Using social media mining for estimating theory of planned behaviour parameters

Marko Tkalčič[12], Bruce Ferwerda[1], Markus Schedl[1], Cynthia Leim[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 (Ekipa2), Ljubljana, Slovenia

Abstract. In this position paper we present the scenario of users who do not attend classical concerts. 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 microblogs 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. The study suggested that some of the barriers that users provided (e.g. social aspects, 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 [markot TODO]: generic description of the TPB model
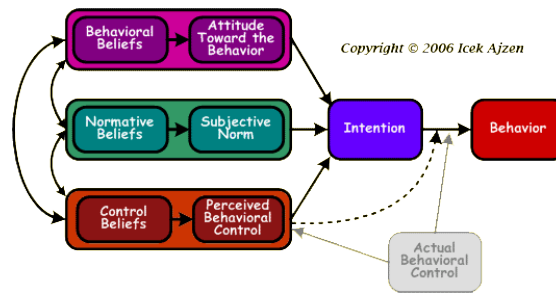

Fig. 1. The Theory of Planned Behaviour model (source: (1))

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 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. These concert are usually advertised, so 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.

We plan to make predictions of the numerical values of the two parameters by social media mining, more specifically, through microblog (i.e. twitter) mining. 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 tecnical 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
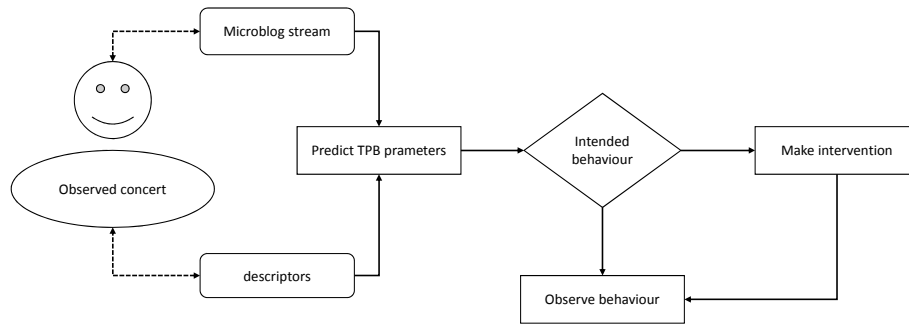5. compare it to the predicted behaviour

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 uf 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 features from Tab. **??** we will 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

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

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 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

Bibliography

[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`