Preface: EMPIRE 2014 - 2nd Workshop on Emotions and Personality in Personalized Services

Marko Tkalčič[1 2], Berardina De Carolis[3], Marco de Gemmis[3], Ante Odić[4], and Andrej Košir[2]

[1] Johannes Kepler University Department of Computational Perception, Linz, Austria
[2] University of Ljubljana Faculty of electrical engineering, Ljubljana, Slovenia
[3] University of Bari Aldo Moro, Department of Computer Science, Italy
[4] Outfit7 (Slovenian subsidiary Ekipa2 d.o.o.), Ljubljana, 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 second edition of the workshop has nine accepted papers.

Keywords: user modeling, emotions, personality

1 Introduction

The 2nd workshop on Emotions and Personality in Personalized Services[5] is taking place in conjunction with the UMAP 2014[6] conference in Aalborg, Denmark, as a full-day workshop on 11. July 2014.

Personality and emotions shape our daily lives by having a strong influence on our preferences, decisions and behaviour in general. Hence, personalized systems that want to adapt to end users need to be aware of the user's personality and/or emotions to perform well. Affective factors may include long-term personality traits or shorter-term states ranging from affect dispositions, attitudes (liking, loving, hating etc.), interpersonal stances (distant, cold, warm etc.), moods (cheerful, irritable, depressed etc.) or real emotions.

Recently, there have been extensive studies on the role of personality on user preferences, gaming styles and learning styles. Furthermore, some studies showed that it is possible to extract personality information about a user without annoying questionnaires, by analyzing the publicly available user's social media feeds. Also, the affective computing community has developed sophisticated techniques that allow for accurate and unobtrusive emotion detection. Generally, emotions can be used in personalized systems in two ways: (i) either to change the emotion

[5] `http://empire2014.wordpress.com/`
[6] `http://www.um.org/umap2014/`

(or mood, e.g. from a negative to a positive) or (ii) to sustain the current emotion (e.g. keep a user 'charged' while doing sports). Recent studies showed that such information can be used in various personalized systems like emotion-aware recommender systems.

2 Contributions

Kompna and Bieliková [1] present the results of a study that compared a standard group recommender system with the proposed modeling of groups with a graph of influences where the vertices are the users and the edges represent connections based on personality, context and relationship.

Pesek et al. [2] present a dataset of the subjects affective responses to audio stimuli. The affective responses are modeled in two dimensions: color and music perception.

The Theory of Planned Behaviour (TBP) modeling of user is present in two submissions [3,7]. Košir et al. [3] modeled the decision-making process of film viewers using TPB to predict the genre of the film to be viewed by the active user. Tkalčič et al. [7] use the TPB user modeling approach to design a persuasive system that will try to persuade users to attend classical music concerts.

Ferwerda and Schedl [4] lay out a proposal for a music recommender system based on personality and emotions. The proposal is geared towards the scenario of the anticipation of emotion self-regulation with music.

The relations between emotions expressed through posts on Facebook status updates and the users' age, gender and personality is being studied in the submission by Farnadi et al. [5].

De Carolis and Ferilli [7] argue that there is relationship between user daily routines and mood. To this end they devised a mobile app that collects affective data and calendar entries. They present their preliminary results.

Cena et al. [8] lay out a proposal for a personal informatics system that focuses on the collection of affective data for self-reflection and self-knowledge.

Chin and Wright [9] discuss the issues of a method to predict personality parameters from the observed user's social media.

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, Rosta Farzan (from University of Pittsburg, USA) and Robert Jäschke (from University of Hannover, Germany) for their guidance during the workshop organization.

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

- Aleksander Valjamae, Linköping University, Sweden
- Floriana Grasso, Liverpool University, UK
- Francesco Ricci, Free University of Bolzano, Italy
- Giovanni Semeraro, Universita di Bari, Italy
- Ioannis Arapakis, Yahoo! BCN, Spain
- Ivan Cantador, Universidad Autónoma de Madrid, Spain
- Li Chen, Hong Kong Baptist University, Hong Kong
- Luca Chittaro, University of Udine, Italy
- Maria Nunes, University of Sergipe, Brazil
- Matt Dennis, University of Aberdeen, UK
- Mehdi Elahi, Free University of Bolzano, Italy
- Mohammad Soleymani, Imperial College, UK
- Neal Lathia, Cambridge University, UK
- Olga Santos, UNED, Spain
- Oliver Brdizcka, PARC, USA
- Pasquale Lops, Universita di Bari, Italy
- Pearl Pu, EPFL, Switzerland
- Rong Hu, EPFL, Switzerland
- Viviana Patti, University of Torino, Italy
- Yi-Hsuan Yang, Yang Academia Sinica, Taipei

4 References

1. Michal Kompan and Maria Bielikova, *Social Structure and Personality Enhanced Group Recommendation*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073
2. Matevž Pesek, Primož Godec, Mojca Poredoš, Gregor Strle, Jože Guna, Emilija Stojmenova, Matevž Pogačnik and Matija Marolt, *Gathering a dataset of multi-modal mood-dependent perceptual responses to music*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073
3. Andrej Kosir, Ante Odić, Marko Tkalcic and Matija Svetina, *Human Decisions in User Modeling: Motivation, Procedure and Example Application*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073
4. Bruce Ferwerda and Markus Schedl, *Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073

5. Golnoosh Farnadi, Geetha Sitaraman, Mehrdad Rohani, Michal Kosinski, David Stillwell, Marie-Francine Moens, Sergio Davalos and Martine De Cock, *How are you doing? Emotions and Personality in Facebook*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073
6. Marko Tkalčič, Bruce Ferwerda, Markus Schedl, Cynthia Liem, Mark Melenhorst, Ante Odić and Andrej Košir, *Using Social Media Mining for Estimating Theory of Planned Behaviour Parameters*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073
7. Berardina De Carolis and Stefano Ferilli, *Towards Learning Relations between User Daily Routines and Mood*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073
8. Federica Cena, Ilaria Lombardi, Amon Rapp and Federico Sarzotti, *Self-monitoring of Emotions: a novel Personal Informatics Solution for an Enhanced Self-Reporting*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073
9. David N. Chin and William R. Wright, *Social Media Sources for Personality Profiling*, Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, July 2014, CEUR-WS.org, ISSN 1613-0073

Social Structure and Personality Enhanced Group Recommendation

Michal Kompan and Mária Bieliková

Faculty of Informatics and Information Technologies,
Slovak University of Technology in Bratislava,
Ilkovičova 2, 842 16 Bratislava, Slovakia
{michal.kompan, maria.bielikova}@stuba.sk

Abstract. The social aspects of the group members are usually omitted in today's group recommenders. In this paper we propose novel approach for the inter-group processes modeling, while the friendship type, user's personality and the group context is considered in order to reflect the group member influence. Moreover, the bi-directional emotional contagion is reflected by the spreading activation over the influence graph. In this manner we are able to compute adjusted ratings for recommended items which reflect and simulates the real users' preferences.

Keywords: satisfaction modelling, group recommendation, personality.

1 Introduction and related work

Hand by hand with the social network increase we can observe the users interaction increase over the Web. The Web became the tool for the friends' activities discussion or organization. Thus the increasing trend of group recommenders' popularity can be expected in the future.

The group recommendation extends the standard single-user recommendation to the whole group by maximizing the usefulness function. The group recommendation often aggregates the single-user preferences based on various aggregation strategies [1,3], while the minimal satisfaction guarantee is the one of the important differences. The history of the group recommenders is connected to the TV domain or movies domain or the holidays and tours recommendation.

Based on the domain, various group characteristics can be observed, while the most important seems to be the group duration (stable vs. temporary), the group structure and the activeness. From the recommendation point of view the number of recommended items (one or sequence) plays important role for the recommendation method selection. Nowadays, the social aspects of individuals become more and more important. The group often includes various user's personalities and types, while not only horizontal but vertical social structure can be observed. These aspects play crucial role in the group recommendation process, as the whole group satisfaction is based on the inter-group processes and members attitudes [9]. Often the predicted individual satisfaction, based on which the group recommendations are mostly generated differs to real satisfaction of user in the group (he/she is influenced and liable to change attitudes). The social structure (including emotional contagion) is in nowadays group recommenders considered only between two users if at all. Personal Impact Model proposed in [6] identifies user's impact on the group based on the previous activity, while the reason for the impact is hidden, thus it is unsuitable for the changing groups. Gartrell proposed a recommender based on relationship types and intensity between two users [2], while the proposed approach does not spread the impact and emotional contagion to other group members. Mashoff introduced a model for the satisfaction modeling [7]. Pro-posed satisfaction function uses the user emotions and sequence satisfaction impact. Quijano-Sanchez considers personality types and trust factors [10]. The results significantly improve the basic recommender and support the suitability of the social factors integration, while the history, the group context and the emotion influence propagation within the group should be considered when designing group recommenders.

2 Group influence

The group influence is based on the user to user individual influences. These are dependent on various personal aspects – personality, social status or relationship type. As the inter-group processes have the bi-directional character, we propose spreading activation based satisfaction function, which can be used for the rating prediction adjustment. Based on this adjusted rating we are able to choose appropriate items or sequence of items for specific group.

Similarly, the group members can be considered as the influence source. By using adjusted concept of influence modeling within the group recommendation, we are able to model real-life intergroup processes.

Group-based influence graph is an ordered pair $G = (V, E)$ where V is a set of vertices which represent group members and E set of weighted oriented edges connecting two vertexes (group members), which are associated to the influence presence between two users. The edges weights can be computed based on various social and personality based approaches (e.g. relationship type, personality).

For every group, the influence graph is constructed and each user's satisfaction (predicted rating for single-user) spread over this graph (Figure 1.). The vertexes represent group's members, while the edges refer to the emotional contagion between all pairs of users. This contagion is computed based on the social interaction type (friends, partners etc.), the single-user's personality (detected based on NEO-FFI by Costa and McCrane [8]) and the users' conflict score (obtained based on TKI by Thomas and Kilmann [11]). Moreover, the weight is adjusted with respect to the special occasions or actual context (e.g., birthdays) while these have stronger influence to the group members usually ($relationship \times personality \times context$).

Vertexes (representing group members) have assigned initial satisfaction influence values, which are derived in a standard manner (collaborative/content based recommendation prediction). The edges are present between users whose are in relationship, in other words between users with emotional contagion. Every edge has a weight, which refers to the level of emotional contagion between two users.

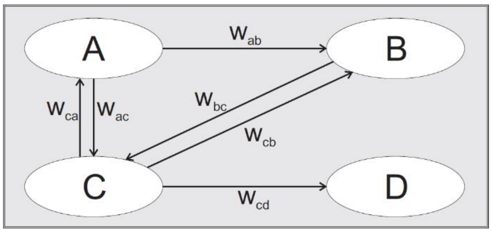

Fig. 1. Graph representation of group satisfaction modeling. Vertexes (A-D) represent users and edges the user's influence computed based on relationship, personality and actual context.

User satisfaction is the internal state of user u which describes the attitude of user to specific item i or to the all provided recommendations. This state is based on single user preferences adjusted by actual user context - emotion, mood, group members, sequence of recommended items etc.

The concept of context influence for single-user recommendations, we investigated in [5]. Based on obtained results we propose to enhance this concept to the group recommendation and introduce the group members as one of the sources for context.

In the group recommendations, often the sequence of items instead of single item is recommended. Moreover, the regularly repeated events can be considered as the sequence recommendations (time limitation required). The satisfaction does not depend only on the emotional contagion between the group members (modelled as the spreading activation through group influence graph), but on the time (emotion decrease) and previously experienced items either. Based on this assumption (previously experienced items influences user's attitudes as well) we suggest providing spreading activation several times for every item in the recommended sequence order history. The final adjusted rating is defined as follows:

$$r_{u\in Users, i\in Items} = \kappa\iota\left(\frac{\sum_{j=1}^{|RI_u|-1}\left(\left(\log_{|RI_u|-1}\sqrt{j+1}\right)sp\left(RI_{u_j},u\right)\right)}{|RI_u|-1}\right) + (1-\iota)\,sp(i,u) \quad (1)$$

where i represents the item actually experienced, and u is the user, whose satisfaction is computed. The user emotional contagion is expressed as $sp(i,u)$ - the spreading activation in user's u influence graph for predicted item i. RI_u refers to the recent items - items rated by user u previously (within one recommendation sequence). By using symmetric rating scales e.g., $< 5; 5 >$ we are able to model positive and the negative emotional contagion respectively ($\kappa = 2.631$ - compensation of logarithm for the used rating scale). The previous experienced items (sequence) is combined with the ι weight. The time emotion decrease is considered from the beginning of the sequence (log). The principle of proposed approach is presented in Figure 2.

Thanks to two basic parts – modelling of influence for actual experienced item and considering the previous history, we are able to model intergroup processes and to simulate various sequences for particular group. This allows us to maximize the group satisfaction.

3 Evaluation and results

We evaluated proposed novel satisfaction modeling (based on the group members' personalities and relationships) in the controlled experiment, where we were comparing proposed approach to the standard group recommenders (considering Average or Additive aggregation strategy).

Hypothesis. Proposed approach help us to model real life inter-group processes thus we expect that - group recommendation based on the proposed novel satisfaction modeling increases precision of group recommendation generated based on standard approaches.

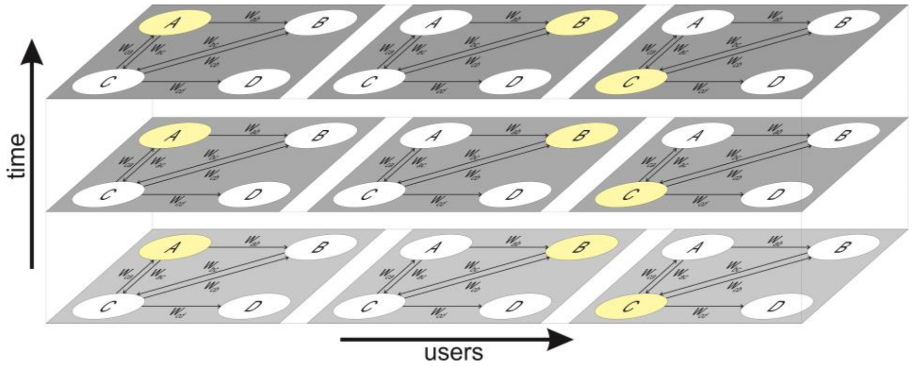

Fig. 2. Proposed satisfaction modeling for users (A,B,C,D) and several experienced items (time – each row of graphs represents one item from the recommendation history or sequence).

Participants. The total of 9 students of faculty were asked to simulate the group recommendation task in the movie domain - to choose movies which they want to watch together.

Process. Firstly, we focus on collecting personal characteristics and social structure information about participants. We asked experiment participants to complete personality based questionnaire - Neuroticism-Extroversion-Openness Five Factor Inventory (NEO-FFI), which is based on the original NEO-I measurement. Similarly, participants were asked to complete the Thomas-Killman Conflict Mode Instrument (TKI) [11] which refers to the users' competitive personality aspects. Next, each of the participants have to rate 20 movies on the rating scale 1-5 in order to model each user preferences.

In the next phase, we simulated group recommendation task. Experiment participants were spilt into groups of sizes 3 and 6 users (3 groups of size 3 and 2 groups of size 6 users) and asked to choose 3 and 5 films they want to watch together.

For each group, the influence graph was constructed, while 4 personal characteristics were considered - Competing (from TKI model) and Neuroticism, Extraversion and Cooperativeness (from NEO-FFI model). As we believe these characterises express user potential for influence other group members. For each characteristic the percentile for the population was computed and the intensity of specific characteristic was assigned (three intensity levels - low = 0.5, average = 1.0, high = 1.5). The final weight of user to user connection was computed by multiplying of intensity levels for 4 obtained characteristics.

Results. After collecting all information about participants and simulating the group recommendation task, we generated recommendations based on proposed satisfaction modeling approach and standard group recommender. We computed Precision@3 and Precition@5 as the groups were asked to choose 3 and 5 movies to recommend. Proposed recommendation approach was implemented as the standard Top-N recommender based on the Additive and Average aggregation strategy, while ratings adjusted with proposed satisfaction modelling and raw group members' ratings were used in the comparison.

As we can see (Figure 3), proposed approach generally outperforms standard group recommendation. Our results indicate, that proposed approach scored higher when less items are recommended (Precision@3). In the group recommendation this is a desired behaviour as the group does not usually demands lot of items to experience.

Obtained results suggest, that proposed approach helps us to model real inter-group influence processes which result to the satisfaction change over group members based on the their personal characteristic. This change is important in the aggregation step, while users change their single-user preferences within the group observably.

In our experiments, the weights were computed based on the 3 basic levels of influence we expected. In the real settings, when large amount of users is available, these ratings can be learned by the machine learning techniques which should even more increase performance of our proposed method.

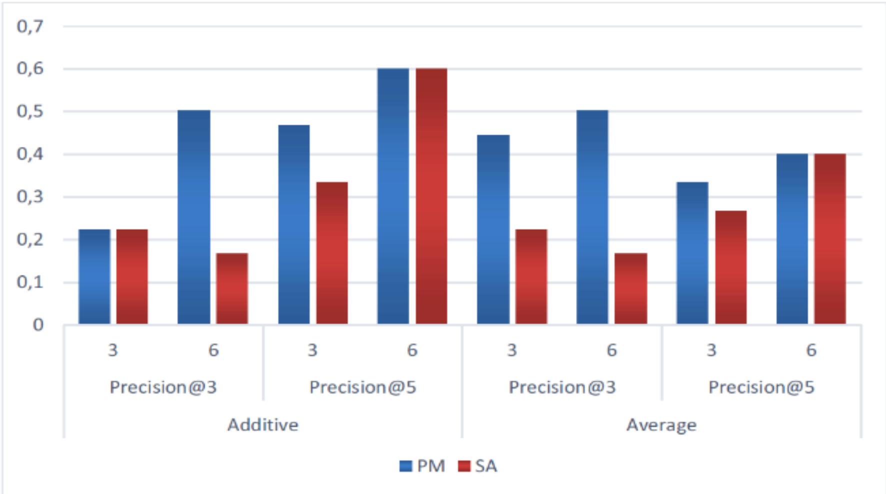

Fig. 3. Comparison of proposed influence enhanced group recommender (PM) and standard group recommender (SA). The Additive and Average aggregation strategy for group size 3 and 6 is compared.

4 Conclusions

As there are various group types, various influences can be considered in the group recommendation. When the group structure or the social characteristics are available, the consideration of such information by the influence modelling outperforms the standard group recommendation. Our proposed approach for the satisfaction modeling (adjustment of predicted ratings) models real life inter-group processes. As the emotional contagion is the bi-directional process, the spreading activation over the group influence graph allows us to reflect actual influence. Because the sequence of previously experienced items is important as well – the history of recommendation influences actual user state - proposed approach balances the present and history in order to compute adjusted rating of predicted item which in the next step is used for the recommendation. As we have shown proposed approach improves the precision of group recommender approaches and can be used for sequence optimization for group of users.

Acknowledgements. This work was partially supported by grants No. VG1/0675/11, APVV-0208-10 and by project University Science Park of STU Bratislava, ITMS 26240220084, within the Research and Development Operational Programme, co-funded by the European Regional Development Fund.

References

1. Baltrunas, L., Ricci, F.: Group Recommendations with Rank Aggregation and Collaborative Filtering. Vol. 1, pp. 119-126 (2010)
2. Gartrell, M., Xing, X., Lv, Q., Beach, A., Han, R., Mishra, S., Seada, K.: Enhancing group recommendation by incorporating social relationship interactions. In Proc. of the 16th ACM int. conf. on Supp. group work (GROUP '10). ACM, pp. 97-106 (2010)
3. Jameson, A., Smyth, B.: Recommendation to groups. In Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors, The adaptive web. Springer-Verlag, pp. 596-627 (2007)
4. Kompan, M., Bieliková, M.: Personalized Recommendation for Individual Users Based on the Group Recommendation Principles. In Studies in Informatics and Control. - Vol. 22, No. 3, pp. 331-342 (2013)
5. Kompan, M., Bieliková, M.: Context-based Satisfaction Modelling for Personalized Recommendations. In proc. Of the 8th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), IEEE, pp. 33-38, (2013)
6. Liu, X., Tian, Y., Ye, M., Lee, W.: Exploring personal impact for group recommendation. In Proc. of the 21st ACM int. conf. on Inf. and know. manag. ACM, pp. 674-683 (2012)
7. Masthoff, J., Gatt, A.: In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. In UMUAI. Springer, pp. 281–319 (2006)
8. McCrae, R. R. and Costa, P. T.: A contemplated revision of the NEO Five-Factor Inventory. Personality and Individual Differences, 36(3):587–596 (2004)
9. Postmes, T., Jetten, J.: Individuality and the Group. Advances in Social Identity. Sage Publications Ltd, London (2006)
10. Quijano-Sanchez, L., Recio-Garcia, J.A., Diaz-Agudo, B., Jimenez-Diaz, G.: Social factors in group recommender systems. ACM Trans. Intell. Syst. Tech. 4. ACM, pages 30 (2013)
11. Thomas, K. W., Kilmann, R. Thomas-Kilmann: Conflict Mode Instrument. Xicom. (1997)

Gathering a dataset of multi-modal mood-dependent perceptual responses to music

Matevž Pesek[1], Primož Godec[1], Mojca Poredoš[1], Gregor Strle[2], Jože Guna[3], Emilija Stojmenova[3], Matevž Pogačnik[3] and Matija Marolt[1]

[1]University of Ljubljana,
Faculty of Computer and Information Science
{matevz.pesek, matija.marolt}@fri.uni-lj.si
{primoz.godec, mojca.poredos}@lgm.fri.uni-lj.si

[2]Scientfic Research Centre of the Slovenian Academy of Sciences and Arts
Institute of Ethnomusicology gregor.strle@zrc-sazu.si

[3]University of Ljubljana,
Faculty of Electrotechnics
{joze.guna, emilija.stojmenova, matevz.pogacnik}@fe.uni-lj.si

Abstract. The paper presents a new dataset that captures the effect of mood on visual and auditory perception of music. With an online survey, we have collected a dataset of over 6600 responses capturing users' mood, emotions evoked and expressed by music and the perception of color with regard to emotions and music. We describe the methodology of gathering the responses and present two new models for capturing users' emotional states: the *MoodGraph* and *MoodStripe*. Also, general research questions and goals, as well as possible future applications of the collected dataset, are being discussed.

Keywords: color perception, human computer interaction, mood estimation, music information retrieval

1 Introduction

The complexity of human cognitive processes forces researchers to focus on specific narrow problems within their field of expertise. By selectively partitioning the multi-modality of human perception on the respective field-specific features, many other aspects are excluded or presumed to have no impact on the results. For example, much of music information retrieval (MIR) research to date has been *systems-based*, conducted in laboratory environment by assuming some objective "ground truth" (such as genre classification) and ignoring user context and individual preferences [12]. Such models or systems are "one-dimensional", restricted to some simple notion of perception, and fail to capture "real-world" scenarios.

1.1 Aim and scope

In order to propose a system capable of multi-dimensional modeling of human perception, we have considered various perceptual modalities and built a dataset for multi-dimensional model. Our work connects approaches in the fields of human-computer interaction (HCI) and MIR, with an intention to incorporate feature types from both fields into the model. Our aim is to evaluate two aspects of human perception: the color and music perception and concurrently bind these to the emotions induced and perceived from those modalities. By extracting the information about the user's emotional state, we attempt to link the variable components of one's personality which presumably impact the decision-making processes of a human. Following paragraphs provide a brief overview of closely related problem tasks in the fields of HCI, MIR and music visualization.

An impressive amount of attention has been given to user modeling and personalization of user experience (UX) in different application domains. Originating from product-oriented problems, the UX has become one of the most important factors in scientific fields including computer graphics and HCI [1]. In MIR research, on the other hand, there have been only a handful of studies focusing on user satisfaction by including her personal preferences into the decision making model [12]. These typically include music features representing the music genre, artists, instrumentation and mood [16].

