

A two-step, user-centered approach to personalized tourist recommendations

Valeria Orso
University of Padova
Padova, Italy
valeria.orso@unipd.it

Alessandra Varotto
University of Padova
Padova, Italy
alevarotto@hotmail.it

Stefano Rodaro
University of Padova
Padova, Italy
stefano.rodaro821@gmail.com

Anna Spagnolli
University of Padova
Padova, Italy
anna.spagnolli@unipd.it

Giulio Jacucci
University of Helsinki
Helsinki, Finland
giulio.jacucci@helsinki.fi

Salvatore Andolina
Aalto University
Helsinki, Finland
salvatore.andolina@aalto.fi

Jukka Leino
University of Helsinki
Helsinki, Finland
jukkatapanileino@gmail.com

Luciano Gamberini
University of Padova
Padova, Italy
luciano.gamberini@unipd.it

ABSTRACT

Geo-localized, mobile applications can simplify a tourist visit, making the relevant Points of Interest more easily and promptly discernible to users. At the same time, such solutions must avoid creating unfitting or rigid user profiles that impoverish the users' options instead of refining them. Currently, user profiles in recommender systems rely on dimensions whose relevance to the user is more often presumed than empirically defined. To avoid this drawback, we build our recommendation system in a two-step process, where profile parameters are evaluated preliminarily and separately from the recommendations themselves. We describe this two-step evaluation process including an initial survey ($N=206$), and a subsequent controlled study ($N=24$). We conclude by emphasizing the benefit and generalizability of the approach.

CCS CONCEPTS

• **Human-centered computing** → **User studies**;

KEYWORDS

Tourist applications, user validation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIItaly '17, September 18–20, 2017, Cagliari, Italy

© 2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-5237-6/17/09...\$15.00

<https://doi.org/10.1145/3125571.3125594>

ACM Reference format:

Valeria Orso, Alessandra Varotto, Stefano Rodaro, Anna Spagnolli, Giulio Jacucci, Salvatore Andolina, Jukka Leino, and Luciano Gamberini. 2017. A two-step, user-centered approach to personalized tourist recommendations. In *Proceedings of CHIItaly '17, Cagliari, Italy, September 18–20, 2017*, 5 pages. <https://doi.org/10.1145/3125571.3125594>

1 INTRODUCTION

Current mobile technologies allow each individual to undertake a visit with an endowment of navigation and survival tools that would have been unthinkable some decades ago. Such technologies allow visitors to directly contact local people without intermediaries, to translate language, and to receive expert advice. All users are then enabled to develop with autonomous resources their knowledge of a territory without necessarily relying on professional travel agencies and organizers. Location-based services that provide personal recommendations about Points of Interest (PoIs) are a key feature of mobile applications for tourists [7, 9, 15, 17]. They are personalized recommender systems, which tailor the search result to the users' characteristics [8, 12] instead of generically ranking the items [4, 5]. Information about the users' preferences is stored in a user profile [10], which is a proxy of the user's demographics, expertise, interests, usage behaviors and intentions [3]. Appropriateness of the profile and of the profiling dimensions on which recommendations are built is a crucial requisite for an effective personalization service, since the system filters out information that is supposedly irrelevant to the user. Despite users' satisfaction with the recommendations does not necessarily correlate with their accuracy [14], the typical assessment of a recommender system aims at proving the quality of the recommendations

using statistical methods, without the involvement of users [6]. Even less effort is devoted to the evaluation of the relevance of the attributes on which the recommender system relies [9]. Typically, they are derived from the same structure adopted in the database that the recommender system is mining. For instance, [2] use Tripadvisor tourist reviews and rely on the profile attributes categorized by that web service, i.e., age, gender, country of origin, goal of travel, and travel partners. Similarly, [20] rely on the profile categories present in the tourist emergency report on which the algorithms are run: object, environment, activity, event, result. Specific research effort devoted to defining such categories is very rare. The result is that in an attempt to save a few resources and time, the recommender system might be based on presuppositions that are never tested with the target users, contrary to all user-centered principles that have been advocated in HCI so far, and that would ultimately preserve systems from rejection and market failure. In the present work we attempted a different approach by first grounding the user profiling attributes on data gathered in the field and secondly by validating the actual relevance of the recommendations provided by the system in a controlled user study. In the remainder of the paper, within the space allotted, we report on the two-step procedure followed to validate the users' profiling characteristics and to assess the relevance of the recommendations returned by the system.

2 STEP 1: PROFILING DIMENSIONS

In order to create user-centered profiling dimensions for tourist recommendations we consulted with tourists themselves, and stakeholders, checked local reports on tourist data and read scientific publications on tourist preferences. We were looking for tourist characteristics that account for visiting preferences. Once these characteristics were drafted, we went back to tourists to evaluate their ability to account for visiting preferences.

