

As test users will be select people who reflect the generic population of moviegoers. They will be asked if they want to participate. The required number of users in test and control group will be estimated based on a-priori statistical power analysis.

4 Selected Elements of the Experiment

In the previous section, we described only the basic procedure of the experimental scenario without the details of scenario. However, we have made some assumptions on which the described scenario is based. We will discuss these assumptions and decisions below.

4.1 Role of Emotionally Neutral Video

Users that will use our system will certainly not all be in the same initial mood. This can be very critical for the control of our experiment. Therefore, we induced the neutral emotions to users by watching a short video clip. Users watched an emotionally neutral video at the beginning of the experiment before interaction with the system starts. The result of the use of this video is the approximately the same initial emotional state of all users. Video is documentary clip from a National Geographic and was already used in [22] and [23]. Clip is portraying a fish at the Great Barrier Reef.

4.2 Role of the Human Operator

The human operator is used to provide ground truth action recognition, social signal extraction and system feedback to the user in real time. He is not additionally trained to recognize the gestures, because we use simple movements. On the other hand, social signals are not simple to extract, so we need a trained person who will able to recognize the social signal from various perspectives. However, the consistency of recognitions made by the human operator will be tested with additional human operators estimating their inter-agreement using standard statistical procedures.

4.3 System Sound Feedback

The feedback from the system to the user is necessary. We assume that the user's emotional response is much less distinctive, if he does not know how his/her social signals and gestures are interpreted than if he knows. This can lead to an unpleasant user experience and consequently to useless test results. Based on that, we decided on text-to-speech synthesis system for the Slovenian language [24] with predefined sentences. The system plays a sound feedback when human operator recognizes user's gestures or social signals. The texts for the test group of users are: *"I am offering you four diverse items.", "I am offering you three similar items.",* and *"I see you have chosen the item you like."* The texts for control group are: *"I am offering you three similar items."* and *"I see you have chosen the item you like."*

4.4 Video Selection Functions

The whole test scenario includes TV remote, mobile phone and video-on-demand system selection. We limit our experiment only to the video-on-demand sub-scenario. Video-on-demand simulates an event in the video rental store or at home. The user wishes to get a video but he is not sure which one. The support person provides him with four videos and he expresses an opinion. If he is not satisfied at all, it provides him with four completely new items. If he picks one out, that one stays on and three similar ones are added. This is repeated until a final selection is made. Therefore we need three video selection functions provided by conversational RS:

```
[hA,hB,hC,hD] = getInitialItems(),
```
(2)

```
[hS,hA,hB,hC] = getSimilarItems(hS,h1,h2,h3),          (3)
[hA,hB,hC,hD] = getDiverseItems (hDi,h1,h2,h3).        (4)
```

Function (2) provides four videos for the first screen. It should diversely cover the whole matrix factorization space. Function (3) provides four videos that are similar to hS (selected video); one of them is hS. It narrows the search. Function (4) provides four videos that are not similar to h1, h2, h3 and h4. It expands the search. The function should diversely cover all factorized space of videos except those covered by h1, h2, h3 and h4. Distance metric that measures similarity among movies is based on matrix factorization space.

A conversational recommender system with no previous knowledge about the user is used. Functions `getInitialItems()`, `getSimilarItems()`, and `getDiverseItems()` based on matrix factorization feature space [25] of the LDOS-CoMoDa research dataset [17], [18]. We do not use all videos from the LDOS-CoMoDa dataset. Our subset contains over 300 videos (trailers of movies).

4.5 Role of Gestures and Social Signals

We use gestures to control the system and social signals to find out if the user is hesitating when selecting the content. Based on latter, the system expands or narrows the search. Therefore, there is only one social signal transmitted of two classes {hesitation, no hesitation} about the content the user sees. The absence of hesitation means that user is confident in his decision. In our case this is the same as the user is not hesitating. Social signal is used only to decide on diverse or similar new items.

The user uses gestures to pick up the best video or say I do not like them. With a gesture, the user also makes his first decision (select video-on-demand service) and final decision (select video he wants to watch).

4.6 Data Tracked During the Experiment

All the tracked data are stored in two files. The first includes the information about activities and social signals of the user recognized/extracted by human operator. We track the time when the activity/social signal starts and when it stops. In the second file, the feedback of the recommender system is stored. The whole interaction between user and system is also recorded. The inter-rater agreement of human operators will be tested based on these videos.