Mood has been extensively analyzed within several scientific fields, not only in the fields of psychology and cognitive science [2,10], but also in the fields of HCI [15] and in recent years in MIR [8,12]. Some of the approaches that use mood for personalization capture mood by eye-tracking and visual recognition (with cameras), while others rely on manual input of relevant information by interacting with the interface. In MIR, mood estimation from music is a relatively new task [7], generally evaluated independently from other user-dependent tasks. However, we argue that for mood estimation to be successful, it needs to be integrated with other user-dependent tasks, such as music recommendation task [14]. The evaluation metrics for music recommendation often rely on play-list comparison. The task is formalized as a production of an ordered list of songs, given a query and a set of source data. "The results are evaluated against a ground truth derived from a second source or human evaluation"[1]. While this may seem as a valid evaluation procedure, it discards any personal information about the user, which could, in our opinion, have great impact both on the results and on the functionality and personalization of the recommender system in practical use.

A notable effort has been put into visualizing the music data on multiple levels: audio signal, symbolic representations and meta-data [9]. Often, these representations contain interpretable psychological dimensions, built upon relationships between multiple qualitative dimensions of represented entities and additional data in form of shape, size or color [3]. Color tone mappings are also applied onto the frequency, pitch or other spectral components [11], in order to

[1] `http://www.music-ir.org/mirex/wiki/2009:Music_Recommendation`

describe the audio features of the music [18]. However, such representations significantly differ, also due to the choice of the color set, which is typically picked randomly by the creator. Furthermore, there is no standard problem task evaluating the "correctness" of music visualization, since the evaluation procedure would be unclear due to the amount of factors possibly influencing the ground truth produced by a person. Thus, the proposed visualization techniques often serve merely as the visually appealing aspect of the representation. Nevertheless, we believe that certain uniformity exist in the general perception of connections between colors and music.

1.2 Research questions

Based on the proposed model and subsequent studies, our goals are set along the following research questions:

- Does users' mood impact their perception of color and music?
- Is the relationship between color and emotions uniform or individual? To what extent can such relationship be influenced by mood and personal characteristics?
- Does the correlation between sets of perceived (expressed by music) and induced (evoked by music) emotions depend on the personal musical preferences?
- Are there *emotionally ambiguous* music excerpts?

Our general hypothesis is that perceived emotions of a music excerpt are expected to be similar across listeners, whereas the induced emotions are expected to be correlated across groups of songs with similar features (such as genre) and users with similar personal characteristics. Furthermore, if there is a significant variance between sets of perceived and induced emotions, the latter could be used to implicate user's satisfaction of the given selection. We assume such relationship can be integrated into music recommender system, significantly improving current state-of-the-art emotion-based music recommenders. Another aspect, largely ignored by existing recommenders, is the effect of music on user's current mood. If results prove such effect to be significant, there is a real motivation for constructing progressive recommender system, dynamically adjusting music recommendations according to emotionally charged music changes selected by the user. We plan to implement these aspects for playlist generation and evaluate the usefulness of such approach with regard to current state-of-the-art recommenders that do not consider such variability.

From HCI perspective, there are a number of possibilities, where user's mood can be usefully taken into account. One of the most obvious ones is improvement of recommender algorithms for personalized selection of multimedia content (music, movies, books, etc....). Standard personalization techniques have reached the stage where user context is taken into account in addition to user's general preferences. The user's context is usually comprised of a number of parameters such as user's mood, location, time of day, activity, etc. While the number of available movies for example, has reached many thousands, the recommended ones can be consequently counted in hundreds. Taking user's mood into account can further reduce the number of recommended movies to ten or twenty, which makes the recommendation system much more user friendly.

The paper is structured as follows: section 2 provides a detailed methodology of the survey design and the proposed *MoodGraph* emotion model, while section 3 elaborates on the gathered dataset and discusses possible future applications.

2 Methodology

We started our survey design with a preliminary questionnaire (see subsection 2.1), which provided some basic guidelines for the overall design. The following subsections describe the structure of the survey, divided into three parts. These are intended to capture user's emotional state, the perception of colors and corresponding emotions, and emotions perceived and induced from music, along with the corresponding color.

2.1 Preliminary study for bias exclusion

As there are no generally accepted sets of textual descriptors (labels) for various emotional states, beyond the set of basic emotions proposed by [5], many studies choose labeled sets intuitively, with no further explanation, e.g. [17]. In contrast, we performed an initial study of existing emotion sets used within psychology and music research in order to establish a relevant set of emotion labels. In order to eliminate the cultural and lingual bias on the labeling, we performed our survey in Slovenian language for Slovenian-speaking participants. The preliminary questionnaire asked the user to describe their current emotional state through a set of 48 emotion labels, each with an intensity-scale from 1 (inactive) to 7 (active). The questionnaire was solved by 63 participants. Principal component analysis of the data revealed that first three components explained 64% of the variance in the dataset. The majority of the 17 emotion labels for our survey were then chosen from the first three components. To capture the relationship between colors and emotions, we evaluated the effectiveness of the continuous color wheel for choosing colors. Responses indicated the continuous color scale to be too complex for some users and a modified discrete-scale version was chosen instead. The discrete color wheel provides the user with a choice of 49 colors displayed on large tiles, enabling the user to pick the best match. All 49 colors have been chosen to represent a common color spectrum of basic colors and provide a good balance - a trade-off between the complexity of the full continuous color wheel and the limitations of choosing a smaller subset of colors.

2.2 Structure of the survey

We divided our online survey into three parts. The first part contains basic demographic questions, including questions regarding musical experience. The second part investigates user's current mood and her/his perception of relationships between color and emotions. Part three consists of ten 15-second long music excerpts and questions related to perceived and induced emotions, as well user's perception of relationship between color and individual music excerpt.

2.3 Part one - Demographic data

In order to evaluate how personal characteristics of users impact their music and color perception, we designed a set of questions, shown in Table 1. The set contains three demographic-oriented questions, two questions related to the musical background, including user's preferences and interests in music, and two questions inquiring about medications and drug use (we consider them necessary in order to detect possible discrepancies).

Table 1. Questions from the first part of our survey. Each question is provided with a set/range of possible answers (2nd column) and with comments where needed (third column). The responses provide some background information about the user's music experience and preferences.

Question	Range	Comments
Age	[5, 99]	*in years*
Gender	{Male, Female}	
Area of living	{city, rural area}	
Music school attendance	[0, 20]	*in years, 0 - meaning not attending*
Instrument playing or singing	[0, 20]	*in years, 0 - meaning not attending*
Usage of drugs	{yes, no}	*Is participant using drugs*
Influence of drugs	{yes, no}	*Is participant under the influence of drugs when filling the survey*
Genre preference	{Classical, Opera, Country, Folk, Latin, Dance / Disco, Electronic, RnB/Soul, Hip Hop/Rap, Reggae, Pop, Rock, Alternative, Metal, Blues, Jazz, Vocal, Easy Listening, New Age, Punk}	*up to three preferred genres (at least one) can be selected starting with the most preferred genre.*
Time listening to the music	{less than 1, 1-2, 2-3, more than 3}	*in hours per day*

The introduction of a larger set of demographic questions has also been considered. However, as the focus of our research is investigation of the interplay between colors, music and emotions, a very demanding task in itself, the decision has been taken not to put additional stress on the user by conducting lengthy demographic investigation. The amount of time required to finish the survey was averaged under 10 minutes.

2.4 Part two - mood, emotions and colors

This part of the survey is designed to capture information about user's current mood, her/his perception of relationship between color and emotions, and evaluation of the latter in terms of pleasantness and activeness. Here, mood is defined as a longer lasting, but less intense emotional state (generally unaffected by current situation or particular stimulus), compared to emotion (which is usually directly affected by a particular event or stimulus), though such definition is still arguable as both terms are often used interchangeably (see e.g. [6,13]). We use term 'emotional state' in places where both mood and emotion are being referred to. The structure of the second part of the survey is outlined in Table 2.

Table 2. The second part of the survey inquiring about the perception of emotions and colors. The fourth question is visually divided into three blocks, with six, six and five emotions in each block. The fifth question is presented to the user in a combination of a word describing the mood and a color wheel for user input (see Figure 2).

Question	Range	Comments
Current mood	Valence/Arousal space	*User selects her/his mood in a 2 dimensional space according to the pleasantness and activeness of the mood - see Figure 1*
Color of the mood	Color wheel	*User chooses the color tone most reflecting her/his current mood - see Figure 2*
Perception of emotions	{fear, energetic, angry, relaxed, happiness, sadness, liveliness, joy, disappointment, discontent}	*User places the set of emotions onto valence/arousal space - see Figure 3*
Emotional state	{active, wide awake, drowsy, inactive, miserable, discontent, disappointment, relaxed, happiness, cheerful, joyous, satisfied, sleepy, sad, calm, angry}	*User places all emotions onto a stripe with a continuous scale ranging from unexpressed to highly expressed - see figure 4*
Colors of emotions	{energetic, discontent, sad, disappointed, relaxed, angry, fearful, happy, joyous, lively}	*For each word in a set, user selects a color most resembling the described emotion.*

First, the participants are asked to describe their current emotional state. Instead of having them choose a set of emotional labels, they performed this task by placing a point in the valence-arousal space (Figure 1). This is a standard mood estimation approach, also frequently used for the estimation of perceived emotions in music. The same approach was used for self-estimation in the *MoodStripe*. We believe, that the valence-arousal space is a fast and intuitive way for the user to describe her/his current emotional state.

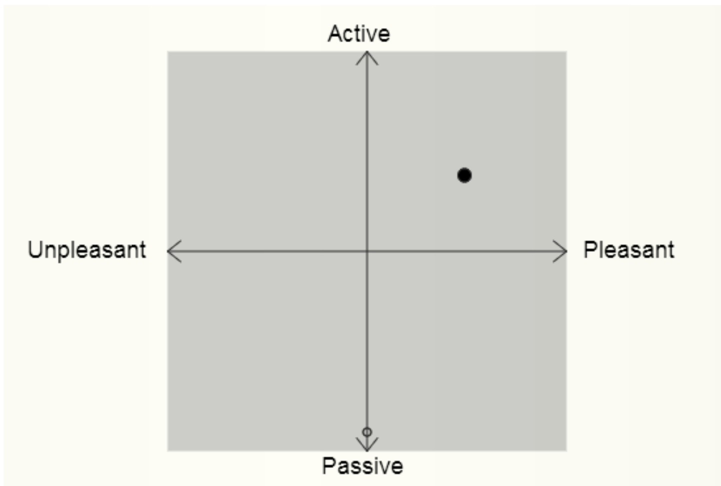

Fig. 1. Valence-arousal space for estimation of user's emotional state. The graph axes are marked *Uncomfortable* and *Comfortable* for the abscissa, and *Passive* and *Active* for the ordinate values.

Questions two and five refer to a color wheel, shown in Figure 2. Participants were asked to choose the best-matching color for particular emotion. The color wheel contains a total of 49 possible color selections (see 2.1).

Question three was designed to assess how users perceive the pleasantness and activeness of emotions in the valence/arousal space, called *MoodGraph* (Figure 3). The users were asked to place a set of ten emotion labels in the *MoodGraph*, according to their perception of pleasantness and activeness of individual emotion. Although we assume that emotions will form distinctive clusters in the valence/arousal space across users (thus, gathering the *stereotypical* representation of the emotions), we nevertheless wish to evaluate the variability of the emotion labels placements in terms of their activeness and pleasantness, and compare with the results in part three, where users described musical excerpts in a similar manner.

The fourth question was designed to capture users' current emotional state, by annotating 17 emotions on a scale from unexpressed to highly expressed. To make the task more intuitive and keep in line with the overall UX (user experience) of the survey, we designed a new input modality, *MoodStripe* (Figure 4). As in *MoodGraph*, the user drags emotion labels onto space representing

Fig. 2. A color wheel with total of 49 colors presented. The black dot indicates the selected color.

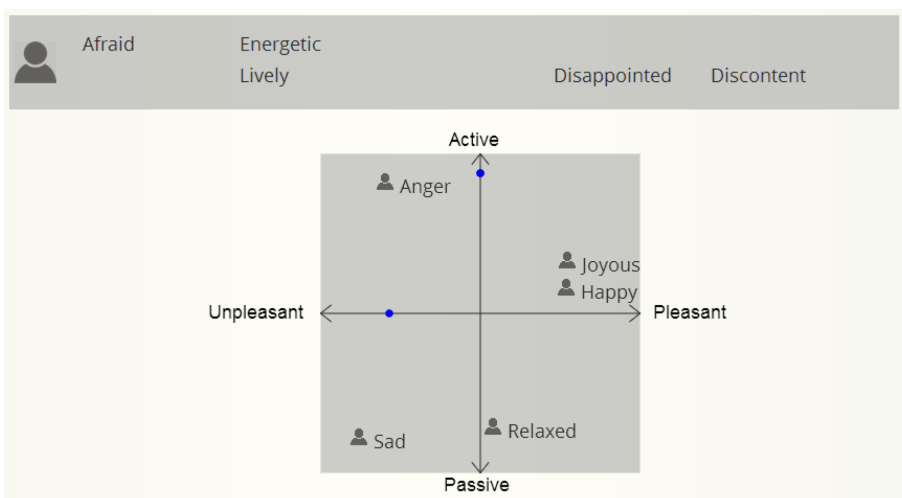

Fig. 3. The *MoodGraph*: emotions are dragged from the container above the graph onto the valence/arousal space. Blue dots indicate the position of the selected emotion in the valence/arousal space.

expressivity of emotions. To ensure the ease the positioning, the emotions are divided into three separate *MoodStripes*.

Both novel user input modalities, the *MoodGraph* and *MoodStripe*, could easily be replaced by a set of ordinal scales, implemented as radio buttons for each emotion. However, we believe that our approach is more efficient, intuitive

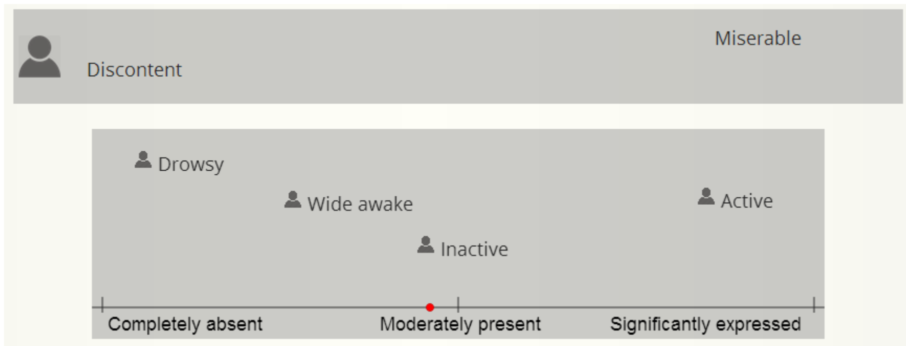

Fig. 4. The *MoodStripe*: to express their emotional state, users position a set of emotions on a scale from unexpressed on the left to highly expressed on the right. The scale is marked from *absent* over *present* in the middle, to *expressed* the right.

and puts less stress on the user. We provided an optional feedback form after users finished the survey and preliminary analysis reveals very positive remarks about the proposed innovative input modalities.

2.5 Part three - music in relation to colors and emotions

In part three of the survey, users are asked to complete two tasks on a set of ten randomly selected 15-second long music excerpts. After listening to an excerpt, the user is first asked to choose the color (from the color wheel) that best represents current music excerpt (Figure 2). Next, the user is asked to place a set of emotion labels on the *MoodGraph*, according to two categories: a category of induced emotions and a category for perceived emotions. The category of induced emotion labels is marked with a person icon, while the perceived emotion labels are represented with a note icon. The user may place any number (but at least one from each category) of emotions on the *MoodGraph*. An example is shown in Figure 5.

3 Contributions and future work on the collected dataset

This survey was published in March. With the help of social networks we have gathered responses of 952 users to date. This number is much higher than expected, particularly as the survey was only distributed among Slovenian speaking population. Users provided 6609 mood/color-perception responses for the 200 music excerpts used in the survey.

To achieve the most relevant results in regard to the emotional influence of music, we've carefully selected the songs with an intention to cover the whole variety of genres, and at the same time avoid popular (or most known songs), that could bias users' ratings. The database includes 80 songs from the online service

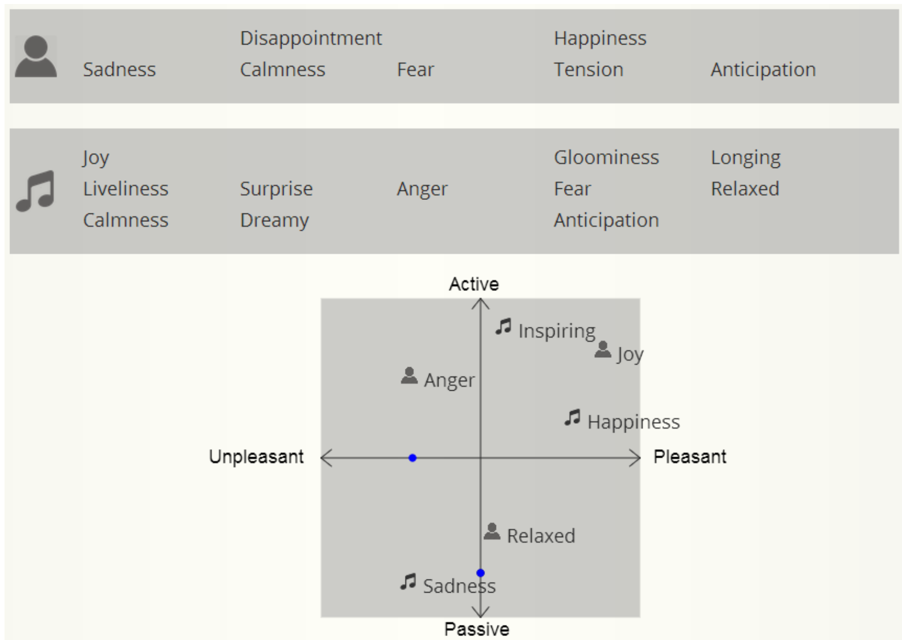

Fig. 5. An example of two-category *MoodGraph*. Induced emotions are marked with a person icon, perceived emotions with a note icon.

Jamendo[2], which is widely used by less-known artists. We have also included a part of a previously collected dataset [4] containing 80 excerpts of film music, a subset of 20 Slovenian folk music excerpts and 20 music excerpts from past International Computer Music Conference proceedings containing contemporary electro-acoustic music. The diversity of the chosen sets will allow us to analyse the differences of responses and draw possible conclusions by extracting the audio liable features. By gathering a vast amount of responses (each excerpt has on average 33 responses), including induced and perceived emotions and relationships between music excerpts and color, we've provided a foundation for future research on person's emotional state, perception of music and color, and complex relationships between perceived and induced emotions in music.

To our knowledge, no currently available music-mood dataset has such a high ratio of user annotations per music excerpt. By analysing the data provided by *Google analytics*, we've established that the average time to complete the survey was 8.27 minutes, thus we reached our goal of a less than 10 minute long survey.

We intend to make the entire dataset available to the public, including musical excerpts, the data on users' emotional states, their placement of emotions within the valence/arousal space, their perceived and induced emotional responses to music and their perception of color in relation to emotions and music.

[2] `http://www.jamendo.com`

This will open new possibilities for evaluating and re-evaluating mood estimation and music recommendation approaches on a well annotated dataset, where the ground truth lies in the statistically significant amount of responses per song, rather than relying on annotations of a small number of users.

Shortly, we will also publish an English version of the survey, where an additional set of responses will be gathered. We also intend to enlarge the number of music excerpts in the music dataset and provide it to the users who have already participated in the study. Thus, we hope to further extend and diversify the music collection.

The information collected during this research will possibly provide the basis for the realization of the following research goals and future work:

- Previously introduced mood estimation algorithms can be evaluated by weighting the correctness of their predictions with the distribution of perceived-emotion responses for a music excerpt. A new mood estimation algorithm will be developed, building upon the newly obtained data.
- We will explore the modelling of relations between music and the corresponding colors chosen by users in the survey. Results may be useful for music visualization, provided that correlations between audio and visual perception will be consistent enough.
- A user interface for music recommendation will be developed, presenting recommendations in a visual manner, possibly raising the user satisfaction by reducing the textual information needed to be presented to the user. The interface will include personal characteristics and their variability in the decision model.
- The dataset can also be used in several other domains. Responses that link color and mood perception based on the user's emotional state can be used independently.

References

1. William Albert and Thomas Tullis. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Google eBook)*. Newnes, 2013.
2. Tiziano Colibazzi, Jonathan Posner, Zhishun Wang, Daniel Gorman, Andrew Gerber, Shan Yu, Hongtu Zhu, Alayar Kangarlu, Yunsuo Duan, James A Russell, and Bradley S Peterson. Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion (Washington, D.C.)*, 10(3):377–89, June 2010.
3. Justin Donaldson and Paul Lamere. Using Visualizations for Music Discovery. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, page Tutorial, 2009.
4. T. Eerola and J. K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1):18–49, August 2010.
5. Paul Ekman. An argument for basic emotions. *Cognition and Emotion*, 169-200(6), 1992.

6. Alf Gabrielsson. Emotion Perceived and Emotion Felt: Same or Different? *Musicae Scientiae*, 5(1_suppl):123–147, September 2002.
7. Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull. Music emotion recognition: A state of the art review. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 255–266, Utrecht, 2010.
8. Cyril Laurier, Owen Meyers, Joan Serrà, Martin Blech, Perfecto Herrera, and Xavier Serra. Indexing music by mood: design and integration of an automatic content-based annotator. *Multimedia Tools and Applications*, 48(1):161–184, October 2009.
9. Nicola Orio. Music Retrieval: A Tutorial and Review. *Foundations and Trends® in Information Retrieval*, 1(1):1–90, 2006.
10. N A Remmington, L R Fabrigar, and P S Visser. Reexamining the circumplex model of affect. *Journal of personality and social psychology*, 79(2):286–300, August 2000.
11. Shigeki Sagayama and Keigo Takahashi. Specmurt anasylis: A piano-roll-visualization of polyphonic music signal by deconvolution of log-frequency spectrum. In *ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing*, Jeju, Korea, 2004.
12. Markus Schedl, Arthur Flexer, and Julián Urbano. The neglected user in music information retrieval research. *Journal of Intelligent Information Systems*, 41(3):523–539, July 2013.
13. K. R. Scherer and M. R. Zentner. Emotional effects of music: production rules. In P. N. Juslin and J. A Sloboda, editors, *Music and emotion*. Oxford University Press, New York, 2001.
14. Y. Song, S. Dixon, and M. Pearce. A survey of music recommendation systems and future perspectives. In *Proc. 9th Int. Symp. Computer Music Modelling and Retrieval (CMMR)*, pages 395–410, London, 2012.
15. Marko Tkalcic, Ante Odic, Andrej Kosir, and Jurij Tasic. Affective Labeling in a Content-Based Recommender System for Images. *IEEE Transactions on Multimedia*, 15(2):391–400, February 2013.
16. Jill Palzkill Woelfer. The role of music in the lives of homeless young people. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 367–372, Porto, 2012.
17. B. Wu, S. Wun, C. Lee, and A. Horner. Spectral correlates in emotion labeling of sustained musical instrument tones. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 415–421, 2013.
18. Ho-Hsiang Wu and Juan P. Bello. Audio-based music visualization for music structure analysis. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, 2010.