Preliminary profiling dimensions

54 semi-structured interviews (mean age=44, $SD=17$, 28F) were carried out with city visitors. Respondents were approached in three central touristic areas by interviewers wearing a university badge. Visitors were explained the purpose and procedure of the interview. If they agreed to be interviewed, they signed the informed consent. Respondents were asked to describe their touristic experience in the city, the sights visited, the motivation of the visit and length of the stay, as well as some demographic information. The interviews lasted about 5 minutes each, were audio-recorded and then transcribed. Stakeholders were seven people from the public and private sectors who are in touch with tourists and visitors on a regular basis: the Director of the Tourism Promotion Office, the Tourism Councilor, two tourist guides,

a representative of the tourist promotion consortium, a representative of a local Shopkeepers and Restaurateurs Association and the head of regional Convention and Visitors Bureau. Respondents were asked to sign the informed consent and were then interviewed with a semi-structured protocol for about 30 minutes. The interviews focused on the visitor types in that city. Each interview was video-recorded and then transcribed. All interview transcripts were analyzed along with scientific literature and local tourist statistics in order to find relevant information about visitors' different preferences. Two researchers worked individually on the whole set of data and then compared their notes to identify a common list of emerging themes. Such information allowed to identify the preliminary attributes characterizing the different types of tourists and their preferences, i.e., age, provenance, travel purpose, budget, length of stay, season, special needs, transportation means and interests.

Validated profiling dimensions

The preliminary profiling dimensions and a list of visiting preferences were put into a 40-item questionnaire. The initial part of the questionnaire addressed the respondents' demographics (items 1-6), and the visit purpose, length, organization and planning (items 7-13). Items in this section were mainly a multiple-choice format. The second part of the questionnaire (items 14-40) investigated tourists' motivations and preferences about their travel. In this section, respondents answered by selecting their level of agreement with each statement on a 6-point scale. The researcher, wearing a university badge, approached the visitors in the city center and asked them to fill in a brief, anonymous questionnaire. If the approached visitor agreed to participate, s/he first signed the informed consent and then filled in the questionnaire.

Participants

In total, data from 206 respondents (mean age=44, $SD=16$, 120 F) were analyzed. 43 questionnaires were excluded due to incompleteness or unreliable response style.

Analysis

The effect of the preliminary profile attributes identified (i.e., age, provenance, travel purpose, budget, length of stay, season, special needs, transportation means and interests) on the participants' preferences, as reported in the questionnaire, was examined. The analyses aimed at assessing in which cases a given dimension was effective in affecting visit preferences (e.g., younger visitors would favor cost-effective accommodation and activities). Various data analysis techniques were used based on the question format and, consequently, the types of data collected: contingency tables and count of frequencies, ANOVA and non-parametric tests

(Mann-Whitney, Kruskal-Wallis). Table 1 reports the dimensions and preferences relations that resulted significant from this analysis.

3 STEP 2: USERS' ASSESSMENT OF PERCEIVED RELEVANCE

The second step in evaluating the recommender system was to assess the relevance of the recommendations generated. We hypothesized that the recommendations based on the personal profile would be rated as more relevant than those generated on other criteria. In particular, the recommendation system had a graph-based retrieval model using graph layers associated with content as data sources, and it included four variants: one using only the tags layer (T), one combining tags and personal information overlays (TP), one combining the layers of tags and social rating (TR), and one using tags, social ratings and personal data layers (TRP). The system run on a database containing real data referring to the target city (approximately 600 PoIs). Profiling dimensions were entered by the user him/herself, a useful method in new recommenders that cannot have a history of prior users' behavior [12, 13]. Social rating was instead based on how users rated the PoI contained in the system, and was also entered by users. This study aimed at evaluating users' perceived relevance of the recommendation, with a similar approach of other evaluation studies [1, 5, 6, 11, 16, 18]. However, differently from other researches, the test was run as a controlled laboratory study to control important factors such as the identity of the person actually carrying out the evaluation, and the seriousness with which the evaluation is completed (e.g., whether skimming through the list of recommendations). This method allowed to compare the four variants of the recommendation algorithm as one independent variable in a within-subjects experimental design.

Materials

A brief initial questionnaire collected background information (name, age, education) and participants' knowledge of the main landmarks of the city. The relevance of the recommendations was then assessed using an on-line questionnaire presented on a tablet, in which participants rated the relevance of each PoI recommended by the system on a 5-point scale. In order to have all respondents referring their judgments to the same situation, a task scenario was devised. Such scenario depicted an unambiguous and realistic situation, in which a visitor expressed interest in the religious heritage of the city.