4.7 Inter-rater Agreement of Human Operators

The social signal of hesitation can be expressed in several different ways (with facial expression, head movement, shoulder movement, etc.). All forms of this kind of social signal are difficult to determine in advance so it is necessary to check the consistency of recognitions made by a human operator. We will use additional human operators who will estimate gestures and social signals based on a recorded video of user interaction. The result of test of inter-rater agreement will be presented with a coefficient of internal consistency (Cronbach's alpha).

4.8 Evaluation of the Impact of Social Signals

Evaluation of the impact of SS is based on comparison between the test and control group of users. These two groups of users will be tested in order to allow realistic estimation of effect size of the impact of social signals. Determination of the intensity of impact of social signals on user decisions in a user interaction with a system is the basic aim of this comparison.

In the test group of users, induced social signals during interaction with a communication device are taken into account. In control group of users, social signals are not taken into account. Our main task is therefore the determination of the size of the impact of social signals on a user's decisions during a communication scenario based on comparison between the test and control group of users. Comparison will be based on two questionnaires, one before interaction (pre-questionnaire) and one after interaction (post-questionnaire). For both of them, we will measure psychometric characteristics such as reliability and variability. If there is an impact of social signals on the user's decisions, the comparison between both groups of users must show the difference in contentment with the selected content, in contentment with the system, and in the user's interaction time with a system.

Pre-questionnaire consists of 17 questions and based on the 7-point Likert scale proposed in [26] and [27]. Post-questionnaire consists of 24 questions and like-wise based on a 7-point Likert scale.

5 Discussion

The expected result of the experiment is increase of statistically significant user satisfaction with a video-on-demand service when social signals are taken into account. Satisfaction may be reflected directly through faster selection of video or indirectly through results of the post-questionnaire. However, there are more factors that have an impact on user decisions, not only the social signals. We have included those factors that are expected to influence on QoE.

6 Conclusions

The proposed experimental design will be used for determination of the impact of social signals on user satisfaction with selected content and not on user's decisions or user's satisfaction with the system. The user selects among four videos projected on a screen. With simple gestures, he chooses only one. Together with gestures, the user also expresses social signals. Our work is focused on the social signal of hesitation. If the user is hesitating, the recommender system in the background offers him four diverse items according to the selected one. If not, the recommender system offers him three similar items. Most of evaluation is based on questionnaires. User fills in questionnaire before and after interaction. Tracked data about expressed social signals and gestures are also used for evaluation.

One of the advantages of the proposed design is the use of social signals in interaction. Consequently, this can increase user satisfaction with a video-on-demand system. Videos that are suggested to users are not selected randomly but recommended based on a recommender system that uses data from the LDOS-CoMoDa dataset. The next advantage is the use of an emotionally neutral video at the start of experiment. Based on that, we can get a more similar initial mood of the users.

The experiment is extremely sensitive to unknown or uncontrolled factors of a user's decision making process and that could be one of the drawbacks of the proposed design. The impact of the social signals will be measured based on a statistical analysis. A recommender system with no previous knowledge about the user is used. This is a realistic assumption for new system users with no applicable history of movie selections. The advantage of such a system is that the system is not provide additional uncontrolled parameters.

Our future plan is to implement described experimental design on a sufficiently large set of users. One of the future tasks could also be the testing of new groups of users where videos will be suggested randomly between functions that recommend similar and diverse items. This group will be compared with the control group of users where the similar items are always suggested. Based on the comparison, we can get the information regarding whether our assumption that the control group of users always gets the similar items is correct.

Acknowledgments. Operation part financed by the European Union, European Social Fund.