Human decisions in user modeling: motivation, procedure and example application

Andrej Košir[1], Ante Odić[2], Marko Tkalčič[3], and Matija Svetina[4]

[1] University of Ljubljana, Faculty of electrical engineering
Tržaška 25, 1000 Ljubljana, Slovenia
andrej.kosir@fe.uni-lj.si,
[2] OutFit 7 (Slovenian subsidiary Ekipa2 d.o.o.),
Ameriška 8, 1000 Ljubljana, Slovenia
ante.odic@outfit7.com,
[3] Johannes Kepler University
Altenberger Strasse 69, 4040 Linz, Austria
marko.tkalcic@jku.at
[4] University of Ljubljana, Faculty of arts
Tržaška 2, 1000 Ljubljana, Slovenia
m.svetina@ff.uni-lj.si

Abstract. The aim of this study was to investigate the use and potential of the psychological theory of human-behavior modeling, called the Theory of Planned Behavior (TPB), in a user-modeling domain. We performed a user experiment involving a well-studied problem of user modeling, i.e., a recommender system (RS) for movies. As a part of the TPB, a survey to estimate the behavioral, normative and control beliefs regarding movie selection was designed. Using the participants' responses an Ajzen model for movie genres was built and evaluated. An existing public dataset for context-aware movie recommendation, CoMoDa, was used to evaluate the proposed method. The results showed that the TPB approach led to an interesting explanation of movie genre selection. Among others, the potential applications of the TPB in recommender systems and the architecture of such a RS were addressed. Questions about the potential applications of the TPB in the user-modeling domain and its limitations and drawbacks were discussed.

Keywords: Theory of planned behavior, Behavior interpretation, Recommender system

1 Introduction

User-modeling and user-adaptation techniques have received much attention in recent decades as a way to tackle the problem of human-computer interaction across a broad range of communication services. Recommender systems, as an element of this user modeling, are today a part of most services that involve content or service selection made by end users. Many user-adaptation tasks can be seen as a problem of the effective recommendation of a predefined set of entities. Several different directions of algorithm are under development due to the fact that effective user adaptation is very much dependent on the domain of recommendations. However, several drawbacks of the existing user-modeling techniques are only partially solved, such as the problem of intrusive end-user data acquisition, end-user privacy protection, the problem of diversity of the RS, etc.

Human-behavior modeling has been an intensive research field in psychology for several decades. The Theory of Planned Behavior [1] is particularly appealing in user modeling and adaptation for several reasons. First, the behavior model is relatively easily interpreted in several domains in such a way that the available domain knowledge can be utilized. Second, the procedure of building the Ajzen model for a given domain is a well-defined procedure (we present it in Sec. 2). Third, the prediction model is not predefined but can be selected according to the domain knowledge. Fourth, there are a large number of modeling cases providing rich past experiences resulting in effective modeling guidelines.

We present the procedure for Ajzen model building, including how to select predefined behaviors and demonstrate the model on a dataset of real users. We discuss the potential of this type of psychological modeling of human behavior in user-adaptation procedures. The discussion also addresses the constraints and issues of further development with regard to the implementation of the TPB into the RS.

1.1 Related work

The usual reasoning in RS procedures is to build the model of a user u according to his/her past treatment of the domain items h. These items are multimedia-content items, tourist destinations, selected food, etc. No underlying mechanisms that govern the user's interaction with the system are normally taken into account. For example, the Netflix prize-winning algorithm Matrix Factorization (MF) has no model of users or items that is based on the users or item features (metadata).

Recommender systems The main goal of RSs is to predict ratings for items that the user has not yet consumed. Based on these predicted ratings, suitable items (those with high predicted ratings) are selected and provided as the recommendations.

Content-based (CB) recommender systems [18] analyze the items' descriptions in order to learn the user's preference for specific types of items. The prediction of the unseen item is based on ratings for similar items provided by the same user. In Collaborative Filtering (**CF**) strategies the prediction of the unseen item is based on the opinion of users with similar tastes [20]. This approach ignores the items' metadata, so cross-domain recommendations are possible (e.g., books, movies, music, etc.) by employing cross-domain techniques [11]. After the Netflix prize competition [7], Matrix Factorization has become a popular CF technique [14]. However, according to [5], for the user whose tastes are unusual compared to the population, the similarity compared to other users will be poor, which will result in poor recommendations for such a user.

This kind of model construction has a number of constraints. One way to address these issues is to gain additional knowledge with regard to the underlying mechanisms of the user's interaction with the system. One such model is the Lazy User Theory [21]. Here the authors developed a theory that the user will most often choose the solution that will fulfill his/her information needs with the least effort. Such an assumption allows us to explain selection factors using multivariate statistics, but it also assumes that a user has a clearly defined goal while seeking information. However, it seems that this simple and strong hypothesis is not valid in many situations of a user's interaction with information systems due to the fact that modern users employ these systems with no specific goal. A different theory, i.e., TPB, seems more promising in this context. The theory and the rationale for applying it in the RS context are provided below.

Theory of Planned Behavior (TPB) The pioneering work on the Theory of Planned Behavior was carried out by Icek Ajzen [2], and so the model suggested by the TPB is usually called the Ajzen model.

The Ajzen model was introduced as a complete model for explaining human behavior and is based on a large number of behavior studies. According to the TPB, human behaviors are influenced by attitudes towards their behavior, by subjective norms regarding their behavior, and perceived behavior control [2]. The structure of TPB is depicted at Fig. 1 where the aggregated model is presented. Behavior is domain specific; in this study we selected the behavior as the selection of a movie with a given genre. Attitudes are beliefs that one person has about the outcomes of the behavior (seeing the selected movie) and are divided into cognitive, emotional and behavioral. Subjective norms are related to beliefs about the expectations of others and the wish to comply with them. Behavior control relates to the ability one has to perform the preselected behavior and this directly affects the decision about the behavior.

There are several areas where human decision making is of key importance and which are exhaustively studied using TPB models, such as outdoor recreation activities [8], decisions related to high-school studies [9], public-transport habits [6], health-related behavior [4], consumer attitudes and behavior [3], employers' hiring intentions [12], job satisfAction [13], adoption of wireless-sensor-network services in households [17], factors influencing the intention to watch online-video advertising [16] and mobile-phone usage while driving a car [22].

The common goal of these studies is not only to be able to predict human decisions but also to understand the underlying mechanism of these decisions. These explanations are then used to create a new theory or to modify existing ones in order to provide further insights into the targeted domain.

1.2 Problem statement

The goal of this paper is to provide a rationale for using the Theory of Planned Behavior (TPB) in user-modeling applications. We present the background of the TPB and outline the procedures for the acquisition of the TPB parameters. As a proof-of-concept we present the results of an experiment where we used the TPB model in a recommender system for movies. Since the creating of a final version of the effective and valid questionnaire is beyond the scope of this preliminary work, conclusions from the model are only partially valid. The TPB was meant to model planned behaviour and was, to our knowledge not yet used in predicting movie preferences. We address a new account in using psychological driven theory in user modeling. The present paper is an attempt to gain initial evaluation of such an approach.

2 The procedure of model building

We list below a procedure for TPB model building to collect the most relevant guidelines and potential errors for the UM community.

1. *Define a set of behaviors.* This is the most important step in the whole modeling task. Prior to it the reason why we apply the TPB must be clarified. In this paper's given example the reason was to further understand the mechanisms of movie genre selection. We therefore assume that movie genre selection is influenced by cognitive, emotional and behavioral control, and by social norms. So we can expect that understanding the reasons for these variabilities would provide the insight that we can utilize to improve the accuracy of the user model in the movie recommender. The behaviors are required to be discriminable with reasonable user data. The behavior definition should rely on an end-user data analysis and on the clear goal of the modeling itself.
2. *TPB questionnaire construction.* The next step is to design a questionnaire for the end users in order to estimate the parameters of the model. It must meet the requirements set by the TPB. We group them into three groups with respect to: behavioral beliefs (about the consequences of the behavior), normative beliefs (about the expectations of others) and control beliefs (about factors that affect the performance of the behavior). Therefore, this construction requires an in-depth domain knowledge of the selected behaviors. The basis of all the questions is the defined behaviors (see step 1.). The next issue addressed is the specification of the end-user population. Five to six questions are then formulated to assess each of the constructs (attitude, norms, control and intention).
3. *Select and build the prediction model.* According to the constructed questionnaire and the set of predefined behaviors, a prediction model is selected. First, the criteria variable indicating the true behavior is constructed. In our example of movie-genre selection, for the first criteria the variable is computed from the previous movie selection of the targeted end users (see Sub. 3.3). For the second model, the criteria variable is simply the genre indicator of the most likely selected genre by this end user. Next, the model itself is selected. Typically, the first option considered is a multivariate linear regression model (MVR), if the predictor and criteria variables fit the

requirements. Other options include linear discriminant analysis (LDA), the logit regression model, canonical regression, structural equation modeling, etc. In general, there is no limitation from the TPB imposed on the model selection. However, the explanation power of the selected model also matters, since the interpretation of the fitted model may provide useful hints for a further improvement of the user-adaptation procedure.

4. *Interpret the model.* The interpretation of the models is based on a standard interpretation of selected models. For instance, the linear-regression model is interpreted according to the sign and the magnitude of the estimated normalized model coefficients, etc.

3 Materials and Methods

3.1 Participants

In our experiment we had 28 subjects, aged between 17 and 38 years old (18 males and 10 females). Each of the subjects filled in a TPB questionnaire using GDrive forms. The users were selected from contributors of the movie ratings in the contextual movie dataset CoMoDa [15].

3.2 Instruments

The constructed TPB questionnaire consists of 49 questions related to beliefs regarding movie selection and consumption according to the TPB and is available on-line. The filling time was 10 to 12 minutes. Most of the answers were 5-level Likert scales, i.e., Not important 1 - 5 Important (17), Not really 1 - 5 Very much (15), Never 1 - 5 Very often (2), ratings 1 - 5 (3) and enter nonegative number (7). One question was No - Yes, one question was predefined genre selection and some required a free-text answer (3). The questionnaire is available at `www.ldos.si/ComodaTPBv01.html`.

3.3 Construction of criteria

We describe the ground-truth user behavior (see Fig. 1) with two criteria variables determined from the user's known previous movie selections and ratings that they provided for the CoMoDa dataset [15]. Each of the rated movies in the dataset has three genres assigned to it. The first criteria variable is the genre scores denoted by $\mathrm{gS}(u, g)$ where u is the user and $g \in \{\mathrm{Drama}, \mathrm{Action}, \mathrm{Comedy}\}$ is the movie genre. It is defined as a ratio between the number of movies selected having the genre g and the number of all the genre (movie) selections. For example, if a user u has rated 45 Drama movies and provides 91 ratings for the database, we have $\mathrm{gS}(u, \mathrm{Drama}) = 45/(3{\cdot}91) = 0.165$, since every movie selection means a selection of three genres.

The next criteria variable we introduce is genre membership $\mathrm{gC}(u)$, where u is the user. The indexes of the genres (also the behaviors in our case) are

$\mathrm{Id}_{\mathrm{Drama}} = 1$, $\mathrm{Id}_{\mathrm{Action}} = 2$ and $\mathrm{Id}_{\mathrm{Comedy}} = 3$. $\mathrm{gC}(u)$ is defined as the index of the user preferred genre for which the user's expected rating is the highest. These expected scores are computed from the user's previous genre ratings. Here we assume that the user has rated mostly movies with the genres that he/she prefers since the ratings in the dataset are collected for movies that the user chose to see, and not based on our recommendation. For example, if a user u has rated 45 Drama movies and his/her average rating for these movies is 3.82, while the average ratings for Action and Comedy are lower, we set $\mathrm{gC}(u) = \mathrm{Id}_{\mathrm{Drama}} = 1$.

3.4 Construction of predictor variables

To allow the explanation of the contributions of the three beliefs (behavioral, normative and control, see Subsec. 1.1 and Fig. 1) of the TPB we decided on the hierarchical model. As depicted in Fig. 1, we fit the following models:

1. Each of the three beliefs is regressed to a score showing the contribution of each of the beliefs to the selection of behavior (movie with a given genre). The criteria variable used is gS and this yields nine models. In these models the predictor variables are the answers to questions assigned to the modeled belief;
2. Aggregated model: the prediction of the scores for each of the three beliefs obtained (in the previous step) are used as predictor variables to model the selected behavior. We introduce the aggregated model in order to estimate the relative effect of each of the three sub-models to the analyzed behavior. This is required when the next version of the TPB questionnaire is constructed (balancing among the end user effort when answering the questionnaire).

The regression model we selected depends on the criteria variables used. For genre scores gS we selected multiple linear regression (MVR) as the first choice for linear continuous variable prediction model, and for genre membership we used linear discriminant analysis as an optimal linear classifier.

Factors that affect the decision for each of selected genres may vary (e.g. for selecting drama the main actor may be important, while for selecting actions the movie director may be important). Hence, we decided to use multivariate models with different slopes, (i.e. different regression coefficients for each of the selected behaviors), which yields the triple (s_D, s_A, s_C) computed by inserting the users' answers into the MVR models for the genres Drama, Action and Comedy, respectively.

4 Results

For each of the six models that we fitted and tested (three models from attributes, one from norms and control, and the top level aggregate model) we fitted the MVR model (resulting in the model coefficients β_k and the proportion of the explained variances R^2) and we performed the linear discriminant analysis (resulting in the discriminant weights w_k and the separability s in terms of the Fisher discriminant analysis [19]). We do not report the results for all six models, but only for the cognitive attributes (selected for demonstrating how to interpret the results) and for the aggregate model which summarizes the whole TPB model. Since this is the initial version of the TPB questionnaire, the analysis of within-questionnaire correlations needs to be performed. We list maximal and typical correlations showing that our questionnaire needs to be upgraded to assure the required low level of correlations, see also 4.3.

4.1 Selected sub-model: the cognitive dimension of attitudes

This sub-model explains the role and contribution of the user's answers to eight questions $Q_1 - Q_8$ regarding the cognitive dimension of attitudes (i.e. the selection of a movie of one of the selected genres). In Fig. 1, which depicts the aggregated model, this sub-model is located on the top of the three sub-models explaining the user's cognitive dimension of attributes.

The proportion of the explained variance for the cognitive dimension of attitudes is $R^2 = 0.48$. This high value allows us to interpret the beta coefficients of the model and determine the relative importance of the cognitive attitude in the overall mode as discussed in Subsec. 4.2.

The normalized MVR model coefficients are listed at Tab. 1. The maximal correlation between answers in a cognitive attitude group of question is $r = 0.49$, typical ones are $r \sim 0.1$. The coefficient β_0 represents the offset of the resulting model, while the coefficients β_k, $1 \leq k \leq 8$ correspond to the questions Q_k. The significant β coefficients at risk level $\alpha = 0.05$ are marked with *. We

Genre	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Drama	0.32	0.02	-0.03	-0.01	-0.02	0.01	0.00	-0.01	0.00
Action	3.43	0.09	-0.11*	0.10*	0.05	-0.08	0.13*	-0.02	-0.02
Comedy	3.69	0.48*	-0.34*	0.05	0.07	-0.04	-0.12*	-0.11*	0.06

Table 1. MVR coefficients of the cognitive dimension of the attitudes predictors, $R^2 = 0.48$.

observe that none of the coefficient that model the genre Drama is significant and therefore no conclusion can be made here. This is most probably due to the relatively low sample size used to fit the model. Regarding the genre Action, the coefficients representing Q_2 =*How important for you is the story in the movie?*, Q_3 =*How important for you is the movie's genre?* and Q_6 =*How important for you are the special effects?* are significant. Since β_2 is negative, the users that do not care much about the story of the movie are more likely to select the Drama genre. The positive coefficients β_3 and β_6 show that the users that cared about the genre and the special effects are more likely to select the Drama genre.

In the same way we interpret the selection of Comedy movies. The large positive coefficient $\beta_1 = 0.48$ representing Q_1 =*How important for you is the main* *actor of the movie?* indicates that the main actor is the most important factor in selecting the Comedy genre, while the coefficient β_2 representing Q_1 =*How important for you is the story in the movie?* indicates that the story has very little relevance in selecting the Comedy genre. The coefficients $\beta_6 = -0.12$ and β_7 representing the questions Q_6 =*How important for you are the special effects?* and Q_7 =*How important is an attractive trailer?*, respectively, indicate the low relevance of the special effects and of the trailer in the selection of the Comedy genre.

We analyzed the separability of the genre selection behaviors by LDA. The Fisher separability of the cognitive dimension of attitudes is $s = 0.81$ which means a moderate separability. The LDA coefficients separating the given pairs of genres are listed in Tab. 2. The significant coefficients at the risk level $\alpha = 0.05$ are labeled by *. A significant contribution to the separation of the genres Drama and Action are obtained from Q_3 and Q_6 etc.

Genre pair	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
Drama/Action	-0.87	-0.82	-3.26*	-0.66	-0.42	-3.54*	1.68	-0.01
Drama/Comedy	-5.67*	-0.49	0.94	0.09	-1.48	-1.23	1.19	-1.56
Action/Comedy	-4.80*	0.32	4.19*	0.75	-1.07	2.31	-0.49	-1.55

Table 2. Linear discriminant coefficients of the cognitive dimension of attitudes predictors.

To summarize the interpretation, the questions and underlying decision factors that are relevant in both models (MVR and LDA) are regarded as the most important. These are the Q_1, Q_3 and Q_6 factors. The question Q_1 is the most important one since it is involved in the discriminant function with highest magnitudes.

4.2 Aggregated TPB model

We computed the scores predicted by each of the beliefs, norm and control models and used them as a stand-alone predictors for the three behaviors to build a hierarchical model. Each of the underlying models contributed three scores, one for each of the behaviors. The obtained aggregated model achieves a large proportion of explained variance $R^2 = 0.89$ and a good Fisher separability value of 1.008. The aggregated model with R^2 of the sub-models are depicted in Table 3.

We do not list and interpret the coefficients of MVR and LDA for the aggregated model here. We summarize the whole model by a list of the explained variances and Fisher separabilities in Tab. 3. All the listed R^2 values including the aggregated $R^2 = 0.89$ are statistically significant and they indicate the relative weight of each sub-model in the movie selection of the genre. Behavior control contributed the least and the cognitive aspect of attributes contributed the most to the whole model. Note that for the sake of simplicity we did not

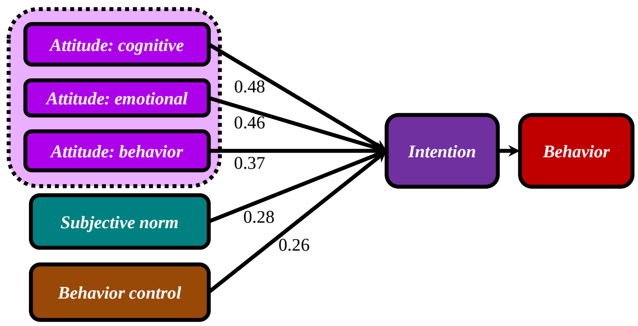

Fig. 1. Aggregated model of genre selection, $R^2 = 0.89$.

regress behavior intentions but directly the behaviors themselves. The models allows to regress the intentions but this it is beyond the scope of this paper. The listed R^2 was estimated simply by direct estimation of the aggregated model's R^2.

	Attr: cog.	Attr: emot.	Attr: behav.	Norms	Cont.
$\mathbf{R}^2$	0.49	0.46	0.37	0.28	0.26
Separability	0.81	1.03	0.60	0.44	0.50

Table 3. Proportion of the Explained variance R^2 and Fisher separability for the aggregated model.

4.3 The correlation structure of the questionnaire

In Subsec. 4.1 we listed maximal correlation among end user's answers regarding cognitive attitude. The value is much to high to allow the stable interpretation of estimated TPB model weights (betas). The same is true for the correlations of other groups of questions (emotion attitude maximal corr. $r = 0.69$, behavior attitude maximal corr. $r = 0.81$, subjective norm maximal corr. $r = 0.65$ and behavior control maximal corr. $r = 0.69$).

On the other hand, the majority of these correlations are relatively low. The factor analysis applied separately to each three groups of questions (attitude, norms, controls) revealed a simple structure allowing us to remove a small number of highly correlated questions while preserving the assumed aspects of end users behavior. Due to the space limitations, we do not list factor matrices here.

5 Discussion

The main goal of this paper was to introduce the Theory of Planned Behavior into user modeling. Below we discuss the most relevant issues regarding the advantages, limitations, and further development of the RS backed with psychology-based research.

What are the benefits of implementing the TPB in user modeling? One of the main reasons for the application of the TPB in a given domain is to further extend the understanding of underlying mechanisms that govern the way that users make their decisions. In the field of user modeling, this understanding relates to two aspects. The first one is understanding user adaptation as a whole (for example, what are the relevant factors in movie-item selection) and can be summarized from the results of the prediction model fitting. The second one is about the individual user's mechanism (what are the individual factors in these selections for a given user) and we explain it from the individual user's responses to the survey questions, together with the model-fitting results.

The explanation of the TPB model is very dependent on the selected prediction model (MVR in our case). This means that models with little or no explanation power (black boxes) are of less interest in a TPB.

When the TPB is applicable? For statistical and machine-learning reasons, the first requirement for the successful application of the TPB is that the predefined behaviors are separable by the selected model. In our case, genres that are not separable by the user's attributes, norms and beliefs cannot be well modeled and the fitted model would provide misleading results.

The separability of modeled behaviors also limits the number of these behaviors. It is clear that in practice several hundreds or thousands of behaviors cannot be separable (the user data acquisition would not tolerate it in the first place). This leads to an important limitation of the TPB in user modeling, meaning that the treatment of individual items cannot be defined as a behavior, but that these items must be grouped in a smaller number of groups or the definition of behaviors is based on a completely different aspect of user adaptation.

What are the options for integrating TPB models into a RS? The role and integration of the TPB into a user-adaptation procedure is mainly determined by the definition of behaviors. As already indicated, user Actions related to individual content items as behaviors are not a good choice.

What are the concerns of TPB user data acquisition? The theory and practice of the TPB shows that the surveys required to fit the TPB model accurately enough are relatively long and they also demand a considerable effort from the respondent (end user) to provide relevant answers. In the context of user modeling, this means that the user-data acquisition is relatively intrusive. On the other hand, since the user's attributes, norms and beliefs are changing very slowly with time, it is sufficient for the user to complete the survey only once a year. However, the sampling period may vary significantly according to the domain and also according to an individual user's practice. In our example, attributes, norms and beliefs toward movie-genre selection may change faster for those users who see more movies in a given amount of time.

Does the TPB allow cross-domain user modeling? The cross-domain of user-adaptation techniques is of great interest. The question is can TPB models, in particular the Ajzen model, ensure cross-domain capabilities in terms that the attributes, norms and beliefs of the end user estimated in one domain (for example, movie selection) are at least in part valid for the other domain (for example, tourist-destination selection). Unfortunately, in general the answer is no. The reason for this is simply the fact that the survey used to estimate these attributes, norms and beliefs must be very specifically related to the domain of behaviors. For instance, the relevance of certain factors is asked for movies or for tourist destinations and not about some general user opinion common to both domains. However, the research on life-styles [10] indicates that there are strong relations among human behaviors in different domains.

What are benefits of using the TPB as user modeling technique? After the above listed considerations one could argue what are the benefits of the introduction of TPB that are not available from advanced statistics and machine learning algorithms. We see the following benefits of TPB in user modeling domain:

1. *Explanation of the underlying mechanisms.* A deeper understanding of the processes accompanying user adaptation usually leads to more effective adaptation procedures, more appropriate evaluation measures and procedures, and fresh ideas about how to implement the user-adaptation results for end users;

2. *Guidelines for survey-question formulations.* According to the previous point reasoning, the TPB further provides explicit guidelines for user-data-acquisition survey construction. It is important to note that there is a large number of data-driven studies in several domains that support the theory of the TPB and the description of behavioral, normative and control beliefs. This allows us to construct more effective surveys, resulting in more accurate user data at the same level of intrusion;

3. *Study of cross-domain user adaptation.* Cross-domain user adaptation is one way to reduce the intrusion of user-data acquisition and to design more effective user-adaptation techniques. As already indicated, no cross-domain of the estimated beliefs in the TPB is guaranteed. However, the studies related to life-styles shows the potential to link the correlated user-behavior patterns in a way that allows us to make conclusions about end-user beliefs from the original to the correlated domain. This is related to our future work plans.

6 Conclusion and further work

The work presented in this paper aims at establishing the relevance of the psychological human-decision modeling Theory of Planned Behavior (TPB) into the field of user modeling. The study contributes to the models that are applicable in user modeling, in particular to the explanation of these models.