Participants

A total of 24 participants (mean age 26, $SD=6$, 12 F) volunteered in the study, and received no remuneration. Participants reported a fairly good knowledge of religious buildings

($M=3.42$ on a scale ranging from 1 to 5), historical landmarks ($M=3.58$) and cultural sites ($M=3.92$).

Procedure

The study consisted of two sessions, a few weeks apart. The first session served to feed the system with the necessary data, i.e., user profile information and social ratings; the second session was meant to collect users' comparative ratings of the recommendations provided by the 4 versions of the system. Session 1 started when participants gave their informed consent to partake in both sessions and filled in a brief questionnaire collecting background information (i.e., demographic information, familiarity with mobile tourist applications and previous knowledge of the main landmarks of the city). Next, they rated the attractiveness of each PoI presented on a 5-point scale. The PoIs in the database belonged to three areas of the target city, namely Area 1, Area 2 and Area 3. The participants who rated the PoIs belonging to Areas 1 and 2 in Session 1, would rate the relevance of the PoIs belonging only to Area 3 in Session 2, and viceversa (PoIs belonging to Area 2 were not returned in Session 2). By doing this, we avoided situations in which respondents were recommended the same items they had previously rated. In Session 2, participants were asked to query the system for recommendations using all of the four system variants. Once the system had returned the recommendations, participants were asked to examine the list of PoIs and to rate the relevance of each recommended item on a 5-point scale, based on the scenario. They were asked to imagine to be visiting the city alone for religious purposes. The order with which they run the four different algorithms was counterbalanced in a Latin Square design.

Results

After the first session, we compared the ratings assigned to religious vs. non-religious PoIs. The Wilcoxon test revealed no significant differences. In particular, the average rating for the religious PoIs in Area 3 was 1.7 ($SD=1.27$) and that for the non-religious PoIs was 1.46 ($SD=1.05$), $z=1.41$, $p=.16$. Similarly, the average evaluation for the religious landmarks in the Area 1 was 2.02 ($SD=1.21$) and that for the non-religious ones was 1.99 ($SD=1.37$), $z=.16$, $p=.91$. Therefore, it was ascertained that conditions TP and TR in the second session would have actually differed, since the latter would not prioritize religious PoIs, while the former would (being based on the user profile as a religious visitor). After the second session, the perceived relevance of the recommended PoIs in the four algorithms was compared. The analysis was limited to the ratings assigned to the first three items in the list, since they were those prioritized by the algorithm. A repeated-measures ANOVA was run, with the system variant as a four-level factor. The analysis showed a significant difference in the four

VISITOR'S ATTRIBUTE S		Age					Provenance				Purpose		Familiarity		Company							
		under 25	26 to 34	35 to 54	55 to 64	Over 65	West-Eur	East-Eur	Usa Canada	Australia	South-America	Asia	Religious	Non Religious	First time	Returning visitor	Alone	Partner	Family	Friends-Relatives-	Tour Operator	Colleagues
Budget (budget necessary to enjoy a certain tourist attraction or information to access promotions and discounts)	Budget-friendly	**	**			**																
	Budget-free			**	**																	
Fun (information about entertaining activities)	Nightlife	**	**																			
	Food							*								*		*	*			*
	Shopping															*		*	*			*
Areas (nature of a certain area of the city)	Dense					*																
	Distributed																					
	S. Anthony						*				*											
Attractions popularity	Major attractions							*														
	Best kept secrets																					
Flexibility	Open						*		*					**	*				*			*
	Booked																				*	
Special attractions	Children																	*				
	Religion								*	*												

Table 1: Profile attributes and related preferences, * $p < .05$, ** $p < .001$

algorithms $F(3, 69) = 4.99, p = .01$ (Figure 1). Post-test comparisons corrected for Bonferroni revealed that the relevance ratings were statistically higher for TP ($M = 2.79, SD = 1.01$) than for TR ($M = 2.43, SD = .86; p = .04$). In addition, the relevance rating for TP was significantly higher also compared to TRP ($M = 2.51, SD = .95; p = .01$). All other comparisons were statistically non significant.

4 DISCUSSION AND CONCLUSIONS

In the present paper we reported on a two-step procedure for the identification of user-centered profiling dimensions. The first step aimed at applying a user centered design approach to establish the core dimensions differentiating users to the purpose of providing personalized recommendations. The second step assessed whether the profile information

effectively contributed to return pointed recommendations for users. Although our study was contextualized to a recommendation engine in the touristic domain, our approach can be applied to other scenarios of use, e.g., recommendation of books. We advocate a user-centered approach in every step of the design cycle, to avoid building the recommender system on false premises. Although a user-centered approach involves additional effort in recruiting users and analyzing data compared with an evaluation carried out by developers or recommendation accuracy only, it has been long established in HCI that without it user models risk to have very poor ecological validity. Of course a limit of this approach is that in the real context of a visit tourists are attracted by POIs that match some other expectations or can have ambivalent/shifting visit purposes [1, 19]; but then