References

1. Pantic, M., Nijholt, A., Pentland, A., and Huang, T.S.: Human-Centred Intelligent Human-Computer Interaction (HCI²): how far are we from attaining it?. International Journal of Autonomous and Adaptive Communications Systems 1(2), 168-187 (2008)
2. Sidnell, J., Stivers, T.(eds.): The Handbook of Conversation Analysis. Wiley-Blakwell, UK (2013)
3. Håkansson, M.: Human-Computer Interaction. `http://www.sics.se/fal/kurser/isd/` Cited 19 February 2013
4. Lew, M., Bakker, E.M., Sebe, N., Huang, T.S.: Human-computer intelligent interaction: a survey. In: Proceedings of the 2007 IEEE international conference on Human-computer interaction, pp. 1-5. Springer, (2007)
5. Moon, AJ., Panton, B., H.F.M., Van der Loos, M., Croft, E.: Using Hesitation Gestures for Safe and Ethical Human-Robot Interaction. In: Proceedings of the ICRA 2010, pp. 11-13. IEEE, (2010)
6. Pentland, A.: Social Signal Processing. IEEE Signal Proc. Mag. 24(4), 108-111 (2007)
7. Vinciarelli, A., Slamin, H., Pantic, M.: Social Signal Processing: Understanding social interactions through nonverbal behavior analysis. In: Proceedings of the Computer Vision and Pattern Recognition Workshops, pp. 42-49. IEEE, (2009)

8. Jokinen, K. and Allwood, J.: Hesitation in Intercultural Communication: Some Observations and Analyses on Interpreting Shoulder Shrugging. In Ishida, T.(ed.) Culture and Computing-Computing and Communication for Crosscultural Interaction. Springer, Kyoto (2010)
9. Vinciarelli, A., Pantic, M., Bourlard, H.: Social Signal Processing: Survey of an Emerging Domain. Image Vision Comput. 27(12), 1743-1759 (2009)
10. Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., Schroeder, M.: Bridging the gap between social animal and unsocial machine: A survey of social signal processing. IEEE Transactions on Affective Computing 3(1), 69-87 (2012)
11. Bousmalis, K., Morency, L., Pantic, M.: Modeling Hidden Dynamics of Multimodal Cues for Spontaneous Agreement and Disagreement Recognition. In: Proceedings of the Conference on Automatic Face and Gesture Recognition, pp. 746-752. IEEE, (2011)
12. Sun, X., Nijholt, A., Truong, K.P., Pantic, M.: Automatic visual mimicry expression analysis in interpersonal interaction. In: Proceedings CVPRW'11, pp. 40-46. IEEE, (2012)
13. Moon, A., Parker, C.A.C., Croft, E.A., Van der Loos, H.F.M.: Did you see it hesitate? - Empirically grounded design of hesitation trajectories for collaborative robots. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1994-1999. IEEE, (2011)
14. Mu, X., Chen, Y., Yang, J., Jiang, J.: An improved similarity algorithm based on hesitation degree for user-based collaborative filtering. In Cai, Z. et al. (eds.) Advances in Computation and Intelligence. Springer, Wuhan (2010)
15. Ricci, F. et al. (eds.) Recommender Systems Handbook. Springer, New York (2011)
16. Resnick, P., Varian, H.R.: Recommender Systems. Commun. ACM 40(3), 56-58 (1997)
17. Košir, A., Odić, A., Kunaver, M., Tkalčič, M., Tasič, J.F.: LDOS - CoMoDa Dataset. `http://212.235.187.145/spletnastran/raziskave/um/comoda/comoda.php` Cited 22 February 2013
18. Košir, A., Odić, A., Kunaver, M., Tkalčič, M., Tasič, J.F.: Database for contextual personalization. Elektrotehniški vestnik 78(5), 270-274 (2011)
19. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, California (1978)
20. Montgomery, D.C.: Design and Analysis of Experiments. John Wiley & Sons, Hoboken (2009)
21. Costa, P.T., McCrae, R.R.: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual. Psychological Assessment Resources (1992)
22. Gino, F., Schweiitzer, M.E.: Blinded by Anger or Feeling the Love: How Emotions Influence Advice Taking. J. Appl. Psychol. 93(5), 1165-1173 (2008)
23. Lerner, J.S., Small, D.A., Loewenstein, G.: Heart strings and purse strings: Carry-over effects of emotion on economic transactions. Psychol. Sci. 15(5), 337-340 (2004)
24. Justin, T., Pobar, M., Ipšić, I., Mihelič,F., Žibert, J.: A Bilingual HMM-Based Speech Synthesis System for Closely Related Languages. LNCS 7499, 543-550 (2012)
25. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD, pp. 426-434. ACM, (2008)
26. Dawes, J.G.: Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5 Point, 7 Point and 10 Point Scales. Int. J. Market Res. 50(1), 61-78 (2008)
27. Finstad, K.: Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales. Journal of Usability Studies 5(3), 104-110 (2010)