Our results show that the application of the TPB in the area of recommender systems allows a further insight into the underlying process of the user's decision making, i.e., into factors that affect these decisions. These insights can be used to address several issues, such as effective user-data acquisition, understanding and mitigating the reasons for unacceptable recommendations, etc. As an important part of this research performed by an interdisciplinary team, including engineers, mathematicians, and psychologists, are the guidelines for the future applications of the TPB in different areas of user modeling. They include behavior selection, user-questionnaire construction, criteria variable construction, regression-model selection and fitting, and an explanation of the obtained results.

Despite the limitations of the proposed modeling, our study showed that such modeling improves our understanding of the user-adaptation process. It is not meant as a replacement for the existing user-modeling models (for example, Matrix Factorization in movie recommendations) but as a predictor of end-user behaviors that affects the whole process. Such behaviors influence the selection of the device he/she uses to consume the recommended service, etc. Furthermore, in the discussion section we addressed several issues relevant for the application of the TPB in the user-modeling domain.

The obvious further work is the application of the upgraded TPB questionnaire, confirming its validity and reporting the TPB model results in terms of the explanation of why users selects movie genres as they do. In this way, we allow the next development steps as indicated in the discussion.

References

1. Ajzen, I.: TPB homepage, accessed 10.01.2014., `http://people.umass.edu/aizen/index.html`
2. Ajzen, I.: The theory of planned behavior. Organizational Behavior and Human Decision Processes 50(2), 179–211 (Dec 1991)
3. Ajzen, I.: Consumer attitudes and behavior. In: Curtis, Paul, Frank (eds.) Handbook of consumer psychology, pp. 525–548. Lawrence Erlbaum Associates (2008)
4. Ajzen, I., Manstead, A.S.R.: Changing health-related behaviours: An approach based on the theory of planned behaviour. In: Hewstone, M., Schut, H.A.W., De Wit, J.B.F., Van Den Bos, K., Stroebe, M.S. (eds.) The scope of social psychology: Theory and applications, pp. 43–63. Psychology Press (2007)
5. Balabanović, M., Shoham, Y.: Fab: content-based, collaborative recommendation. Communications of the ACM 40(3), 66–72 (1997)
6. Bamberg, S., Ajzen, I., Schmidt, P.: Choice of Travel Mode in the Theory of Planned Behavior: The Roles of Past Behavior, Habit, and Reasoned Action (2003)
7. Bennett, J., Lanning, S.: The netflix prize. In: Proceedings of KDD Cup and Workshop. vol. 2007, p. 35 (2007)
8. Daigle, J.J., Hrubes, D., Ajzen, I.: A Comparative Study of Beliefs, Attitudes, and Values Among Hunters, Wildlife Viewers, and Other Outdoor Recreationists (2002)
9. Davis, L.E., Ajzen, I., Saunders, J., Williams, T.: The decision of African American students to complete high school: An application of the theory of planned behavior. (2002)

10. Dewberry, C., Juanchich, M., Narendran, S.: Decision-making competence in everyday life: The roles of general cognitive styles, decision-making styles and personality. Personality and Individual Differences 55(7), 783–788 (Oct 2013)
11. Fernández-Tobías, I., Cantador, I., Kaminskas, M., Ricci, F.: Cross-domain recommender systems: A survey of the state of the art (2012)
12. Fraser, R.T., Johnson, K., Hebert, J., Ajzen, I., Copeland, J., Brown, P., Chan, F.: Understanding employers' hiring intentions in relation to qualified workers with disabilities: preliminary findings. Journal of occupational rehabilitation 20(4), 420–426 (2010)
13. Icek Ajzen, I.A.: Job Satisfaction, Effort, and Performance: A Reasoned Action Perspective (2011)
14. Koren, Y.: Factorization Meets the Neighborhood : a Multifaceted Collaborative Filtering Model. In: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 426–434 (2008)
15. LDOS: Comoda context rich movie recommender dataset (Jul 2009), `http://212.235.187.145/spletnastran/raziskave/um/comoda/comoda.php`
16. Lee, J., Lee, M.: Factors Influencing the Intention to Watch Online Video Advertising (Oct 2011)
17. Lin, F.R., Lu, I.Y., Hsieh, P.S.: Understanding the Adoption of Wireless Sensor Network Service in Households. In: 2011 International Joint Conference on Service Sciences. pp. 218–222. IEEE (May 2011)
18. Pazzani, M., Billsus, D.: Content-based recommendation systems. The adaptive web pp. 325–341 (2007)
19. Rencher, A.C., Christensen, W.F.: Methods of Multivariate Analysis (Wiley Series in Probability and Statistics). Wiley, 3 edn. (7 2012)
20. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM conference on Computer supported cooperative work. pp. 175–186. ACM (1994)
21. Tetard, F., Collan, M.: Lazy user theory: A dynamic model to understand user selection of products and services. 2013 46th Hawaii International Conference on System Sciences 0, 1–9 (2009)
22. Walsh, S.P., White, K.M., Hyde, M.K., Watson, B.: Dialling and driving: factors influencing intentions to use a mobile phone while driving. Accident; analysis and prevention 40(6), 1893–900 (Nov 2008)

Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal

Bruce Ferwerda and Markus Schedl

Department of Computational Perception,
Johannes Kepler University (JKU),
Altenberger Straße 69, A-4040 Linz, Austria
bruce.ferwerda@jku.at,markus.schedl@jku.at

Abstract. This position paper describes the initial research assumptions to improve music recommendations by including personality and emotional states. By including these psychological factors, we believe that the accuracy of the recommendation can be enhanced. We will give attention to how people use music to regulate their emotional states, and how this regulation is related to their personality. Furthermore, we will focus on how to acquire data from social media (i.e., microblogging sites such as Twitter) to predict the current emotional state of users. Finally, we will discuss how we plan to connect the correct emotionally laden music pieces to support the emotion regulation style of users.

Keywords: Music recommender systems, personality, emotional states, emotion regulation

1 Introduction

Research on recommender systems have shown increased interest to incorporate psychological aspects. Especially the relationship between personality and user preferences has gained a lot of attention. For example, knowledge about the influence of personality traits on music taste [25], and diversity in item recommendations [39] have been exploited to improve the user tailored recommendation. As personality is defined as the individual differences in enduring emotional, interpersonal, experiential, attitudinal and motivational styles [12, 17], one can expect to be able to infer much more based on personality traits to improve the recommendation.

The goal of this project is to improve music recommendations by incorporating additional psychological factors. More specifically, we focus on emotional states and their relationship with personality to infer music taste and preferences. By knowing the user's current emotional state, a system can anticipate its recommendation with an emotionally laden song that is in line with the user's style of emotion regulation (e.g., changing or maintaining their emotional state).

In the following sections we give a brief introduction about what is known about personality and emotional states, and work towards how we are planning to use it to improve recommendations.

2 Personality

Personality has shown to be an enduring factor that influences an individual's behavior [13], interest, and tastes [14, 25]. As personality plays such a prominent role in shaping human preferences, one can expect similar patterns (i.e., behavior, interest, and tastes) to emerge between similar personality traits [2].

Different models have been created to categorize personality, where the five-factor model (FFM) is most well known and widely used [17]. The FFM consists of five general dimensions that describe personality. Each of the five dimensions consist clusters of correlated primary factors. Table 1 shows the general dimensions with the corresponding primary factors.

General dimensions	Primary factors
Openness	artistic, curious, imaginative, insightful, original, wide interest
Conscientiousness	efficient, organized, planful, reliable, responsible, thorough
Extraversion	active, assertive, energetic, enthusiastic, outgoing, talkative
Agreeableness	appreciative, forgiving, generous, kind, sympathetic, trusting
Neuroticism	anxious, self-pitying, tense, touchy, unstable, worrying

Table 1. Five-factor model

There is an emerging interest in how personality relates to user preferences in different domains. This provide valuable information for the development of domain specific recommender systems. Knowing someone's personality can help to infer their preferences [23–25], and can therefore contribute to a more accurate recommendation. For example, music preferences were found to be correlated with personality traits [25]. Rentfrow and Gosling [25] categorized music pieces into 4 music-preference dimensions (reflective and complex, intense and rebellious, upbeat and conventional, and energetic and rhythmic), and found correlations with the FFM general dimensions, such as, a relation between energetic and rhythmic music and extraversion and agreeableness.

The prediction of personality parameters is starting to establish by either using implicit acquisition (e.g., personality prediction by extracting data from social media [8, 16, 22]), or explicit acquisition by letting users answering a personality quiz [10]. Although the implicit method is unobtrusive, accuracy is compromised as it dependents on the quality of the source (e.g., frequency of expressing on social media). On the other hand, the explicit method is more accurate, but intrusive and time consuming.

3 Emotional States

We can find emotions in every facet of our life, such as during: decision making, objective and subjective thinking, creativity. To categorize the emotional states we experience, Ekman [5] defined six basic emotional states in which we can categorize experienced emotion: anger, disgust, fear, happiness, sadness, and surprise. Others on the other hand believe that emotions are a mix of dimensions of emotional states [35].

To deal with our emotional states throughout the day, we adapt different strategies. Parkinson and Totterdell [21] defined 162 different strategies (e.g., exercising, music listening, taking a bath). Especially listening to music plays an important role. Research has found that music is the second most strategy used [36, 7]. It can change, create, maintain, or enhance emotions [3]. This suggest that music can play an important supportive role when people dealing with their emotions in daily life.

Just as with personality, there is also an implicit (e.g., blog text) and an explicit way to detect emotion with the same drawbacks. Although the implicit detection has advanced, it goes without saying that automatic capturing of online emotional states remain challenging. As Scherer [28] noted "The inherent fuzziness and the constant evolution of language categories as well as inter-language, inter-cultural, and inter-individual differences make it difficult to define central working concepts".

4 Personality & Emotional States

How we regulate our emotions have been investigated with relation to our personality. Of particular interest are the neuroticism and extraversion dimensions. These dimensions are associated with experiencing negative and positive affect consistently. For example, Tamir [32] found that people scoring high on the neuroticism dimension tend to increase their level of worry. Similarly, people who score low on the extraversion dimension tend to be less motivated to increase their happiness [33].

While most studies are focusing on personality traits in relation to emotion regulation, there is a small area that argues that the emotion regulation style can be explained by one's implicit theory of emotion. In other words whether someone beliefs that emotions are fixed (entity theorist), or more malleable (incremental theorist). Entity theorists experience more negative emotions, that is, less favorable emotion experiences, lower well-being, greater depression, more loneliness, and poorer social adjustments compared to incremental theorists [34].

Music has the ability to induce intense emotions (positive and negative) [40]. Some studies have investigated how, and whether the emotion that consist in music is used by people in their emotion regulation. Thoma et al. [38] categorized different music pieces on valence and arousal, and found that different pieces were preferred depending on the emotionally laden situation. Similarly, Van Goethem and Sloboda [7] found that people use music to support their regulation strategy. For example, music is used to help to distract from the affect or situation, or can help to think about it in a rational way. Despite findings on an individual level (i.e., personality) and how music is used as a regulation strategy, there is still a gap in connecting these two. That is, it is still unknown how music is used to regulate emotions on an individual level.

5 How to improve recommendations?

As music plays a role in emotion regulation of people, and the way how people regulate their emotions seem to be dependent on their personality (or their implicit theories of emotion), the music that people use to support their emotion regulation may also be dependent on their personality.

Whereas personality is usually used to alleviate the cold-start problem in recommender systems (i.e., new users and sparse data sets) [11], or to determine the amount of diversity in the recommendation [39], including emotional states can help to improve music recommendations on the fly. Currently, music recommender systems anticipate their subsequent recommendations on the music that the user currently listens to. The recommendation is based on similarity by comparing what others with similar taste have listened before (collaborative filtering), or by matching properties (e.g., genre, artist) of the music pieces (content-based filtering). This can result in that recommendations given may fit the user's taste, but may not match the user's actual *needs* at that moment. For example, the systems knows that a user likes Beyoncé. Beyoncé has a range of different emotionally laden songs from up-tempo to ballads. By knowing the user's emotion at a specific moment, the recommender system can anticipate and propose a piece of emotion-laden music that lies within the taste of the user that can support the regulation of the experienced emotion.

5.1 A Scenario

Anna is a 22 year old student. When she listens to music, she often makes use of an online radio. This online radio knows Anna's taste so it can anticipate on the next song to play for Anna. Besides knowing Anna's taste, the system knows that Anna is a little bit neurotic.

On one day Anna is at home, listening to an up-tempo song of Beyoncé. The next song that the radio put in the cue is another up-tempo song, but this time by Katy Perry. Suddenly Anna receives some bad news that makes her sad. She post her feelings on Twitter. The radio system notices this and based on her personality (neuroticsm), it adjust the song in the cue. Instead of playing an up-tempo song, it replaces it with a sad song of Katy Perry. By knowing how Anna likes to regulate her emotions, the system can anticipate the play-list accordingly.

6 Proposal

In the following sections we discuss the initial ideas that we have to improve recommender systems by incorporating the user's personality and current emotional state. We start with describing how we plan to investigate the relationship between personality and emotion regulation through music. After that we discuss the methods for the personality and emotion acquisition from social media, and finally we discuss how we plan to find the emotionally laden songs. For the incorporation of emotion and personality, we assume that system already initiated the user's music taste.

6.1 Step 1: User Study

The first step would be to investigate how people prefer to regulate their emotions and the relationship with their personality. For example, people scoring high on neuroticism tend to increase their level of worry [32]. Therefore, they may not want to listen to music that tries to change their worry state, but want music that is in line with that state instead.

Although there is much research done on the inducing effect of emotional laden music [4, 27], not much is known about how people use music to regulate their emotions. We plan to conduct a online user study using participants on Amazon Mechanical Turk. In this user study we will use the set of film clips (see Figure 1 for the experiment work flow), developed by Hewig et al. [9] to induce one of the basic emotional states. Presenting film clips is one of the methods that is frequently used to induce emotions in psychological experiments. For the user study, we will assign participants randomly to a emotionally laden film clip (anger, disgust, fear, happiness, sadness, or neutral). After showing participants the film clip, we will ask as a control the emotion the film clip induced. In the next step we will present different emotionally laden music fragments (anger, fear, happiness, sadness, and tenderness) and ask participants the likelihood that they would listen to such music when being in the just induced emotional state. The music fragments we will use are categorized by Eerola and Vuoskoski [4] based on the basic emotions they bear. As a control question we will ask in addition what kind of emotion the music pieces induce. To conclude we will ask the FFM questions, implicit theory of emotions questions, and demographics (i.e., age and gender). This will give us information about how music is used in different emotional states and how this is related to personality traits.

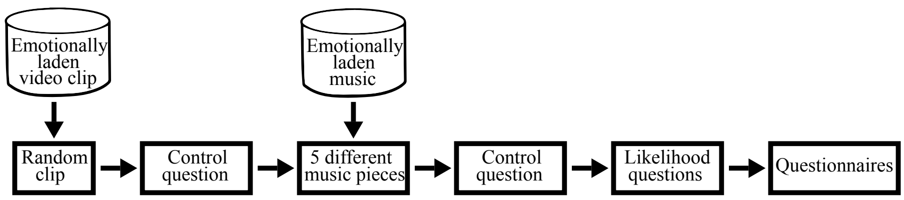

Fig. 1. User study work flow

With the results of the aforementioned user study we will create a model to predict the emotionally laden music pieces that users would like to listen to when in a certain emotional state. A second user study will be carried out to create the dataset for testing the model. The dataset will contain data about their current emotional states of users and the emotional laden songs they want to listen to. A 10-fold cross-validation method will be used for validation.

Once the model is created and verified, we will know how people prefer to regulate their emotional states with music throughout the day, and how this is related to their personality. In the next step we will move toward the extraction of personality and emotional states from social media.

6.2 Step 2: Personality & Emotional State Acquisition

We will use Twitter as our main source to extract the personality and emotional state parameters from. Tweets are crawled by using the Twitter API. Furthermore, we will limit ourselves to tweets with English as the main language.

Personality Acquisition For the acquisition of personality we will work toward an implicit detection of the parameters, i.e., without the need of a questionnaire. Results of previous research of extracting personality parameters from tweets are promising. Golbeck et al. [8] were able to predict personality parameters from Twitter within 11%-18% of their actual value by looking at the content of users' tweets. Other research done by Quercia et al. [22] were able to estimate personality parameters (RMSE below 0.88 on a [1, 5] scale) by only looking at the users' characteristics (e.g., listeners, popular, highly-read, and influential users). We plan to explore the techniques used by prior research and possibly combining them to improve predictions. Another direction that may be worth taking into account would be to incorporate historical tweets that reflect listening behavior of users. Rentfrow and Gosling [25] found relations between personality and music genres. By looking at historical music tweets of users, we are able to extract the genre of the song which in turn can provide us personality information.

Emotional State Acquisition Although we realize that emotional states are not expressed constantly, we do believe that social media is a platform that is increasingly used by users to express themselves. This includes emotional states depicting personal (e.g., anger, frustrations) to global topics (e.g., politics, sport events) [1, 37].

The acquisition of the emotional states from textual collections of user-generated data on the web has been well established (for an overview of this field see [20]). Results indicate that emotional indicators can be extracted accurately. However, acquisition of these indicators from microblogging sites has been done scarcely. Most of the studies focus on the polarity (positive, negative, or neutral) [20] or try to include the magnitude of the emotion (mild and strong) [30]. Only a few have tried to categorize microblogging text based on existing emotional categorizations [1, 29].

One approach that we bear in mind that to build upon is the use of emotion lexicons. These lexicons consist of terms related to an emotion. Several lexicons have been created based on different emotion categorizations and have been tested on tweets. Such as, Sintsova, Musat, and Pu [29] created a lexicon compatible with the Geneva Emotion Wheel categorization of emotions, Roberts et al. [26] based their emotion lexicon on Ekman's categorization, and Suttles, and Ide [31] on Plutchik's. Especially the work of Roberts et al. [26] would be suitable to build upon as the Ekman's categorization is on the basis of our work. We will be able to complement the predictability of emotions by including metadata that have shown to consist of emotional indicators, for example, hashtags [18, 19], traditional emoticons [6], and emoji [26].

6.3 Step 3: Emotion Classification of Songs

The user study (see §6.1) will give us insights in how emotionally laden songs are used in the emotion regulation process. For the system to be able to anticipate its recommendation, we need to find the right emotional annotated song.

The field of emotion classification in music is still evolving (see for an overview [15]). Currently, different methods are used to annotate music pieces on their emotion: direct human annotation (e.g., surveys, social tags, games), indirect human annotation (e.g., web documents, social tag clouds, lyrics), or content-based analysis (e.g., audio, images, video). As Kim et al. [15] noted "Recognizing musical mood remains a challenging problem primarily due to the inherent ambiguities of human emotions." To find the right emotionally laden song within a collection in this project, we will initially turn to the tags provided by Last.FM website. Last.FM currently provide songs with the tags happy, sad, angry, and relaxed. Based on these tags we can make a first attempt to match emotionally laden songs with the user's way of regulating their emotional state.

7 Conclusion

By including the user's current emotional state, we propose that music recommendations can be improved. Our next efforts will be to investigate how people regulate their emotions with music. That is, what kind of emotionally laden music people are listening when being in a specific emotional state. Additionally, we will investigate how emotion regulation with music is related to their personality.

For the acquisition of personality and emotion, we will focus on microblogging sites. As social media generates a constant stream of communication, we believe that microblogging sites as Twitter are suitable to extract personality and emotional states of users. Although accurate results are achieved from textual collections of user-generated data on the web, analyzing microblogging sites remains challenging. The amount of text is scarce as the text posted on Twitter is limited to 140 characters. However, the ability to express oneself in a short and fast way lend itself to post content more often. To extract personality and emotional states from Twitter feeds we will initially trust on different existing methods and combine them to improve predictability of the parameters.

With the findings of the aforementioned steps, we can start matching music that fits the user's way of emotion regulation. To find suitable music, we will initially rely on the emotional tags that Last.FM provides.

8 Acknowledgment

This research is supported by the European FP7/2007-2013 program through the PHENICX project under grant agreement no. 601166 and by the Austrian Science Fund (FWF): P25655.

References

1. Bollen, J., Mao, H., Pepe, A.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: ICWSM (2011)
2. Burger, J.: Personality. PSY 235 Theories of Personality Series, Cengage Learning (2010)
3. DeNora, T.: Music as a technology of the self. Poetics 27(1), 31–56 (1999)
4. Eerola, T., Vuoskoski, J.K.: A comparison of the discrete and dimensional models of emotion in music. Psychology of Music 39(1), 18–49 (2011)
5. Ekman, P.: An argument for basic emotions. Cog. & Emotion 6, 169–200 (1992)
6. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford pp. 1–12 (2009)
7. van Goethem, A., Sloboda, J.: The functions of music for affect regulation. Musicae scientiae 15(2), 208–228 (2011)
8. Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting personality from twitter. In: Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom). pp. 149–156. IEEE (2011)
9. Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., Bartussek, D.: Brief report: A revised film set for the induction of basic emotions. Cognition & Emotion 19(7), 1095–1109 (2005)
10. Hu, R., Pu, P.: Using personality information in collaborative filtering for new users. Recommender Systems and the Social Web p. 17 (2010)
11. Hu, R., Pu, P.: Enhancing collaborative filtering systems with personality information. In: Proceedings of the fifth ACM conference on Recommender systems. pp. 197–204. ACM (2011)
12. John, O.P., Srivastava, S.: The big five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research 2, 102–138 (1999)
13. Jung, C.: Psychological Types. No. v. 6 in Bollingen Series No. 20, Princeton University Press (1976)
14. Kemp, A.E.: The musical temperament: Psychology and personality of musicians. Oxford University Press (1996)
15. Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A., Turnbull, D.: Music emotion recognition: A state of the art review. In: Proc. ISMIR. pp. 255–266. Citeseer (2010)
16. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences 110(15), 5802–5805 (2013)
17. McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality 60(2), 175–215 (1992)
18. Mohammad, S.M.: # emotional tweets. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation. pp. 246–255. Association for Computational Linguistics (2012)
19. Mohammad, S.M., Kiritchenko, S.: Using hashtags to capture fine emotion categories from tweets. Computational Intelligence (2014)
20. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and trends in information retrieval 2(1-2), 1–135 (2008)

21. Parkinson, B., Totterdell, P.: Classifying affect-regulation strategies. Cognition & Emotion 13(3), 277–303 (1999)
22. Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: 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. IEEE (2011)
23. Rawlings, D., Ciancarelli, V.: Music preference and the five-factor model of the neo personality inventory. Psychology of Music 25(2), 120–132 (1997)
24. Rentfrow, P.J., Goldberg, L.R., Zilca, R.: Listening, watching, and reading: The structure and correlates of entertainment preferences. Journal of personality 79(2), 223–258 (2011)
25. Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlates of music preferences. Journal of personality and social psychology 84(6), 1236 (2003)
26. Roberts, K., Roach, M.A., Johnson, J., Guthrie, J., Harabagiu, S.M.: Empatweet: Annotating and detecting emotions on twitter. In: LREC. pp. 3806–3813 (2012)
27. Scherer, K.R.: Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them? Journal of new music research 33(3), 239–251 (2004)
28. Scherer, K.R.: What are emotions? and how can they be measured? Social science information 44(4), 695–729 (2005)
29. Sintsovaa, V., Musata, C.: Fine-grained emotion recognition in olympic tweets based on human computation. WASSA 2013 12 (2013)
30. Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In: Proceedings of the 2008 ACM symposium on Applied computing. pp. 1556–1560. ACM (2008)
31. Suttles, J., Ide, N.: Distant supervision for emotion classification with discrete binary values. In: Computational Linguistics and Intelligent Text Processing, pp. 121–136. Springer (2013)
32. Tamir, M.: Don't worry, be happy? neuroticism, trait-consistent affect regulation, and performance. Journal of personality and social psychology 89(3), 449 (2005)
33. Tamir, M.: Differential preferences for happiness: Extraversion and trait-consistent emotion regulation. Journal of Personality 77(2), 447–470 (2009)
34. Tamir, M., John, O.P., Srivastava, S., Gross, J.J.: Implicit theories of emotion: affective and social outcomes across a major life transition. Journal of Personality and Social Psychology 92(4), 731 (2007)
35. Thayer, R.E.: The biopsychology of mood and arousal. Oxford University Press (1989)
36. Thayer, R.E., Newman, J.R., McClain, T.M.: Self-regulation of mood: strategies for changing a bad mood, raising energy, and reducing tension. Journal of personality and social psychology 67(5), 910 (1994)
37. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment in twitter events. Journal of the American Society for Information Science and Technology 62(2), 406–418 (2011)
38. Thoma, M.V., Ryf, S., Mohiyeddini, C., Ehlert, U., Nater, U.M.: Emotion regulation through listening to music in everyday situations. Cognition & emotion 26(3), 550–560 (2012)
39. Wu, W., Chen, L., He, L.: Using personality to adjust diversity in recommender systems. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. pp. 225–229. ACM (2013)
40. Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion 8(4), 494 (2008)

How are you doing? Emotions and Personality in Facebook

Golnoosh Farnadi[1,2,3], Geetha Sitaraman[2], Mehrdad Rohani[2], Michal Kosinski[4], David Stillwell[4], Marie-Francine Moens[3], Sergio Davalos[2], and Martine De Cock[1,2]

[1]Dept. of Appl. Math., Comp. Science and Statistics, Ghent University, Belgium
[2]Center for Data Science, University of Washington Tacoma, WA, USA
[3]Dept. of Computer Science, Katholieke Universiteit Leuven, Belgium
[4]The Psychometrics Centre, University of Cambridge, Cambridge, UK
{gfarnadi,sgeetha,rohanim,sergiod,mdecock}@uw.edu
{mk583,ds617}@cam.ac.uk
{golnoosh.farnadi,sien.moens}@cs.kuleuven.be
{golnoosh.farnadi,martine.decock}@ugent.be

Abstract. User generated content on social media sites is a rich source of information about latent variables of their users. Proper mining of this content provides a shortcut to emotion and personality detection of users without filling out questionnaires. This in turn increases the application potential of personalized services that rely on the knowledge of such latent variables. In this paper we contribute to this emerging domain by studying the relation between emotions expressed in approximately 1 million Facebook (FB) status updates and the users' age, gender and personality. Additionally, we investigate the relations between emotion expression and the time when the status updates were posted. In particular, we find that female users are more emotional in their status posts than male users. In addition, we find a relation between age and sharing of emotions. Older FB users share their feelings more often than young users. In terms of seasons, people post about emotions less frequently in summer. On the other hand, December is a time when people are more likely to share their positive feelings with their friends. We also examine the relation between users' personality and their posts. We find that users who have an open personality express their emotions more frequently, while neurotic users are more reserved to share their feelings.