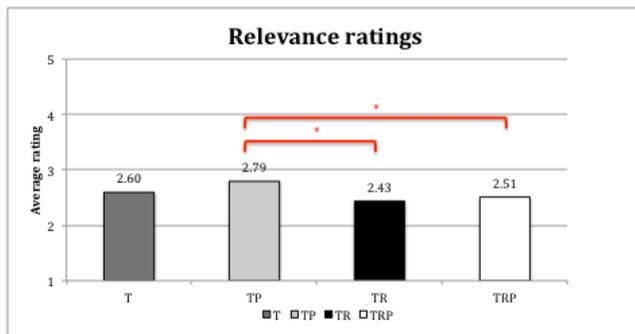


Figure 1: The effect of the different combination of data overlays on the perceived pertinence of the recommendations.
* $p < .05$

this is a matter of how profiles are detected more than of the way in which the profile dimensions are defined in the first place. The importance of involving end users in the assessment of the quality of the recommendations has been already highlighted [14], given the lack of convergence between the evaluations made by computer simulations and by users. Here we suggest to further extend this view and to engage end users also in the designing phase to gather their preferences and needs. We expect that such approach will contribute to develop systems generating recommendations that are relevant in the users' perspective.

ACKNOWLEDGMENTS

The present work was partially funded by European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No 601139 (CultAR).

REFERENCES

- [1] Roger D Abrahams. 1986. Ordinary and extraordinary experience. *The anthropology of experience* (1986), 45–72.
- [2] Atae Rezaei Aghdam, Mostafa Kamalpour, Dong Chen, Alex Tze Hiang Sim, and Jee Mei Hee. 2014. Identifying places of interest for tourists using knowledge discovery techniques. In *Industrial Automation, Information and Communications Technology (IAICT), 2014 International Conference on*. IEEE, 130–134.
- [3] Fedor Bakalov, Marie-Jean Meurs, Birgitta König-Ries, Bahar Sateli, René Witte, Greg Butler, and Adrian Tsang. 2013. An approach to controlling user models and personalization effects in recommender systems. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 49–56.
- [4] Robin Burke, Alexander Felfernig, and Mehmet H Göker. 2011. Recommender systems: An overview. *Ai Magazine* 32, 3 (2011), 13–18.
- [5] Paolo Cremonesi, Francesco Epifania, and Franca Garzotto. 2012. User profiling vs. accuracy in recommender system user experience. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 717–720.
- [6] Paolo Cremonesi, Franca Garzotto, Sara Negro, Alessandro Papadopoulos, and Roberto Turrin. 2011. Comparative evaluation of recommender system quality. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1927–1932.
- [7] A Felfernig, S Gordea, D Jannach, E Teppan, and M Zanker. 2007. A short survey of recommendation technologies in travel and tourism. *OEGAI journal* 25, 7 (2007), 17–22.
- [8] Min Gao, Kecheng Liu, and Zhongfu Wu. 2010. Personalisation in web computing and informatics: Theories, techniques, applications, and future research. *Information Systems Frontiers* 12, 5 (2010), 607–629.
- [9] Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati Pantziou. 2014. Mobile recommender systems in tourism. *Journal of network and computer applications* 39 (2014), 319–333.
- [10] Mouzhi Ge, Dietmar Jannach, Fatih Gedikli, and Martin Hepp. 2012. Effects of the Placement of Diverse Items in Recommendation Lists.. In *ICEIS (2)*. 201–208.
- [11] Rong Hu and Pearl Pu. 2010. A study on user perception of personality-based recommender systems. *User Modeling, Adaptation, and Personalization* (2010), 291–302.
- [12] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. *Recommender systems: an introduction*. Cambridge University Press.
- [13] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4 (2014), 2065–2073.
- [14] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 1097–1101.
- [15] Barbara Neuhofer, Dimitrios Buhalis, and Adele Ladkin. 2014. A typology of technology-enhanced tourism experiences. *International Journal of Tourism Research* 16, 4 (2014), 340–350.
- [16] Jeffrey Parsons, Paul Ralph, and Katherine Gallagher. 2004. Using viewing time to infer user preference in recommender systems. (2004).
- [17] Francesco Ricci. 2010. Mobile recommender systems. *Information Technology & Tourism* 12, 3 (2010), 205–231.
- [18] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction* 22, 4 (2012), 399–439.
- [19] Natan Uriely. 2005. The tourist experience: Conceptual developments. *Annals of Tourism research* 32, 1 (2005), 199–216.
- [20] G. Zhou, J. Lu, C.-Y. Wan, M. D. Yarvis, and J. A. Stankovic. 2008. *Body Sensor Networks*. MIT Press, Cambridge, MA.