Keywords: Emotion detection, NRC lexicon, User modelling, Big Five Personality model, Facebook, Social media, Time factor

1 Introduction

As more and more users are creating their own content on the web, there is a growing application potential for personalization in human computer systems such as personalized information access services, recommender systems, and tailored advertisements. To provide personalization, a wide variety of user information can be processed. Previous personalization research has focused on explicit user characteristics such as demographics, e.g., age, gender, location, or language [1]. Implicit user behavior has also been used, but existing research and applications focus mostly on users' online activities such as clicking behavior, web search history, mouse movement, or the amount of time that a user is looking at a web page [2].

Social media websites provide a unique opportunity for personalized services to use other aspects of user behavior. Besides users' structured information contained in their profiles, e.g., demographics, users produce large amounts of data about themselves in variety of ways including textual (e.g., status updates, blog posts, comments) or audiovisual content (e.g., uploaded photos and videos). Many latent variables such as personalities, emotions and moods — which are typically not explicitly given by users — can be extracted from user generated content (see e.g. [6, 7]).

In this study, we examine the relationship between users' emotions and other characteristics on Facebook. Little work has been done that examines the relation between a user's emotions and other characteristics in social media. In [3] the authors extract emotions from Twitter posts and find correlations with major events in politics and popular culture during a specific time frame, but they focus on the public emotion as a whole and not on feelings or other characteristics of individual users. We detect emotions from users' status updates using the NRC word-emotion lexicon [13], and determine the relation between users' feelings and their demographics (age and gender) and personality. We also extract time features from the time stamp of the status updates to find the relation between users' emotions and time. To the best of our knowledge, no work has been done to find the relations between different emotions and personality with respect to time factors. In [14] the authors study the relation between emotions and time, however their work is based on a questionnaire and not based on social media content. In [12], the authors use SVM classifiers to predict personality using emotion expression in text. For their experiments, they use essays from psychology students, while in this work we focus on emotion expression in Facebook status updates and its relation with users' personality.

1.1 Personality and Emotions

Personality is a fundamental differentiating factor of human behavior. Research in the psychology literature has led to a well established model for personality recognition and description, called the Big Five Personality Model. Five traits can be summarized in the following way [5]:

– **Openness to experience** (Openness) is related to imagination, creativity, curiosity, tolerance, political liberalism, and appreciation for culture. People scoring high on Openness like change, appreciate new and unusual ideas, and have a good sense of aesthetics.
– **Conscientiousness** measures preference for an organized approach to life in contrast to a spontaneous one. Conscientious people are more likely to be well organized, reliable, and consistent. They enjoy planning, seek achievements, and pursue long-term goals. Non-conscientious individuals are generally more

easy-going, spontaneous, and creative. They tend to be more tolerant and less bound by rules and plans.

- **Extroversion** measures a tendency to seek stimulation in the external world, the company of others, and to express positive emotions. Extroverts tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative; they do not mind being at the center of attention, and make new friends more easily. Introverts are more likely to be solitary or reserved and seek environments characterized by lower levels of external stimulation.
- **Agreeableness** relates to a focus on maintaining positive social relations, being friendly, compassionate, and cooperative. Agreeable people tend to trust others and adapt to their needs. Disagreeable people are more focused on themselves, less likely to compromise, and may be less gullible. They also tend to be less bound by social expectations and conventions, and more assertive.
- **Emotional Stability** (reversely referred to as Neuroticism) measures the tendency to experience mood swings and emotions such as guilt, anger, anxiety, and depression. Emotionally unstable (neurotic) people are more likely to experience stress and nervousness, while emotionally stable people (low Neuroticism) tend to be calmer and self-confident.

Personality can affect the decision making process and has been shown to be relevant in the selection of music, movies, TV programs and books. It has been shown that personality affects preference for websites [10], language used in online social media [17], choice of Facebook Likes [11], music taste [16], and content such as movies, TV shows, and books [4].

In addition, it has been shown that users' *emotions* can also be used to detect users' taste at any moment, e.g., sad users are more likely to prefer action movies to watch [9]. Going yet one step further, personalized services can even have an impact on users' feelings. A nice example of this is that watching movies can change users' emotion, e.g., people feel joy when watching comedies or sadness when watching a late night romantic movie [9].

An interesting difference between personality and emotion is that personality is a stable characteristic and emotions are of short term duration. Emotion can be a momental feeling with respect to an object, person, event, or situation. As a consequence, people express a variety of different emotions over a period of time which is not the case for users' personality.

2 Material and Methods

2.1 Dataset

Our results are based on a data set from the myPersonality project [11]. MyPersonality was a popular Facebook application introduced in 2007 in which users took a standard Big Five Factor Model psychometric questionnaire [8] and gave consent to record their responses and Facebook profile. The data set consists of information about users' demographics, friendship links, Facebook activities (e.g., number of group affiliations, page likes, education and work history), status updates and Big Five Personality Scores. However, not all of this information is available for all users. From this data, we produce a data set of 5,865 users for which we have complete information about their age, gender, personality scores and status updates. Table 1 provides details about the data set characteristics.

Table 1: (Table on the left) Characteristics of female and male users in the data set. The entire data set contains 969,035 status updates written by 5,865 users. (Table on the right) Score threshold and number of users for each personality trait. Note that the same user can exhibit more than one personality trait at once.

	Female	Male
# users	3,446	2,419
Average age	26	25
# posts	625,921	343,114
Avg # posts/user	182	142
Min # posts/user	1	1
Max # posts/user	2,428	1,453

Personality	Threshold	# of users
Extroversion	3.60	2,971
Openness	3.80	3,284
Agreeableness	3.55	3,110
Conscientiousness	3.50	3,071
Neuroticism	2.80	2,631

The data set contains a personality score ranging from 1 to 5 for each user and each personality trait. To facilitate further analysis, for each personality trait we split the set of users into those that clearly exhibit the trait and those who do not. To this end we use the same thresholds that were used in the WCPR13 data set.[1] The score threshold and the number of users for each personality trait is presented in Table 1. For instance, in the remainder of this paper, we call a user an extrovert if his Extroversion score is at least 3.60; there are 2,971 such users in our data set. Note that such a binary split of users along the 5 personality dimensions is a fairly crude approach, and that a more fine grained study that considers the sliding scale from Introversion to Extroversion could provide further insights.

2.2 Emotion detection

To detect users' emotions from their status updates, we use the NRC hashtag emotion lexicon [13]. This lexicon contains a 10-dimensional binary emotion vector for 14,177 English words. The 10 dimensions or emotion categories are: *positive*, *negative*, *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, and *trust*. In the NRC lexicon, *positive* and *negative* are actually referred to as sentiments instead of emotions, but in our study we use the terms emotions and feelings loosely and interchangeably to refer to all 10 categories of the NRC lexicon.

A word can convey several emotions at the same time. For instance, according to the NRC lexicon, "happy" represents positive, anticipation, joy, and trust emotions, while "birthday" represents positive, anticipation, joy, and surprise

[1] http://mypersonality.org/wiki/doku.php?id=wcpr13

emotions. In the remainder of this paper, we say that a status update conveys an emotion if it contains at least one term from the lexicon that is associated with that emotion. For example, the status update "thanks to everyone who wished me a happy birthday today" conveys positive, anticipation, joy, trust, and surprise emotions because of the presence of the words "happy" and "birthday". The other words in this particular status update do not convey any emotion according to the NRC lexicon.

Figure 1 presents the frequency of emotions in the posts in our data set. Almost 60% of the status updates express at least one kind of emotion, and the positive emotion is clearly the most prominent one. For completeness, we

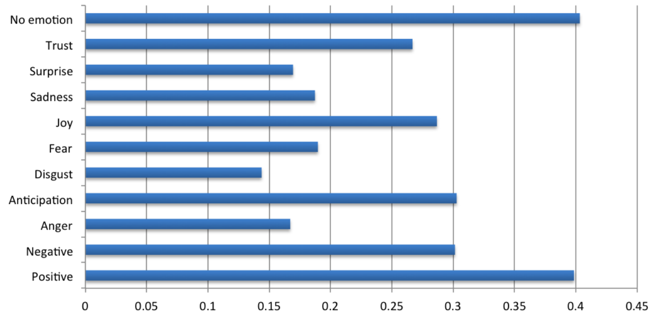

Fig. 1: Emotion frequency in Facebook status updates. Almost 60% of the status updates express at least one kind of emotion from the NRC lexicon, and many posts convey more than one emotion.

point out that to detect emotions we only scan the status updates for exact occurrences of words from the NRC lexicon. We use a bag of words approach and do not consider any misspellings (e.g., hapy or haaaappy), negation (e.g., not good), strength of the emotions using adjectives or adverbs (e.g., very happy vs. happy) or combined words (e.g., long-awaited vs. long awaited). Moreover, any emotions expressed with words that are not present in the NRC lexicon will remain undetected.

3 Results

In the remainder of this paper, let S denote the set of the 969,035 status updates in our study. Furthermore, for each of the 10 emotions 1:positive, 2:negative, 3:anger, 4:anticipation, 5:disgust, 6:fear, 7:joy, 8:sadness, 9:surprise, and 10:trust, let S_i, $i = 1, \ldots, 10$, be the set of status updates that contain at least one word associated with the respective emotion according to the NRC lexicon. As explained in Section 2.2, the sets S_1, S_2, ..., S_{10} are not necessarily disjoint. In addition, we also introduce S_0 as the set of status updates that do not contain a term from the NRC lexicon, i.e. S_0 is the set of status updates that do not convey any emotion. It holds that $S = S_0 \cup S_1 \cup S_2 \cup \ldots \cup S_{10}$.

3.1 Emotion and gender

Let S_f denote the set of status updates written by female authors and S_m the set of status updates by male authors. From Table 1 we know that women post more frequently than men. The probability that a status update is written by a woman is $P(S_f) \approx 0.65$ while the probability that it is written by a man is $P(S_m) \approx 0.35$. To determine the probability that a post conveys a particular emotion, given that it is written by a man or a woman, we calculate

$$P(S_i|S_m) = \frac{|S_i \cap S_m|}{|S_m|} \text{ and } P(S_i|S_f) = \frac{|S_i \cap S_f|}{|S_f|}$$

for $i = 0, 1, \ldots, 10$. The results are visualized in Figure 2.

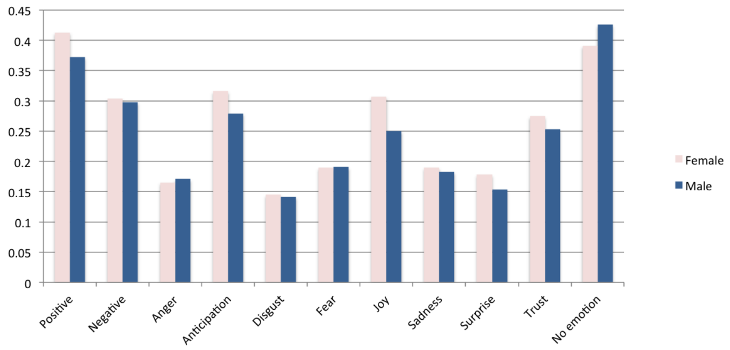

Fig. 2: Probability of occurrences of emotions in status updates from female and male users

Although the differences between both genders are small, we do observe that female users in general express more emotions in their posts. In particular, women are more likely than men to post about positive feelings, joy and anticipation, while men are more likely than women to post status updates that convey anger or no emotion at all.

3.2 Emotion and age

To assess the relation between different age groups and their emotion expression in Facebook, we use five age groups: users younger than 21, users between 21 and 30, users between 31 and 40, users between 41 and 50, and users older than 51. The average age of users in our data set is 26 years old with a standard deviation of 10, suggesting many young users in Facebook. For each age group a, let S_a be the set of status updates written by users from that age group. We calculate the probability of emotion expression for each age group a as $P(S_i|S_a) = \frac{|S_i \cap S_a|}{|S_a|}$ with S_i (for $i = 0, 1, \ldots, 10$) defined as in the beginning of Section 3.

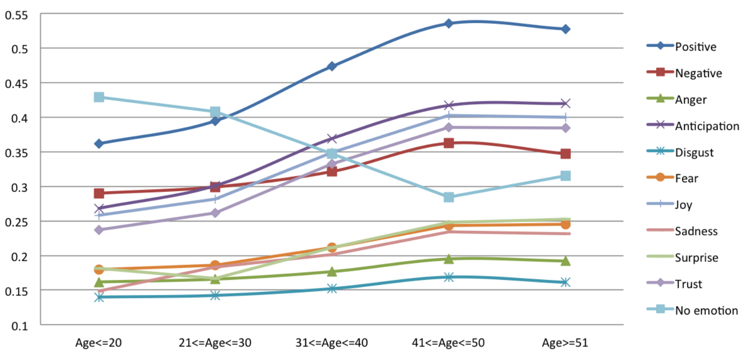

Fig. 3: Probability of occurrences of emotions in status updates from users of different age groups. Users are more likely to post emotions as they get older.

Based on Figure 3, the probability of expression of emotions increases with age. Users post more positive emotions as they get older. We find that older users are more emotional in their posts compared to younger users. Users between 40 to 50 years old have the smallest amount of status updates without emotion expression (less than 30%), which indicates their willingness to share their feelings. On the other hand, more than 40% of young users' posts (users less than 21 years old) are without emotions. This evidence could be caused by their language use and the fact that our dictionary does not contain all possible expressions.

3.3 Emotion and personality

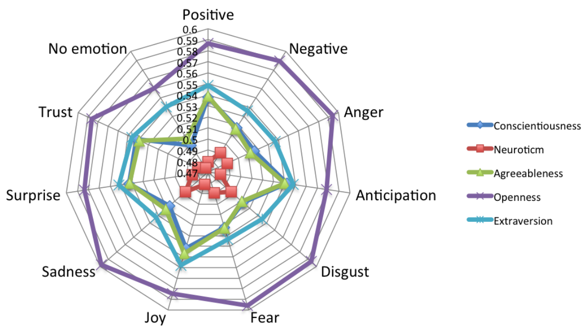

Fig. 4: Probability of occurrences of emotions in status updates from users with different positive personality traits.

Similarly as in the previous sections, for each of the personality traits, we consider the set of status updates written by users who meet the threshold for that personality trait according to Table 1. Using S_p to denote the set of status updates linked in this way to personality trait p, we compute $P(S_p|S_i) = \frac{|S_i \cap S_p|}{|S_i|}$ with S_i (for $i = 0, 1, \ldots, 10$) defined as in the beginning of Section 3. The results are visualized in Figure 4. Similarly, results of $P(\neg S_p|S_i) = \frac{|S_i \cap \neg S_p|}{|S_i|}$ are visulaized in Figure 5.

Neurotic users' posts are less likely to be emotional, while open users' posts convey emotions more frequently than other personalities. After open users, extrovert users express the most emotions in their posts. Interestingly, agreeable users express emotions very similar to conscientious users on Facebook.

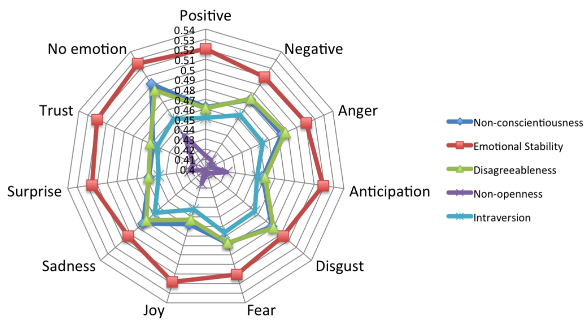

Fig. 5: Probability of occurrences of emotions in status updates from users with different negative personality traits.

Posts containing anticipation are mostly expressed by agreeable, conscientious and extrovert users. Neurotic users use less joy expressions than other personalities and their posts are most likely about disgust, sadness and negative feelings. Sadness appears more than other emotions for neurotic and open users, while joy emotions are expressed most by extrovert, conscientious and agreeable users. Open users also post frequently about their fear and anger.

3.4 Emotion and time

In this section, we investigate the relation between emotion expression and the time stamp of the posts. The graphs in this section depict the conditional probabilities of emotion expression w.r.t. time using $P(S_i|S_t) = \frac{|S_i \cap S_t|}{|S_t|}$, where S_t is the set of status updates posted in a specific time interval. In Figure 6, there are 7 such time intervals, each one corresponding to a day of the week. In Figure 7, the time intervals correspond to the months of the year.

Emotion and day of the week Figure 6 presents that people are more emotional during workdays than weekends. During the weekend (on Saturday and

Sunday), users are less emotional and their posts are more likely without emotion expression. Among other things, the number of posts about trust decreases during the weekend, and a similar observation holds for posts related to fear.

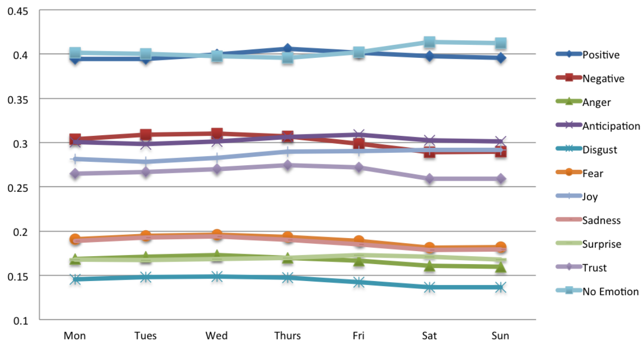

Fig. 6: Probability of occurrences of emotions in status updates depending on the day of the week. Status updates are more likely to contain emotions during workdays than during the weekend.

Facebook status updates are most likely to convey emotions on Thursday. From Friday onwards, the probability of emotion expression decreases. On Saturdays, users are least likely to express any emotions in their posts. Interestingly, the frequency of status updates conveying anger and surprise remains constant from Monday to Thursday. However, on Friday, users express more surprise and become less angry in their posts. In addition, users are more negative during the workdays and less likely express to joy. However, on Saturday and Sunday, users become less negative and more joyful.

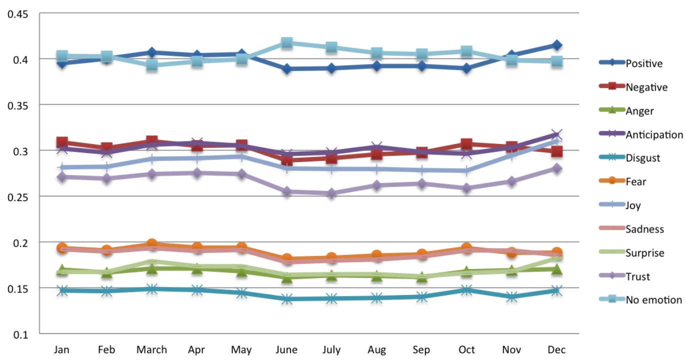

Fig. 7: Probability of occurrences of emotions in status updates depending on the month of the year. Status updates are most likely to contain emotions in December.

Emotion and month of the year Figure 7 presents the probability of emotion expression during different months of the year. Facebook users are more emotional in December; in particular, users are less negative, more joyful, surprised, anticipating and positive compared to other months of the year. This is reflected in posts such as *"Happy holiday", "Happy NYE", "Happy Christmas"* which are very prominent in December and which are tagged as emotion conveying posts by the emotion detection method described in Section 2.2. Although there are no significant changes in emotions during the rest of the year, during the summer months (June, July and August), the amount of positive, fear and trust expressions decreases, and users' posts are least likely to contain any emotion.

3.5 Correlations between features and emotions

Table 2: Pearson Chi Squared test results for on characteristics of users and posts, and emotion categories: *positive (Pos)*, *negative (Neg)*, *anger (Ang)*, *anticipation (Ant)*, *disgust (Dis)*, *fear (Fea)*, *joy (Joy)*, *sadness (Sad)*, *surprise (Sur)*, and *trust (Tru)*.

Features	Pos	Neg	Ang	Ant	Dis	Fea	Joy	Sad	Sur	Tru
Gender	**0**	**0**	**0**	**0**	**0**	0.205	**0**	**0**	**0**	**0**
Age	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Open	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Conscientious	**0**	0.014	0.793	**0**	**0**	**0**	**0**	0.496	**0**	**0**
Extrovert	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Agreeable	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.065	**0**	**0**
Neurotic	0.613	**0**	**0**	0.015	**0**	**0**	**0**	**0**	**0**	0.249
Monday	**0.001**	0.023	0.146	0.050	0.058	0.333	**0**	0.105	0.055	0.081
Tuesday	**0.001**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.029	0.879
Wednesday	0.213	**0**	**0**	0.137	**0**	**0**	**0.001**	**0**	0.220	**0.001**
Thursday	**0**	**0**	**0.008**	**0.002**	**0**	**0**	**0.002**	**0.001**	0.482	**0**
Friday	0.019	0.029	0.517	**0**	0.139	0.566	**0.001**	0.047	**0**	**0**
Saturday	0.437	**0**	**0**	0.891	**0**	**0**	**0**	**0**	0.104	**0**
Sunday	0.029	**0**	**0**	0.170	**0**	**0**	**0**	**0**	0.200	**0**
January	0.039	**0**	0.016	0.704	**0.007**	**0.003**	**0**	**0**	0.066	**0.004**
February	0.432	0.377	0.740	**0.001**	0.035	0.322	**0.006**	0.082	0.094	0.139
March	**0**	**0**	**0.001**	0.019	**0**	**0**	**0.004**	**0**	**0**	**0**
April	**0.003**	0.027	**0.002**	**0.002**	**0.001**	**0.002**	**0.004**	0.025	**0.002**	**0**
May	**0.001**	0.016	0.465	0.269	0.623	**0.004**	**0**	**0.003**	**0.003**	**0**
June	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
July	**0**	**0**	0.011	**0.001**	**0**	**0**	**0**	**0**	**0**	**0**
August	**0**	**0**	**0.001**	0.424	**0**	**0**	**0**	**0**	**0**	**0.001**
September	**0**	0.014	**0**	**0.004**	**0.003**	0.016	**0**	**0.008**	**0**	0.027
October	**0**	**0**	0.614	**0**	**0.001**	**0.003**	**0**	**0.009**	**0.004**	**0**
November	**0.001**	0.130	0.126	0.760	**0.002**	0.410	**0**	**0.003**	0.250	0.513
December	**0**	0.113	**0.004**	**0.004**	0.004	0.211	**0**	0.195	**0**	**0**

To assess the relation between the different features and emotions, we apply the Pearson chi-squared dependence test [15]. Table 2 presents the p-values. The null hypothesis is that features and emotions are independent. The p-values that are lower than the significance level ($p < .01$) denote significant correlations of features with emotions. They are indicated in bold in the table.

Gender is related with all emotion categories except fear. Age is shown to be related to all emotion types. Similarly, Openness and Extroversion are related with all emotion types. Conscientiousness is related to anger, negative and sadness emotions. Agreeableness is not related to sadness. And finally, Neuroticism shows no relation with positive, anticipation and trust emotions.

4 Conclusion and Future Work

In this study, we explored the relation between the emotions of 5,865 Facebook users with their age, gender and personality by using their status updates (almost 1 million posts). We used the NRC hash-tag emotion lexicon to detect emotions from the posts. We also extracted temporal features from the posts' time stamps. Almost 60% of status updates contain at least one type of emotion expression. Positive emotion is expressed with the highest frequency in status updates and disgust is least likely to appear in the status updates of users.

The results confirm a relation between users' characteristics and their emotions. Similar to offline expression, female Facebook users express more emotions in their status updates than male users. Similarly, older users express more emotions in their status updates than younger users. Neurotic users are not very emotional in their status updates, while open users are mostly likely to express their feelings about different subjects. By analyzing the time stamp of the status updates, we examined relations between Facebook posts' time and users' feelings. Interestingly, emotions are more likely to be expressed during the workdays compared to the weekend. The frequency of emotional status updates is lowest during the summer and highest in December.

We found significant correlations between our selected features and users' emotions. In future research, we will develop a model that will predict the most probable upcoming emotion for each user, among other things based on time, demographics and personality. We believe that being able to predict users' emotions and target the end users accordingly would be useful for personalized services.

Aside from the work we have presented in this paper, there is clear potential for more fine grained emotion detectors. Emotion detection in this study has been performed using a lexicon based approach. However, due to the complexity of the status updates, the limited size of the lexicon, and a huge amount of noise in the unnormalized status updates, it is very likely that we have missed many emotion expressions in the status messages. Exploring better techniques to extract emotions not only based on the words, but also based on other features is potentially an open path to explore.

Acknowledgement: This project is part of the SBO Program of the IWT (IWT-SBO-Nr. 110067).

References

1. Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

2. Gediminas Adomavicius and Alexander Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.
3. Johan Bollen, Huina Mao, and Alberto Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *ICWSM*, 2011.
4. Iván Cantador, Ignacio Fernández-Tobías, Alejandro Bellogín, Michal Kosinski, and David Stillwell. 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), at the 21st Conference on User Modeling, Adaptation and Personalization (UMAP 2013)*, 2013.
5. Paul T Costa and Robert R McCrae. The revised NEO personality inventory (NEO-PI-R). *The SAGE Handbook Of Personality Theory And Assessment*, 2:179–198, 2008.
6. Golnoosh Farnadi, Susana Zoghbi, Marie-Francine Moens, and Martine De Cock. Recognising personality traits using facebook status updates. In *Proceedings of the workshop on computational personality recognition (WCPR13) at the 7th international AAAI conference on weblogs and social media (ICWSM13)*, 2013.
7. Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 253–262. ACM, 2011.
8. Lewis R Goldberg, John A Johnson, Herbert W Eber, Robert Hogan, Michael C Ashton, C Robert Cloninger, and Harrison G Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84–96, 2006.
9. Ai Thanh Ho, Ilusca LL Menezes, and Yousra Tagmouti. E-mrs: Emotion-based movie recommender system. In *Proceedings of IADIS e-Commerce Conference. USA: University of Washington Both-ell*, pages 1–8, 2006.
10. Michal Kosinski, Yoram Bachrach, Pushmeet Kohli, David Stillwell, and Thore Graepel. Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, pages 1–24, 2013.
11. Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.
12. Saif M Mohammad and Svetlana Kiritchenko. Using nuances of emotion to identify personality. *arXiv preprint arXiv:1309.6352*, 2013.
13. Saif M Mohammad and Peter D Turney. Nrc emotion lexicon. 2013.
14. Ante Odic, Marko Tkalcic, Jurij F Tasic, and Andrej Košir. Relevant context in a movie recommender system: Users opinion vs. statistical detection. *ACM RecSys*, 12, 2012.
15. Robin L Plackett. Karl pearson and the chi-squared test. *International statistical review*, 51(1):59–72, 1983.
16. Peter J Rentfrow and Samuel D Gosling. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6):1236, 2003.
17. H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791, 2013.

Using Social Media Mining for Estimating Theory of Planned Behaviour Parameters

Marko Tkalčič[1,2], Bruce Ferwerda[1], Markus Schedl[1], Cynthia Liem[3], Mark Melenhorst[3], Ante Odić[4], and Andrej Košir[2]

[1] Johannes Kepler University Department of Computational Perception, Linz, Austria
[2] University of Ljubljana Faculty of electrical engineering, Ljubljana, Slovenia
[3] Delft University of Technology, Multimedia Computing Group, Delft, The Netherlands
[4] Outfit7 (Slovenian subsidiary Ekipa2 d.o.o.), Ljubljana, Slovenia

Abstract. In this position paper we present the scenario of making interventions for increasing the classical music concert-going behaviour of end users. Within the FP7 Phenicx project we are developing a personalized persuasive system that attempts at changing the concert-going behaviour of users. The system is based on the theory of planned behaviour user model for predicting whether a user will attend a concert or not. Our goal is to develop a machine learning algorithm that will extract the user model parameters unobtrusively from the micro-blogs of the users. We plan to perform a user study to build the training dataset and to test the system on real users within the project.

Keywords: Phenicx, theory of planned behaviour, user modeling, classical music

1 Introduction

Classical music is a domain of music that has an image of inaccessibility to non-experts. Hence, lots of potential classical-music-concert goers do not attend concerts. To alleviate this problem we plan to develop a personalized intervention system that will persuade users to attend classical music concerts.

In this position paper we present the work done so far and the plan of the work to be carried out for addressing the issue of limited classical music concert going. The work presented here was carried out within the European FP7 project Phenicx[5] [4].

In order to design a personalized intervention system the users need to be modeled. We selected the Theory of Planned Behaviour (TPB) model [1], since it was designed especially for intervention scenarios. The TPB model parameters will be predicted using social media mining. In the training phase, questionnaires will be submitted to users to train the prediction model. The intervention will be carried out when the attitude and/or the social norm (two of the TPB prameters) will have low values.

[5] `http://phenicx.upf.edu/`

1.1 Related work

To the best of the authors' knowledge there have been no attempts to design persuasive systems for classical music. However, there are two groups of related work for the task at hand: (i) models of intervention (related to changing the concert-going behaviour) and (ii) social-media mining (related to unobtrusively extracting the parameters of the user model).

Models for intervention We performed a user study on barriers and motivators for attending classical music concerts. Through focus group with subjects we collected data about the reasons that prevents them going to concerts and reasons that drive them to classical concerts. The study suggested that some of the barriers that users provided (e.g. social aspects, the user's background knowledge) could be diminished by interventions. A natural choice for the aforementioned barriers appear to be the TPB model [1] as it takes into account personal attitudes towards going to the concert (related to background knowledge) and social norms (related to the social aspects). The TPB model was designed specifically for intervention scenarios where we wish to predict a behaviour and make interventions if it does not fit the desired behaviour [1]. The parameter values for the TPB model are usually acquired through carefully designed questionnaires. However, these questionnaires are intrusive and time consuming, both for the end user as for the designer of the system, as they need to be done on a behaviour-level (i.e. the concert level, in our case).

The Theory of Planned Behaviour Model The TPB was proposed by Ajzen [1]. The model links an observed behaviour with the subject's beliefs (see Fig. 1. The beliefs taken into account by the TPB are (i) the behavioural beliefs, (ii) the normative beliefs and (iii) the control beliefs. The behavioural beliefs influence the attitude of the subject towards the observed behaviour. The normative belief (i.e. the perception of the social norms' pressure) influences the subjective norm. The control beliefs influence the perceived behavioural control. These three norms have an influence on the behavioural intention, which is an indicator of the actual behaviour.

Social-media mining So far there have been no attempts to extract TPB parameters from social media. Hence, we will rely on related work that uses social media to extract similar user parameters. the underlying assumption is that social media, concretely, microblogging (i.e. twitter) in our case, contain a lot of personal parameters that can be used for user modeling. These information are encoded in two aspects of the microblogs: in the content (the microblogs themselves) and in the social relations between users. Some closely related work has successfully managed to extract personality parameters from tweets [5,2]. Twitter has also been a rich data source for predicting trend-setters [6]. A study of Facebook logs showed that it is possible to predict further personal traits like sexual orientation, ethnicity, religious and political views, personality

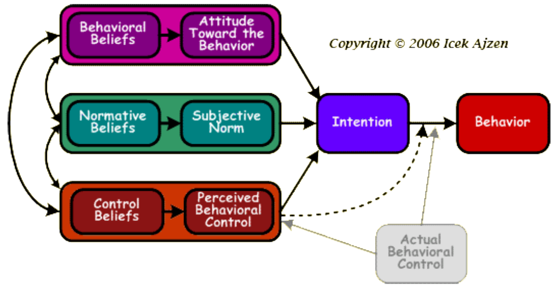

Fig. 1. The Theory of Planned Behaviour model (source: [1])

traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender [3]. This related work (especially the features used and the regression/classification approaches) will serve as the basis for developing our algorithms for predicting TPB parameters.

2 The Scenario

For most of users in urban environments there are classical concerts going on constantly in their vicinity. Most of the medium-to-big cities have concert venues where classical concerts take place. Assuming that these concert are advertised, potential concert goers are aware of them. However, there are issues that prevent these users to attend concerts. By addressing these issues with a personalized intervention we can increase the chance of users to attend classical concerts. The envisioned scenario is that of a personalized application on a mobile device (e.g. a tablet or a mobile phone) that is able to predict whether the observed user has a high or low chance of attending the concert. In case of the low chance, the application should predict the reason for it and perform an intervention.

Given the choice of the TPB model we can focus on three aspects that influence the observed behaviour, i.e. *going to the specific classical music concert*: (i) the user's attitude towards the concert, (ii) the social norm and (iii) the perceived control. Each of these can be a reason for the low chance of attending the concert. More specifically we plan to address interventions in the cases when the first two parameters (i.e. the attitude and the social norm) are low.

Once we determine the numerical values of the two parameters from social media mining, we are able to make an appropriate intervention. In case that the attitude is low, the application should provide more information to make the classical concert more appealing. In case of low social norm, the application should provide messages that persuade the user that her/his peers approve of the classical concert.

3 Experimental design

In this section we present the steps that we plan to undertake in order to build the personalized intervention system. First we present the technical flow of the scenario and then we address each step that needs to be carried out in order to have a functional system.

3.1 TPB for designing pre-concert interventions

The technical scenario is depicted in Fig. 2 and has the following flow:

1. the application takes as input an item (i.e. the observed concert, described with some descriptors/metadata) and a user (described with features extracted from the microblog stream-tweets)
2. for the observed concert and user it predicts the TPB parameters
3. if the predicted intended behaviour is *no* (i.e. the user does not intend to go to a concert) the application generates an intervention (e.g. provides additional info to change the attitude, if low attitude is the problem)
4. observe the actual behaviour
5. compare it to the predicted behaviour

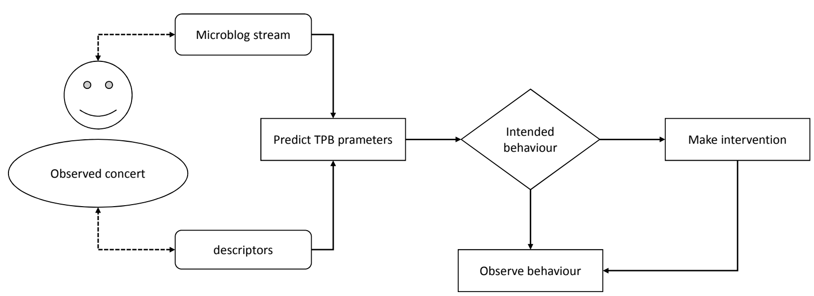

Fig. 2. The experiment flow

Step 1: find relevant concert parameters (descriptors) The first step is to identify the most relevant parameters to describe the concert with. Our wish is to find some simple, one-word, descriptors like composer, performer, type (e.g. symphony, opera . . .). We plan to make a web-questionnaire to identify the parameters that the users think will be important for going to a concert. Out of the top rated parameters by the users we will take the top three and annotate the upcoming concerts with these metadata. The upcoming concerts will be selected from the concerts provided by the Phenicx partners, the Royal Concertgebouw Orchestra (Amsterdam, The Nederlands) and the ESMUC (Barcelona, Spain).

Step 2: Gather TPB data For each of the concerts annotated in the previous step we will design a TPB questionnaire and ask users to fill it in. By doing this, we will be able to get a table with the following columns-variables (features):

- concert related variables
  - performer
  - composer
  - type
- TPB related variables
  - attitude towards behaviour
  - social norm
  - perceived control
  - intended behaviour
- twitter related variables
  - num of tweets
  - num of retweets
  - number of followers
  - number of friends
  - content related features

We need to gather enough data (the number of users and concert is to be defined) to be able to build a model for predicting TPB parameters from twitter data.

Step 3: training the TPB model The underlying assumption for using microblogs for predicting TPB parameters are the following: (i) the twitter social network and blogging behaviour (i.e. original blog postings, retweets) reflect the general social beliefs of the user, (ii) the social network and the content of the microblogs reflect the attitude towards classical music aspects. Using the above mentioned features we plan to train the TPB parameters with regression models for the observed concert/user pairs.

The model will predict the intended behaviour IB with a weighted sum (weights w_1,w_2 and w_3) of the attitude AB, social norm SN and the perceived control PC

$$IB = w_0 + w_1 AB + w_2 SN + w_3 PC \tag{1}$$

3.2 Issues in the experiment

Having nominal attributes for the concert descriptors (i.e. composers, performers, genre) would require big amounts of training data for each of the value of these attributes. Hence, we will have to limit the values of these attributes to just few of them. In the selection of these values we will take into account the number of subjects that we will be able to include in the experiment and the relevant statistical aspects (i.e. effect size, power analysis).

Furthermore, we will not model the perceived control PC and the related weight w_3 from Eq. 1, hence reducing the parameters estimation problem in the training phase to a more viable one.

4 Current data

We have carried out the acquisition of TPB parameters on a small scale for one concert and five subjects. We have crawled the twitter feeds from these users and extracted the following features: *number of followers*, *number of friends*, *number of listed*, *number of statuses*, *number of tweets* and *number of retweets*. The amount of data is too low to draw any conclusions, however, we report it in Fig. 3.

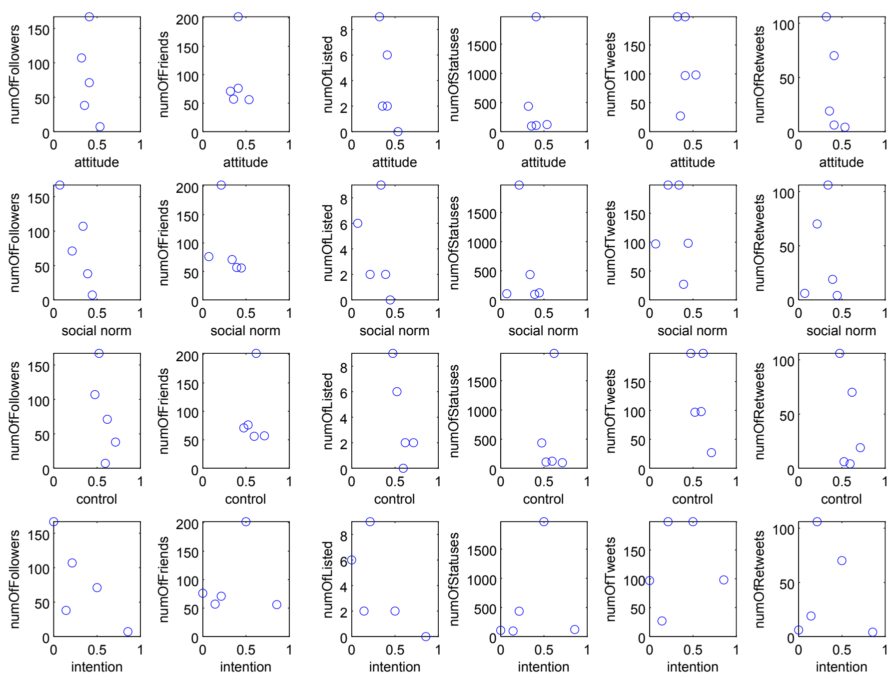

Fig. 3. Distributions of the TPB parameters in the twitter features spaces.

5 Discussion and Conclusion

In this position paper we presented the scenario of making interventions for increasing the classical music concert-going behaviour of end users. The work is being carried out within the FP7 Phenicx project. We showed that the barriers for concert going indicate the choice of the TPB model as a suitable one. We also reviewed related work that uses social media for the extraction of user model parameters. We plan to base our algorithms for the unobtrusive extraction of TPB parameters on these related work. We propose the model for predicting the user behaviour (i.e. going to a concrete concert) and the procedure for estimating the parameters of the model using social media mining and machine learning.

Acknowledgement

This research is supported by the European FP7/2007-2013 programme through the PHENICX project under grant agreement no. 601166 and by the Austrian Science Fund (FWF): P25655.

References

1. Ajzen, I.: The theory of planned behavior. Organizational Behavior and Human Decision Processes 50(2), 179–211 (Dec 1991), http://linkinghub.elsevier.com/retrieve/pii/074959789190020T
2. Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11 p. 253 (2011), http://portal.acm.org/citation.cfm?doid=1979742.1979614
3. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences pp. 2–5 (Mar 2013), http://www.pnas.org/cgi/doi/10.1073/pnas.1218772110
4. Liem, C.C.S., Sterren, R.V.D.: Innovating the Classical Music Experience in the PHENICX Project : Use Cases and Initial User Feedback. 1st International Workshop on Interactive Content Consumption (WSICC) at EuroITV 2013 (2013)
5. Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In: 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing. pp. 180–185. IEEE (Oct 2011), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6113111http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6113111
6. Sha, X., Quercia, D., Michiardi, P., Dell'Amico, M.: Spotting trends. In: Proceedings of the sixth ACM conference on Recommender systems - RecSys '12. p. 51. ACM Press, New York, New York, USA (2012), http://dl.acm.org/citation.cfm?doid=2365952.2365967

Towards Learning Relations between User Daily Routines and Mood

Berardina De Carolis, Stefano Ferilli

Dipartimento di Informatica, Università di Bari
70126 Bari, Italy
berardina.decarolis@uniba.it,stefano.ferilli@uniba.it

Abstract. The main aim of our research is to investigate whether it is possible to find a relation between the user's mood and daily routines. To this aim we developed an application, Smart Calendar, that, thanks to its nature, allows collecting annotated datasets of mobile data, learning user daily routines and relating activities and pattern of sensor data to the user mood. Results of this preliminary phase of the research have shown that it is possible to learn routines and, if present, to find relations with the user mood.

1 Introduction

Pervasive computing applications to be successful should proactively support users in their tasks according to their situation [1]. To this aim, the application should reason on the user and on the context for making prevision about the user's needs and for personalizing the task accordingly. In particular, given data that can be captured by mobile phones over long durations of time, it is possible to discover the emerging behavior of people (including habits and routines) over a long period [2,3]. Examples of applications that could exploit models of the user's daily routines are context-aware reminders, personal assistants, contexts-aware recommender systems, life diaries and lifelogging systems. In a previous stage of our research we focused our effort on learning the daily routines of the user both from sensors present in smart environments and from mobile data [4]. In particular, our approach exploits the WoMan (acronym for 'Workflow Management') system to incrementally learn and refine users routines represented in First-Order Logic (FOL) [5]. The WoMan system is able to deal with non-sequential activities and repeated-tasks, and therefore the approach is suited to represent models of user's routines for predicting the user needs. At the current phase of research we investigate whether it is possible to use our approach to relate patterns of daily routines to affective factors, mood in particular. Mood is an affective state that plays a significant role in our lives. It is more general and less intense than emotion, and depends on broader factors and not necessarily on stimuli [6]. Mood influences our behavior, decisions, communication and preferences. Therefore, we consider the possibility to learn possible relations between mood, activities and routines both in the physical world (activities located in a place such as shopping, driving, and so on) and in the digital world (activities like email, listening to music, posting on facebook, and so on). Starting from the consideration that smartphones have rich information about their owner and that they can help us in collecting a lot of traces (sensor data), many applications have been developed with the purpose of mapping mood to recurrent patterns of sensors data [7,8,9,10]. In our work we aim at discovering the relation between the user's daily location-driven routines and mood states and changes starting from an annotated real life human dataset collected by mobile phones. To this aim we developed an application (SmartCalendar) that allows to incrementally learn contextual models of the user routines and associate them with certain moods. The purpose of SmartCalendar is twofold: on one side it provides to the user a calendar and a ToDoList manager and a context-aware reminder, on the other side it allows gathering annotated dataset of mobile usage data, to learn the daily routines and relate activities to mood that is supplied by the user on request. This phase is necessary in order to collect a dataset that relates variations in the user mood to activities, places, changes in the routine. Results of a first analysis showed that a relation exists and it is possible with our approach to learn such a relation and then use the learned model to predict variations.

2 Incremental Learning of Daily Routines

As far as learning the lifestyle of the user by building models of his daily routines is concerned, it can be seen as a set of processes. Therefore, modeling such routines can then be cast as a process-mining task. A *workflow* model is a formal specification of how a set of tasks can be composed to result in valid processes, allowing compositional schemes such as sequence, parallel, conditional, or iteration. So, in WoMan we decided to learn models that are represented as workflows. In order to understand the example in this paper, we provide here a short description of the employed formalism. In WoMan, a trace element is represented as a 6-tuple **(T,E,W,P,A,O)** where **T** is the time/date the event occurred, **E** is the type of the event (begin of process, end of process, begin of activity, end of activity), **W** is the name of the workflow the process refers to, **P** is a unique identifier for each process execution, **A** is the name of the activity, and **O** is the progressive number of occurrence of *A* in *P*. This is a standard formalism that allows describing explicitly the flow of activities (both sequential and parallel). An example is provided in the following:

```
entry(20121001073047,begin_of_process, monday,rl1,none,none).
entry(20121001073048,context_description, monday,rl1,[mood_valence_neg,
mood_arousal_low, meteo_rain, temp(19.5)],none).
entry(20121001073049,begin_of_activity,monday,rl1,act_Wakeup,1).
entry(20121001073523,end_of_activity,monday,rl1,act_Wakeup,1).
entry(20121001080940,begin_of_activity,monday,rl1,act_KidsToSchool,1).
entry(20121001083002,end_of_activity,monday,rl1,act_KidsToSchool,1).
entry(20121001083010,context_description, monday,rl1,[mood_valence_neutral,
mood_arousal_low, meteo_cloud, temp(20.5)],none).
entry(20121001093028,begin_of_activity,monday,rl1,act_GoToWork,1).
entry(20121001093028,begin_of_activity,monday,rl1,act_Eat,1).
entry(20121001150033,end_of_activity,monday,rl1,act_Eat,1).
```

```
entry(20121001150033,end_of_activity,monday,rl1,act_GoToWork,1).
entry(20121001150036,begin_of_activity,monday,rl1,act_KidsFromSchool,1).
entry(20121001153007,end_of_activity,monday,rl1,act_KidsFromSchool,1).
…
entry(20121002073524,end_of_process,monday,rl1,none,none).
```

As context and activities are detected or entered by the user, the corresponding entries are provided to the WoMan system that, applying the algorithm described in [5], learns activities and relations among them. The task flow of a case is internally expressed in WoMan as a conjunction of ground atoms built on the following predicates:

- **activity(S,T)** : at step *S* task *T* is executed
- **next(S',S")** : step *S"* follows step *S'*

Argument *T* of the *activity/2* predicate is taken from a (fixed and context-dependent) set of constants representing the allowed tasks. Steps are denoted by unique identifiers. Steps are associated to events, and can be implemented as timestamps denoting the associated events. The *next/2* predicate allows to explicit represent parallel executions in the task flow. This avoids the need to infer/guess the parallelism by means of statistical considerations, which may of course be wrong and thus mislead the workflow learning process. Any trace represented in the 6-tuple format previously introduced can be automatically translated into this internal format. For instance, the previous sample trace for the 'monday' would be expressed as:

activity(s0,act_Wakeup), mood_valence_neg(s0), mood_arousal_low(s0), meteo_rain(s0), temp(s0,19.5), next(s0,s1), activity(s1,act_KidsToSchool), mood_valence_neg(s1), mood_arousal_low(s1), meteo_rain(s1), temp(s1,19.5), next(s1,s2), activity(s2,act_GoToWork), mood_valence_neutral(s2), mood_arousal_low(s2), meteo_cloud(s2), temp(s2,20.5), next(s2,s3), activity(s3,act_Eat), mood_valence_neutral(s3), mood_arousal_low(s3), meteo_cloud(s3), temp(s3,20.5), …

The first activity ('Wakeup') is associated to step s0. At that time, the last detected context says that the actor has a negative mood valence and low arousal, it is raining and the temperature is 19.5°C. Activity 'KidsToSchool' is associated to step s1, and has a 'next' relationship to 'Wakeup' as the (only) most recently closed activity. At this time, no changes in the context have been notified to the system, and hence it assumes that at step s1 the context is the same as for the previous step, and so on.

In the WoMan system a workflow structure is described as a conjunction of atoms built on the following predicates:

- **task(t,C)** : task *t* occurs in cases *C*;
- **transition(I,O,p,C)** : transition *p*, that occurs in cases *C*, consists in ending all tasks in *I* (that must be running), and starting the execution of new instances of all tasks in *O*.

Argument *C* represents a history of those tasks/transitions, and thus can be exploited for computing statistics on their use.

WoMan may run in 3 modes. The *learning* mode allows to learn a process model from logs of activities. The *supervision* mode allows to apply a learned model to new cases of the process in order to check that they are compliant with the model. The *prediction* mode allows to apply a learned model to new cases of the process in order to foresee the most likely subsequent activities that the user will perform at a given moment of the execution.

Models are built by WoMan according to the procedure reported in Algorithm 1.

Algorithm 1 Refinement of a workflow model according to a new case

Require: $\mathcal{W}$: workflow model
Require: c: case having FOL description D
  for all $\texttt{activity}(s,t) \in c$ **do**
    if $\exists\ \texttt{task}(t,C) \in \mathcal{W}$ **then**
      $\mathcal{W} \leftarrow (\mathcal{W} \setminus \texttt{task}(t,C)) \cup \{\ \texttt{task}(t,C \cup \{c\})\ \}$ /* update statistics on task t */
    else
      $\mathcal{W} \leftarrow \mathcal{W} \cup \{\ \texttt{task}(t,\{c\})\ \}$ /* insert new task and initialize statistics */
    end if
    refine_precondition($\mathcal{W}, t(s)$:- $D|_s$)
    refine_postcondition($\mathcal{W}, t(s)$:- D)
  end for
  for all $\texttt{next}(s',s'') \in c$ **do**
    $I \leftarrow \{t' \mid \texttt{activity}(s',t') \in c\}$
    $O \leftarrow \{t'' \mid \texttt{activity}(s',t'') \in c\}$
    if $\exists\ \texttt{transition}(I,O,p,C) \in \mathcal{W}$ **then**
      $\mathcal{W} \leftarrow (\mathcal{W} \setminus \texttt{transition}(I,O,t,C)) \cup \{\ \texttt{transition}(I,O,t,C \cup \{c\})\ \}$
      /* update statistics on transition p */
    else
      $p \leftarrow$ generate_fresh_transition_identifier()
      $\mathcal{W} \leftarrow \mathcal{W} \cup \{\ \texttt{transition}(I,O,p,\{c\})\ \}$
      /* insert new transition and initialize statistics */
    end if
  end for

The described approach is fully incremental: it can start with an empty model and learn from one case (while others need a large set of cases to draw significant statistics), and can refine an existing model according to new cases whenever they become available (introducing alternative routes, even adding new tasks if they were never seen in previous cases, and updating the statistics). This peculiarity is an advance to the state-of-the-art, because continuous adaptation of the learned model to the actual practice can be carried out efficiently, effectively and transparently to the users. A graphical representation of a portion of the learned model for the Monday workflow is shown in Figure 1.

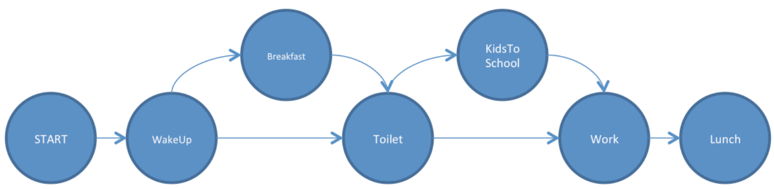

Fig. 1. A graphical representation of a portion of learned workflow for the "Monday" routine.

3 An overview on the SmartCalendar Application

In order to collect a dataset of entries containing annotated data about activities and mood, we implemented the Android application *Smart Calendar*. Besides providing the user with the functionalities of a Calendar, it is a ToDoList and a context-aware Reminder. It allows to learn the activities performed by the user according to two modalities: user-supplied information about performed activities and geo-localization. Each activity in the calendar is related to a place (voluntarily supplied by the user when inserting the activity or using a service that is activated on the bases of the GPS position and time of stay). For instance, if the user is in a supermarket at a certain time and this activity was not entered in the calendar, then the application will ask to the user what he is doing there and will insert the activity in the calendar. Indeed, if the user is in a place at the time expected for doing an activity present in the calendar and stays there for the expected amount of time, the application will consider that activity as done by the user. This is a strong assumption, but asking many information may annoy the user that could abandon using the calendar. However the user may always check the calendar and delete or insert activities. In both cases, SmartCalendar will send the entries about activities and the context to the WoMan system according to the formalism described previously. Figure 2 shows the interfaces for the main tasks of the application.

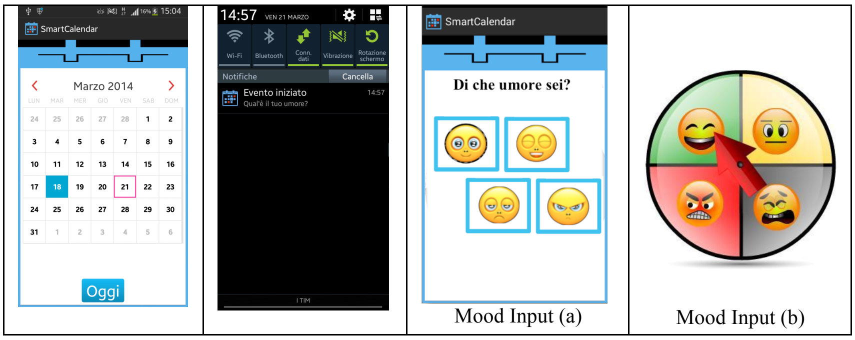

Fig. 2 Interfaces for the main tasks of the Smart Calendar Application

As in many similar applications [9,10], users are asked but not forced to input information about their mood at least three times a day: when the user starts using the telephone, when a not expected activity is performed by the user and after four hours after the last input. This is done using push notifications that are typically not very intrusive. In this way we can capture a user's self-reported mood in relation to activities and places. In deciding how to represent mood information we decided to adopt the two-dimensional approach (valence and arousal) described in the Circumplex model [11]. We developed two prototypes for testing which was a more intuitive and usable interface. In the former (Figure 2a) we used a gallery of faces expressing the set of moods that we considered relevant for our study. In the latter (Figure 2b) we used a circle divided in four quadrants representing the main moods associated to the combinations of the two dimensions of the model and an arrow that allows indicating also intermediate states. In order to avoid confusion and misunderstandings, we put representative faces in the quadrants. In both cases each selection corresponded to a combination of the valence and arousal dimensions. After a formative usability test we decided to use the second interface since it was preferred by 76% of the users. Besides collecting activities of the user related to places, we created a module for logging relevant information regarding application usage, phone calls, email, messages, compass data, and so on.

4 Analysis of the Collected Data

In order to explore the relations between changes in daily routines and the variation in people's mood, we used WoMan on the collected dataset. In total we collected data from 10 users (160 annotated days on average for each user) aged between 28 and 48. Then, first we learned the daily routine models for each working day (Monday-Friday) using WoMan in 'learning' mode (as described in Section 2). In order to have an insight of the learning system's performance, we tested the accuracy of the learned models using a 10-fold cross-validation procedure. For the considered dataset the system reached 85,63% average accuracy with a standard deviation of 10,84. After the learning step, we removed noisy (i.e., infrequent) pieces of the model, and specifically transitions and activities that were encountered in less than 4% of the training cases. Then, we ran WoMan in 'supervision' mode (see again Section 2) using the denoised models and the training cases, in order to collect for each day the warnings returned by WoMan, denoting deviations from the routines. After this, the cases were sorted in chronological order, and a histogram of the warnings in each day was drawn. The curves of mood variations, calculated as a function of valence and arousal, were finally superimposed to this histogram, normalizing the warning bars to their range [-1,1]. Specifically, for each day the average value of each parameter was plotted in these curves. This allowed to visually detect correlations between high or low valence and/or arousal and days with many warnings, which would confirm the influence of mood over routinary behavior (Figure 3).

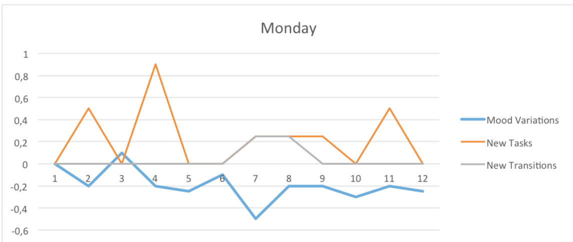

Fig. 3 Routine and mood variations

For sample days in which this correlation was noteworthy, it is possible to draw the plot of the mood parameters during the day, and to superimpose it to the curve of the warning occurrences during the same day, that would further confirm the relationship.

Moreover, we calculated on the dataset the percentage of mood variations in correlation with changes in the routine. In particular, a mood variation occurred in correspondence to a routine variation in 79% of the cases. At present we have not evaluated which factors affect the mood variation in terms of positivity or negativity (i.e. personality traits, type of events or activities, etc.) but this will be the goal of our future work.

5 Conclusions and Future Work

In this paper we reported the first results of the application of the WoMan system to the task of learning daily routines and their relation to mood variation. In order to do so, we developed an application, Smart Calendar, that, besides providing the typical functions of a smart calendar, allows to collect an annotated dataset of activities related to places and context features. In order to investigate on this, we used WoMan in supervision mode, checking the co-occurrence of significant changes in mood and deviations from the routine. Moreover we are combining data collected from mobile data with those that can be detected in indoor situations from indoor sensors of a smart environment [12]. We are aware that collecting personal sensors data may have a high impact on privacy issues. Most of the approaches present in the literature use cryptography, privacy-preserving data mining or store only inferred data from low-level sensors, that are discarded after this inference step. MoodMiner, for instance, hashes all the private user data [13]. At present our application informs the user that his data will be used only for research purposes and will be accessible only by the researchers working on the project. In the near future we plan to implement a stronger privacy policy and to develop a mood inference model that relates events, activities and other contextual factors that can be inferred by mobile data.

References

1. A. Schmidt. Context-Aware Computing: Context-Awareness, Context-Aware User Interfaces, and Implicit Interaction. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed.". Aarhus, Denmark: The Interaction Design Foundation, 2013.
2. K. Farrahi and D. Gatica-Perez, What Did You Do Today? Discovering Daily Routines from Large-Scale Mobile Data, ACM International Conference on Multimedia (ACM MM), Vancouver, Canada, 2008.
3. F. Pianese, X. An, F. Kawsar, and H. Ishizuka. Discovering and predicting user routines by differential analysis of social network traces. In WoWMoM, 2013.
4. H. Jongyi, S. Eui-Ho, J Kim, and J Kim. 2009. Context-aware system for proactive personalized service based on context history. Expert Syst. Appl. 36, 4 (May 2009), 7448-7457.
5. S. Ferilli. WoMan: Logic-based Workflow Learning and Management. IEEE Transaction on Systems, Man and Cybernetics: Systems, ISSN 2168-2216, IEEE Press, 2013.
6. C. J. Beedie, P. C. Terry and A. M. Lane. Distinctions between emotion and mood. Cognition and Emotion, 19(6), 2005.
7. A. Matic, A. Papliatseyeu, S. Gabrielli, V. Osmani, O. Mayora-Ibarra, "Happy or Moody? Why so? Monitoring Daily Routines at Work and Inferring Their Influence on Mood". 5th UbiHealth Workshop, UBICOMP 2010 Conference, Copenhagen, Denmark, 2010.
8. H. Lu et al., “StressSense: Detecting Stress in Unconstrained Acoustic Environments Using Smartphones,” Proc. 14th Int’l Conf. Ubiquitous Computing (UbiComp 12), ACM, 2012, pp. 351–360.
9. Lathia, N., Pejovic, V., Rachuri, K., Mascolo, C., Musolesi, M., and Rentfrow, P. Smartphones for Large-Scale Behaviour Change Interventions. IEEE Pervasive Computing. May 2013.
10. R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong. 2013. MoodScope: building a mood sensor from smartphone usage patterns. InProceeding of the 11th annual international

conference on Mobile systems, applications, and services (MobiSys '13). ACM, New York, NY, USA, 465-466.
11. J. A. Russell. A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), pp. 1161–1178, 1980.
12. S. Ferilli, B. De Carolis, D. Redavid: Logic-Based Incremental Process Mining in Smart Environments. IEA/AIE 2013: 392-401
13. Yuanchao Ma, Bin Xu, Yin Bai, Guodong Sun, Run Zhu. Daily Mood Assessment Based on Mobile Phone Sensing. The 9th International Conference on Wearable and Implantable Body Sensor Networks (BSN2012), May 9-12, 2012, London, UK.

Self-monitoring of Emotions: a novel Personal Informatics Solution for an Enhanced Self-Reporting

Federica Cena, Ilaria Lombardi, Amon Rapp, Federico Sarzotti

Università di Torino – Dipartimento di Informatica
C.so Svizzera, 185 10149 Torino, Italy
cena, lombardi, rapp, sarzotti@di.unito.it

Abstract. Personal Informatics systems help people to collect personally relevant information for the purpose of self-reflection and self-knowledge. Moods and emotions are ones of the most complex data to gather, since a complete automation is impossible due to the need for a cognitive interpretation by the person. The goal of this paper is to present a novel Personal Informatics solution which overcomes some limitations in tracking emotions. To this aim, we propose a complex PI solution based on tangible interface and automatic tracking which support users in tracking their emotions with a limited cognitive load. Our initial efforts on these key directions will be described.

Keywords: Emotion tracking, Personal Informatics, Self-tracking, Tangible Interface

1 Introduction

"How do you feel?" can be a very difficult question to answer. Emotions represent fundamental components of human beings, encompassing physiological, affective, behavioral, and cognitive elements. Differently from moods, emotions are intentional, i.e. imply and involve a relationship with a certain object, and tend to be relatively short lived, while moods are experienced as more diffuse, general and long-term: while moods usually influence which emotions are experienced, emotions often cause or contribute to moods [15]. However, emotions are often contradictory and it is possible to feel different emotions, of different type and intensity, at the same time. Moreover, people tend to report about their emotions on the basis of their belief about them rather than experiences of them [1].

It can be useful to collect emotions for several reasons. Therapists often ask patients to keep track of their changing emotional states (together with other aspects of their daily lives), in order to help with depression and other mood disorders. Recording data helps doctors and patients better understand why symptoms occur, and whether the treatments are working or not. Moreover, people may desire to collect their emotions in order to simply know themselves better, or in order to find correlations among data that suggest a change of their behavior in a particular direction.

Now, new Personal Informatics (PI) tools [2] are making possible to track the ups and downs of people emotional states, and then to aggregate data through some forms of reasoning, feeding them back to user through meaningful visualizations. PI tools track data on different aspects of the daily lives of people (e.g., heart rate, amount of calories burnt, skin temperature, kilometer run, sleeping hours). The data can be gathered automatically in a transparent way with respect to the user or can be self-reported by the user herself, depending of the types of data. It is simple to automatically track physical states (such as glucose level in the blood) or indicators of performance (such as the kilometers run) with the appropriate sensors, while it is almost impossible to collect other data (such as food or dreams). Emotions fall in this second category, since for their nature they are not suited to be automatically detected. In fact, they have a *physical* part (e.g. the arousal, i.e. the physiological reaction to stimuli, which causes changes in physical parameters such as blood pressure, heart rate, temperature), and a *cognitive* part, which interprets the physiological changes and classifies them in a specific emotion [1]. Even if there is a large amount of work trying to infer the emotions from the arousal measured by means of physical parameters, such as heart rate, skin temperatures, eye movements, etc. (see for example Affdex [3] and Emotient [4]), they can only track the physical component of the emotions, but not the cognitive one. Moreover, with the currently available technologies, automatic detection of emotions is very cumbersome for the user, since she must be equipped with a set of invasive devices (such as sensors or wearable devices, like bracelets, helmets, belts, etc.) that make the experience of collecting not natural and not easily applicable to everyday life.

For these reasons, and following psychological literature (e.g. [5]), users' *self-reporting* is necessary to track emotional states, capturing what the user has subjectively experienced. From this perspective, Personal Informatics technologies can be exploited for supporting the self-reporting process, even if they show some practical issues [6].

First, common users may not be so compliant in tracking their own emotions. This issue is also present in clinical settings, where the therapist compels the patient to track her emotions, as Li et al. [7] highlighted. Users can fail to self-monitor themselves due to lack of motivation and time or to forgetfulness.

Second, usually users tend to self-report data after the event to be recorded has occurred. In fact, often it is not feasible for the user to interrupt her activity in order to record what she feels. However, when user reminds of reporting the data, it is often too late to recollect the exact emotional states experienced. This is the case when beliefs trump feelings in self-reporting of emotions [1]. For example, beliefs about the influence of a particular situation (e.g., birthdays are happy), or generalized beliefs about the self (e.g., derived by trait measures of extraversion or neuroticism) or social stereotypes related, for example, to gender (e.g., women are more emotional than men) appear only when the actual experience is relatively inaccessible (later in the time). In fact, memory is reconstructive [8]: with the passage of time, a shift from relatively veridical memories to relatively schematic or stereotypical ones can be observed.

The goal of this work is to find a solution for tracking emotions which addresses these two issues.

First, we aim at *finding new ways for reducing the burden of self-monitoring*. A way for reducing these barriers is to make self-monitoring more fun and enjoyable. We propose to use *tangible interaction* to involve people in self-reporting their emotional states. Tangible User Interfaces (TUIs) leverage physical representation for connecting the digital and physical worlds [9]. TUIs can remind people, by sheer presence, to insert data, motivating users in doing tasks usually perceived as repetitive and burdensome. In fact TUIs showed to be more involving than Graphical User Interfaces when a task is not appealing enough on its own [10], providing a more engaging experience, that can increase the number of repeated activities accomplished by the user [11]. Using TUIs, self-report can become a form of "physical activity", in which objects are manipulated playfully by individuals and contextually gather information about their emotional states.

Second, we aim at *supporting users in the retrospective reconstruction of emotions*. Since people's reports of their emotions reflect whatever information is accessible at the time [1], we aim at providing people with some hints in order to recall the experience where the emotions arose. We will provide user with some other contextual data automatically detected by sensors during the day (time, place, people), in order to recall her the situation in which she experienced an emotion. This would allows her to connect her emotional states to the places visited, the people met and the task accomplished, and, through them, remember what happened to her during the day and report more faithfully her emotions, in a way as much similar as possible to how it actually happened.

In this paper we present a *novel solution for tracking emotions* which reaches these goals. To this aim, we propose a complex PI solution based on tangible interfaces, enhanced with data automatically detected by sensors which support users in tracking their emotions, without a too much cognitive load in the tracking process. Moreover, the solution will be able to integrate all these data and, on the one side, provide users with a meaningful picture of them and, on the other side, find correlations among them. Moreover, we want to enhance the User Model with these correlations that can be used to provide user recommendations about her daily habits in a logic that tries to support her wellness and happiness.

The paper is structured as follows. It first presents a theoretical background of emotion self-reporting in psychology and a brief review on the related work dealing with tool for emotion tracking. Then, our proposed solution is presented and the work in progress is described.

2 Theoretical background

Clinical and experimental psychology has a long tradition in tracking people's emotions. However, emotional experience, defined as the conscious representation of changes in the states of the functional subsystems of the organism, represented by evaluation of an antecedent situation, physiological change, motor expression, motivational effects with prepared action tendencies and subjective feeling state, can only be studied via the introspective report of the subject [5]. In fact, even if it can be possible to obtain physiological measurements and objective measures of the behaviors expressed by the emotions, it is not possible to measure the way in which the subject experiences the physiological and behavioral changes other than through self-report [5]. In psychology several techniques are available for measuring emotions through self-report. A common approach presents participants a checklist of adjectives asking them how well they describe their emotional states [12]: such lists can comprehend terms such as calm, nervous, bored. Another approach relies on dimensional theories of emotion and mood, asking people to rate one or more dimensions of their emotional states, such as arousal (activation) and valence (pleasant/unpleasant) [13]. Sometimes the instructions ask for reports of immediate feelings, sometimes for feelings experienced in a recent period, sometimes for feelings experienced over long periods [14]. However, questions about emotions and mood often refer to past emotional states relying on imperfect and biased memory; otherwise, asking a subject to self-report her emotions when they occur inevitably interrupts the experience [15]. In addition, questionnaires and more in general self-reporting activities commonly employed in psychology are burdensome and time-consuming, making difficult to imagine their usage in the everyday life of people that want to keep trace of their emotional states without having therapeutic motivations or impellent needs.

3 Related works on Emotion tracking

There are a number of applications, research works and technological tools for the emotion tracking, either self-reported or automatic.

Regarding *self-reporting*, a lot of systems collect user's emotions for therapeutic and rehabilitation purposes, such as Mobile Mood Diary [16], a mobile and online symptom tracking tool for adolescents with mental problems.

Commercial applications and devices (e.g., T2 Mood Tracker [17]) (see [18] for an overview) mainly aim at promoting a deeper self-knowledge through a visual exploration of the gathered data. Moreover, they are able to suggest patterns, trends and correlations between emotions changes and habits or occurred events. Other apps, like Mood Panda [19], have an additional social component: users can share their mood with friends and support one another. All these apps force the user to suspend her current task to interact with the phone and this makes tracking burdensome and annoying and the risk exists that in the long term the user gives up.

Many other systems *automatically* track users emotions, such as PSYCHE [20], a personal monitoring system based on textile platforms and portable sensing devices; Fractal [21], which is composed of some gem-like structures and detects wearer's muscle tension and movements, as well as the presence of near people, and adjusts the integrated LED lights pulsing accordingly and the Textile Mirror [22], which is a wall panel made of felt that changes its textural structure according to emotional signals coming from its viewer. An alternative way to detect user's emotions is to interpret facial changes. This is the approach adopted by Affdex [3] and Emotient [4] which is about to release a Google Glass app that analyses subtle pattern changes in a person's face.

Our solution supports users in self-collecting emotions by providing contextual information as hints for remembering emotions. This is different from other systems which are able to automatically collect contextual information and to suggest a relationship between them and the tracked emotions, such as the above mentioned

Mooditoode, which automatically detects position, and Emotion Sense [23], which correlates user's emotions with other factors such as time of day, location, physical activities, phone calls and SMS patterns. Anyway, they do not make the self-reporting activity easier since the user still need to interact with the app to explicitly declare her emotions. Our aim is to facilitate this task for the user, by making it straightforward with a tangible interface.

4. A proposal for a personal informatics system for tracking and remembering emotions

We propose a Personal Informatics system able to support users in self-reporting of emotions. The solution will have the following features:

- it will allow the self-reporting of emotions in an amusing, simple and appealing way by means of a tangible interface
- it will allow to automatically collect contextual aspects related to the emotions: location, time and people in the surrounding when the emotion occurs that will help users in recalling the emotion
- it will provide these contextual information to users to help them in live again emotions in a way as more similar as it actually happened
- it will be able to feed back to users a complex aggregated picture of the emotions of a period of time or of an experience, and correlations among data.

The idea is to create a portable, entertaining and, above all, not burdensome platform that will be composed by several parts: a mobile application on the user smartphone to automatically gather contextual data and some TUIs, i.e. a set of physical objects that the user can manipulate in order to communicate with the system. The TUIs, built on an Arduino board, are used by users to provide her emotions and the moment of the day. It has been decided to monitor 8 different emotions selected as primal emotional states. When user is going to report her emotions, she manually selects in the TUI the time in which the emotion is occurred and the system automatically recollects the context (place and people) in which that emotional state happened, inferring it by e.g. the GPS sensor of user's smartphone, users' social networks (Facebook, Twitter, WhatsApp, Google +), shared calendars (Google calendar, Facebook), etc. Moreover, the platform will be able to receive further information from other devices able to automatically detect physiological indicators (such as heart rate, body temperature, etc.). Data collected will be kept on the server and the user can browse her emotional history, having a representation of her emotional states through different points of view, and inspecting correlations among data.
Fig 1 shows the system architecture. It is composed of two TUI hardware components: the time module and the emotion module. The last one is the core of the system and one of its tasks is to manage the context information. To this aim, the context manager gets the time information that the user sets on the time TUI and infers the context taking into account information coming from different sources such as user's smartphone, social networks, PI tools, etc. Then, the emotion chosen by the user on the dedicated TUIs is collected by the data manager and sent to a remote server with the corresponding context for elaboration.

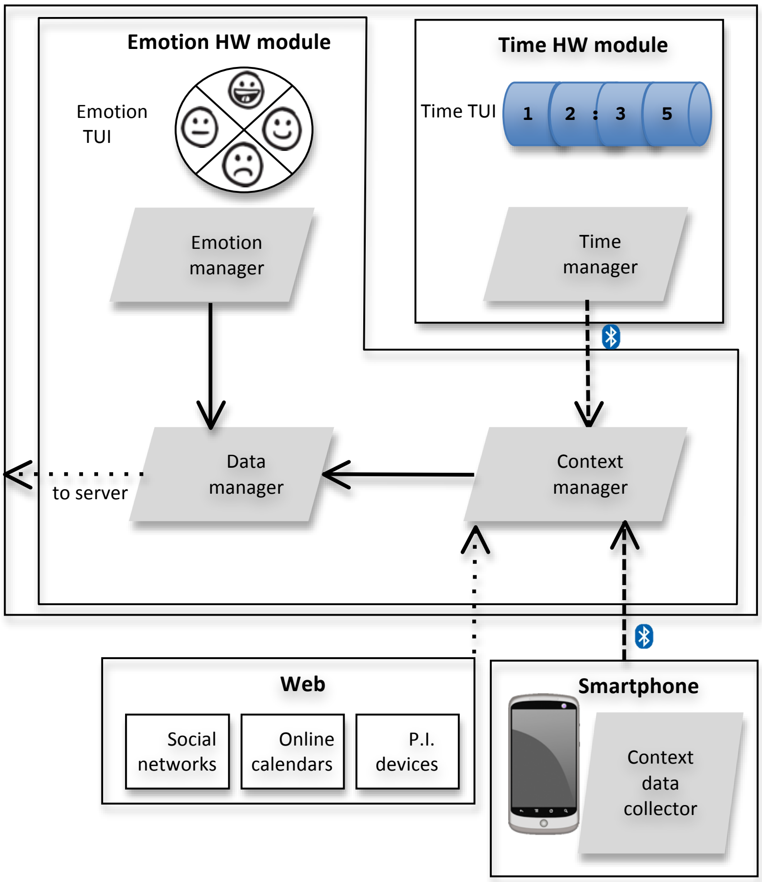

Fig 1. The system architecture

The emotion hardware module communicates via Bluetooth with the smartphone and the time hardware module and via WiFi/Ethernet with the remote server and the data sources exploited to infer the context (social networks, online calendars, personal information tools, etc.).

The final goal of the work will be the design of an enhanced *User Model* to be used for promoting behavior change in some directions (e.g. more healthy lifestyle): i) gathering heterogeneous types of user data (emotions from the prototype but also other data from sensors, from social web activities); ii) reasoning on the data in order to find aggregations and correlations, iii) providing users with personalized *recommendations* in accordance on the UM and meaningful *visualization of data* for rising awareness and motivating people in changing their behavior.

References

[1] Clore, G.L., Robinson, M.D. Knowing our emotions: How do we know what we feel? In Vazire,Wilson (Eds.). Handbook of Self-Knowledge, NY: Guilford Press, (2012), 194-209

[2] Li, I., Dey, A.K., Forlizzi, J.: A Stage-Based Model of Personal Informatics Systems. In: SIGCHI Conf. on Human Factors in Computing Systems. ACM (2010), 557-566

[3] Affdex. Available at: http://www.affdex.com/technology//

[4] Emotient. Available at: http://www.emotient.com/

[5] Wallbott , H. G. and Scherer, K. R. Assessing emotion by questionnaire. In Plutchik, R. and Kellerman, H. (Eds.) Emotion: Theory, research and experience, Vol. 4 The measurement of emotions. San Diego, CA: Academic Press, Inc. (1989), pp. 55-82.

[6] Rapp, A., Cena, F. Self-monitoring and Technology: Challenges and Open Issues in Personal Informatics, to appear in: HCIInternational, Crete, June 2014

[7] Li, I., Dey, A.K., Forlizzi, J.: Understanding My Data, Myself: Supporting Self-Reflection with Ubicomp Technologies. In: Ubiquitous computing. ACM, NY (2011), 405–414

[8] Bartlett. Remembering: A study in experimental and social psychology. Cambridge, UK. Cambridge University Press, 1932

[9] Shaer, O., Hornecker, E., Tangible user interfaces: past, present, and future directions. Found. Trends Hum.-Comput. Interact. 3, 1–2 (2010), 1-137.

[10] Xie, L., Antle, A., Motamedi, N. Are tangibles more fun? Comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In Proc. TEI 2008 (2008), 191–198.

[11] Zuckerman, O. and Gal-Oz, A.. To TUI or not to TUI:. Evaluating Performance and preference in tangible vs. graphical user interfaces. Int. J. Hum.-Comput. Stud. 71, 7-8 (2013), 803-820.

[12] Lorr, M. Models and methods for measurement of mood. In Plutchik, R. and Kellerman, H. (Eds.) Emotion: Theory, research and experience (Vol. 4 The measurement of emotions). San Diego, CA: Academic Press, Inc. (1989), 37-53.

[13] Lang, P. J. The emotion probe. American Psychologist, 50(5), (1995), pp. 372-385.

[14] Plutchik, R. Measuring emotions and their derivatives. In Plutchik, R. and Kellerman, H. (Eds.) Emotion: Theory, research and experience (Vol. 4 The measurement of emotions). Sand Diego, CA: Academic Press, Inc. (1989), pp. 1-33.

[15] Brave, S. and Nass, C. Emotion in human-computer interaction. In The human-computer interaction handbook, Julie A. Jacko and Andrew Sears (Eds.). L. Erlbaum

[16] Matthews, M., Doherty, G.: In the mood: engaging teenagers in psychotherapy using mobile phones. In: CHI2011 ACM, New York, USA (2011), pp. 2947-2956.

[17] T2 Mood Tracker. Available at: http://www.t2.health.mil/apps/t2-mood-tracker

[18] Marcengo, A., Rapp, A.: Visualization of Human Behavior Data: The Quantified Self. Book Chapter in Huang L. H. and Huang, W. (Eds.) Innovative approaches of data visualization and visual analytics. IGI Global, Hershey, PA (2013), pp. 236-265

[19] Mood Panda. Available at: http://www.moodpanda.com/.

[20] Valenza, G., Gentili, C., Lanatà, A. ,Scilingo, E. P.: Mood recognition in bipolar patients through the PSYCHE platform: Preliminary evaluations and perspectives. Art. Int. in Medicine 57, (2013) 49–58

[21] Fractal. Available at: http://www.design.philips.com/about/design/designportfolio/design_futures/fractal.page

[22] Davis, F, Roseway, A., Carroll, E., Czerwinski, M: Actuating mood: design of the textile mirror. In Conf. on Tangible, Embedded and Embodied Interaction, (2013) 99-106

[23] Lathia, N., Pejovic, V., Rachuri, K. K., Mascolo, M., Musolesi, M. Rentfrow, P. J.: Smartphones for large-scale behaviour change interventions. Perv. Comp 12(3), 66-73 (2013)

Social Media Sources for Personality Profiling

David N. Chin and William R. Wright
{chin, wrightwr}@hawaii.edu

University of Hawai'i at Mānoa
Department of Information and Computer Sciences
1680 East-West Road, POST 317, Honolulu, Hawai'i 96822 USA

Abstract. Social media provide a rich source of author-identified text that can be used for personality profiling. However, differences in length and number of entries, syntax, abbreviations, spelling and grammar errors, and topics can affect type and difficulty of preprocessing to extract appropriate text, accuracy of training, time period sampling for training texts, and rate of degradation of accuracy over time. Biases introduced by the topic areas of the social media, author self-selection, and current events affect different social media to varying extents and can bias both the demographics and possibly the personality types of users.

1 Introduction

Many researchers have been able to predict personality from written text [6, 1, 5, 3, 4, 8, 2, 9, 10]. Recent social media outlets provide a rich source for author-identified text that can be used for this purpose. However, the different social media outlets each have different characteristics that will likely affect their effectiveness for personality profiling. For example it is doubtful that any single personality classifier or choice of training features will provide the best results for all social media because of their many differences. This paper will catalog those characteristics of current social media outlets that are expected to affect personality profiling and their implications for personality prediction.

Characteristics of social media likely to affect personality profiling include:

- word length of entries
- number of entries/author
- author identification
- spelling and grammar errors
- topic bias
- time-period bias
- author self-selection bias
- legal access and privacy restrictions
- unusual syntax, usage, abbreviations

The product of the word length of entries and the number of entries per author yields the amount of text available for personality profiling. Relatively small amounts of text can be used to predict personality. For example, [9] found they could achieve more than 81% accuracy with more than 10 email messages. Although [9] do not specify the word length of a typical email message, one study [11] found an average of 3150 characters and a median of 1304 characters for email messages. So 10 email messages would average about 31K characters. [10] were able to achieve an average accuracy of over 80% with essays averaging 787 words/4002 characters in length. However, to train machine learning classifiers, larger amounts of text are preferable.

Another important characteristic of social media is whether authors are identified or not. Some social media types or outlets allow anonymous postings. For example, reviewers on review sites are often anonymous and Myspace allows anonymous members (Facebook does not). Also, some social media forms allow quoting text from other authors, such as including the original email message in a reply email, reTweeting someone else's Tweet, or quoting a previous post in a forum. Automated personality profiling programs need to be able to identify the author of the text to properly attribute personality traits derived from the text. Author identity is also required for accumulating training data because the authors will need to complete personality assessments. Researchers must also consider both the legality and privacy concerns associated with accessing certain social media. For example, the terms of use for Facebook prohibit scraping content via bots. SMS (Short Message Service) messages are typically unavailable in countries with privacy laws.

The type of social media also affects many qualities of the text. Limits on post length such as on Twitter can lead to unusual grammatical usage, which can affect personality profiling. Also different social media may develop abbreviations and other conventions that may not appear in other forms of text media. The particular social media may affect frequency of spelling errors, which can cause problems for text analysis.[1] The social media outlet may also bias the topic, time-period, and author background. For example, web forums typically have a specific topic and the bias in topic may affect how personality correlates with the written text. Likewise, the time-period may affect the relevant topics of discussion in particular social media. Topic differences will affect the word choice, one of the most important features correlated with personality. Finally, the users of social media self-select to use the media and are not representative of the general population. Usage of social media tends to be higher among women, younger users, and urban users [7]. This may be fine if the researcher is interested in the exact demographic represented by the users of the particular social media, but extreme caution is required when trying to generalize any findings to other populations.

2 Social Media Peculiarities

Current social media that can be mined for personality profiling include:

[1] Counting the misspellings offers a useful feature in itself (extraverts tend to leave their text unedited), yet the misspellings make it difficult, on an automated basis, to correctly identify topics, grammatical structures, and of course the words that are being used.

– Emails
– Twitter
– Facebook Status and Wall
– Myspace Bulletins
– Forums
– LinkedIn
– Review sites
– YouTube
– SMS (Short Message Service)
– Blogs and microblogs (e.g., Tumblr)

Email messages have a specified format and requires email-specific preprocessing to extract author text. [9] describes removing headers, signature blocks, and automatic reply quotes. Emails and forums are conversational, so speech acts can be important features that are not found in monolog social media like blogs and to a lesser extent, Twitter (@replies in Twitter allow some conversational threads). Forums often use common forum software, many of which provide an API (Application Programming Interface) to directly extract posts and other information from the forum software without needing to scrape web pages. For example, phpBB, the current most popular forum software provides the REST API to communicate with a phpBB board. Likewise many common blog software packages provide APIs to access blogs and microblogs such as the Blogger JSON API and the Tumblr API.

Review sites tend to be poor choices for personality profiling because authors can be anonymous and reviewers tend to not write very many reviews. The specific purpose of review sites also biases the text making generalization to other contexts very problematic. YouTube suffers from the same anonymous posters problem, lack of sufficient text per poster, and bias based on topic of the posts. LinkedIn does have identifiable authors, but suffers from very little text per author and a very narrow topic area (the career and expertise description).

For some media and occasions, scarcity of time, limits of technology, costs of transmission, purpose of the medium, and etiquette certainly encourage brevity and directness. For example telegrams and radio-communicated Morse Code messages tended to be brief. So are modern SMS messages and Tweets. Blog posts, though, tend to be wordier and more contemplative. It is in such settings, wherein writers express themselves at length, that we expect grammatical choices to vary more, and to tell us more about the personality of the writer. With no rigorous analysis, we will examine a few examples. First from Twitter.com:

> "Still trying to go scuba diving in Mexico for my birthday this summer. Yet, no one is down for the adventure!" * "Why I went #scuba diving with crocs (video). Actual #underwater croc footage starts at 5min 15sec: http://ow.ly/vfczo" * "Wish I was scuba diving chillin with some dolphins and jellyfish instead of being in the presence of these nerds" * "Would you take a scuba dive in the lake that they say the lochness monster is in for 1000 dollars.?" * "Went scuba diving with my dad in the Great Barrier Reef #rad @bvnwnews pic.twitter.com/ahNiLgGGx6" * "@ZoomTV @Ileana_Official scuba dive! I tried once but failed! But will try again soon"

Next some blog posts:

“The vibration became so intense, I could feel it in my bones, and the sound turned into a deafening roar. I could see waterfalls of sand pouring over the coral, and on the sea floor, a few metres below us, cracks began forming and the sand was sucked down. That’s when I realised it was an earthquake. The noise was the sound of the Earth splintering open and grinding against itself.” - Jessica Read, The Guardian, 24 January 2014

“Suspended in limbo, 130 feet from the surface and nearly 100 feet to the sandy bottom, I watched the bubbles. They playfully danced around each other expanding, breaking, conjoining, chaotic, but always up. The ever-changing surface glimmered above — where air meets water, where life meets death. ...My consciousness crept out of its silent prison and I looked at my gauges.” ... “Where had the time gone? I thought. My dive computer started to flash things I had never seen before.” - Kelsey, http://tinyurl.com/kbr84tp

“I was surrounded by more fish than I could count, my eyes unaware of where to look next. I had totally forgotten all about my breathing and equalizing and adjusting my buoyancy level, it all didn’t matter anymore. (Well, I guess I did alright since I made it back, thankfully my guide was there for the constant check.) Schools of fish swam through us constantly, and I found myself honestly enjoying where I was, meters beyond meters below sea level. Even though my ears hurt like hell at some points, the fascinating colors of the life underwater kept me from signaling to my instructor that I wanted to go back on land.” - http://tinyurl.com/mxakoyf

“When I built valves at AirForce, I tested each by pressurizing them in a fixture and tapping the valve stem with a rubber hammer. I had racks of 100 valves at a time, and I went through and did this to each one in turn. That process seated the valve and created a small ring of contact between the synthetic valve and its seat. Sometimes, the valve needed to be hit several times to seat it properly, but it always worked. And it also worked if a valve had a small piece of dirt anywhere in the seals.” - http://tinyurl.com/kqcqgxo

First we immediately note that the Tweets employ very few adjectives. The Tweets contain many errors, even though most of the problems could easily be fixed without exceeding the 140 character limit for a Tweet. Beyond simply determining whether or not the samples are within norms of English usage (does the parser fail?), these errors will complicate attempts to discern individual differences between Twitter users on the basis of grammatical usage. Also the topics and content of Tweets are somewhat constrained by the medium: for example very few Tweets explicitly set forth instructions for users to follow. The broadcast nature of Tweets seems to motivate some users to adopt a style that is maximally accessible to their followers. The wideness of the audience further constrains the expression within short Tweets.

Although they too have a broad audience, blog authors exhibit greater variety in style than the Tweet authors do, perhaps because they have ample space in which to do so coherently. Naturally there are more anaphoric expressions, and much more time is spent on descriptions of objects and processes rather than simply naming them. The use of subordinate clauses varies amongst the blog posts. All these desirable attributes can be applied to personality profiling.

3 Implications for Personality Profiling

Each type and outlet of social media has idiosyncratic characteristics that require specialized processing to extract appropriate author-attributed text. Even after the author-attributed text has been extracted from the social media, there are still other characteristics of each type and outlet of social media that should be taken into account for personality profiling. For example, social conventions can vary across different social media. Emails include various greetings and execute a variety of speech acts that are less common in other social media. Tweets tend to be much more declarative because of the broadcast nature of Twitter. On the other hand, blogs and microblogs are also one-to-many, but exhibit much less of the declarative nature of Tweets. These differences in social conventions can potentially affect the cross-applicability to other social media types for personality classifiers trained in a different type of social media.

Besides different specialized conventions for different social media types, social media types also differ in how quickly these conventions change over time. For example, SMS messages tend to display greater variety over time than perhaps Tweets do, due to rapid evolving of new “text speak” - expressions, abbreviations, emoticons, etc. Although there has been some encouraging success predicting personality from SMS messages [2], a personality predictor depending upon rapidly changing aspects of language may need to be constantly retrained to avoid degradation due to the changes in conventions over time.

Different types of social media are produced under different circumstances and to different audiences. These differences lead to differences in the number of spelling and grammatical errors in the text. SMS messages, and Tweets tend to have the most errors and blogs tend to have the least with other social media types falling in between these extremes. Such differences affect the effort that might be needed to correct errors when processing text for personality profiling. Independent of whether errors are corrected, error rates (before correction) might provide a valuable feature for predicting personality and its value may vary among social media types.

Different kinds of information are conveyed across different social media. Some social media like Twitter are much more topical, dealing more with current events and trends, than other media like blogs and forums, which tend to be on specific topic areas that are less influenced by current events. The specific topic of a blog or forum can influence topicality. For example, political blogs and forums will be much more topical than blogs/forums about topics like parenting that are less influenced by current events. This means that training on text from some social media may need a wider sampling of time periods to avoid over-fitting on topical peculiarities that appear rarely in other time periods.

Different social media differ in the amount of text per author that are typically available. This has implications for the potential accuracy of training personality classifiers. Tweets, SMS text messages, Myspace bulletins and Facebook updates/wall posts tend to have shorter posts. Of these, SMS messages tend to be the most prolific, so many SMS messages over time can easily add up to enough text for accurate training. LinkedIn review sites, YouTube and email tend to have medium sized entries. However LinkedIn entries tend to be single entries, so will not have enough text per person for good training. On the other hand, email is sent constantly over time, so accumulating these can easily provide enough text for training. Review sites may have multiple entries per person, so may provide enough text for training depending on the productivity of the reviewers. However most review sites tend to have very few prolific reviewers, so review sites likely will not have enough authors with enough text for training. Likewise, there may be enough accumulated text accompanying prolific YouTube posters for good training, but more investigation is needed to determine whether there are enough prolific YouTube posters for training. Blogs, Forums, and microblogging present the best sources for large amounts of text by many authors for training purposes.

Users of the different social media types self-select to use that particular form of social media. This will certainly skew the demographics of the studied users. Even more problematic, there are no studies about whether particular personality types are more or less likely to use particular social media. Thus not only are the demographics skewed, which might be corrected with appropriate sampling techniques, but also there might be as yet undocumented biases in personality types introduced because of the selection effect. Likewise, the topic area of a forum, blog, microblog, or YouTube video may bias not only the demographics, but also the personality types of authors.

4 Conclusion

The variety of linguistic expression seen in the blog entries encourages personality profiling applications, whereas inflexible conventions forced by brevity of Tweets tend to narrow the range of linguistic choices. Researchers are then motivated to adapt their methods by focusing on the particular aspects of a given medium that are most useful for personality prediction. Also, some social media, specifically SMS text, exhibit rapid changes in the specialized language employed by users. A classifier trained on unstable aspects of language will quickly degrade in usefulness. This phenomenon enhances the need to identify and exploit those aspects of language usage that change slowly within a given social media context.

References

[1] S. Argamon et al. “Lexical predictors of personality type”. In: *Proceedings, Interface and the Classification Society of North America*. 2005.

[2] T. Holtgraves. “Text messaging, personality, and the social context”. In: *Journal of Research in Personality* 45.1 (2011), pp. 92–99.

[3] F. Mairesse et al. “Using linguistic cues for the automatic recognition of personality in conversation and text”. In: *Journal of Artificial Intelligence Research* 30.1 (2007), pp. 457–500.

[4] S. Nowson. “Identifying more bloggers: Towards large scale personality classification of personal weblogs”. In: *In Proceedings of the International Conference on Weblogs and Social*. Citeseer. 2007.

[5] J. Oberlander and A.J. Gill. “Language with character: A stratified corpus comparison of individual differences in e-mail communication”. In: *Discourse Processes* 42.3 (2006), pp. 239–270.

[6] J.W. Pennebaker and L.A. King. “Linguistic styles: language use as an individual difference.” In: *Journal of personality and social psychology* 77.6 (1999), p. 1296.

[7] Online source PewResearch. *Retrieved*. Apr. 2014. URL: `http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/`.

[8] A. Roshchina et al. “A comparative evaluation of personality estimation algorithms for the twin recommender system”. In: *Proceedings, Workshop on Search and mining user-generated contents*. ACM. 2011, pp. 11–18.

[9] Jianqiang Shen et al. “Understanding Email Writers: Personality Prediction from Email Messages”. In: *User Modeling, Adaptation, and Personalization*. Springer, 2013, pp. 318–330.

[10] William R. Wright and David N. Chin. “Personality profiling from text: Introducing Part-of-speech N-grams”. In: *User Modeling, Adaptation, and Personalization*. Springer, 2014.

[11] Online source www.activityowner.com. *Retrieved*. Apr. 2014. URL: `http://www.activityowner.com/2009/03/14/how-many-characters-are-there-in-a-typical-email-message/`.