Smoothly Extending e-Tourism Services with Personalized Recommendations: A Case Study

Paolo Cremonesi and Franca Garzotto

Department of Electronics, Information, and Bioengineering - Politecnico di Milano
Via Ponzio 34/5 – Milan (Italy)
{paolo.cremonesi,franca.garzotto}@polimi.it

Abstract. Our research explores the influence of recommendations on the quality of the user experience (UX) in the e-tourism domain.. We are interested in the effects of smoothly introducing recommenders in *existing* commercial e-tourism system and to explore the benefits of recommendations in different conditions of *availability* of tourism services (which has a dynamic nature and typically depends on tourism flows in different seasons). The paper presents a wide empirical study (240 participants) that addresses the above issues and has been carried on in cooperation with a large hotel reservation provider (*Venere.com* – a company of *Expedia* Inc.).

Keywords: Recommender systems, E-tourism, Evaluation.

1 Introduction

Theoretical arguments and empirical evidence (mostly in the domain of e-commerce [15]) suggest that Recommender Systems (RSs) improve the user experience (UX) with web services that offer large amounts of digital content. This paper explores this issue in the context of online tourism services. Our work takes into account one peculiar features of the tourism domain, the *dynamic nature* of the availability of services (e.g., accommodations). In online hotel booking, for example, it is likely that most accommodations are fully booked during high-season, and the user has difficulty in finding the solution (s)he is looking for. Users may interpret the scarcity of resources in a given period as a weakness of the catalogue of services and ascribe the phenomenon to the service provider rather than objective contingent situations. This experience can reduce trust on the provider, and even induce users to leave the current online service and try a different one. Hence it is interesting to study the effects of RSs in different conditions of resources availability and to investigate how personalized recommendations play in relation the potentially negative effects of scarcity of resources. In addition, differently from previous works in e-tourism evaluating RSs systems that are *totally* built "*ex novo*", in terms of interface and functionality, we are interested to explore the effects of RSs on the UX from a "conservative" perspective, i.e., when recommendations represent the *only* modification to the interface and functional characteristics of an *existing* service. To address this issue, we carried on an

C. Huemer and P. Lops (Eds.): EC-Web 2013, LNBIP 152, pp. 174–181, 2013.

empirical study on the hotel e-booking service of *Venere.com* (www.venere.com), a company of the Expedia group). Venere.com is one of the worldwide leaders in the hotel booking market, featuring more than 120,000 hotels, bed and breakfasts and vacation rentals in 30,000 destinations worldwide. The goal of our study was to empirically explore the effects, on the UX with Venere.com, of the smooth, "conservative" introduction of recommendations generated with *four different recommendation algorithms* and in *two different conditions of room availability*.

2 Related work

The potential benefits of RSs in e-tourism have motivated some domain-specific researches. Some works used travel recommender systems to emulate offline travel agents [4,12] or to recommend product bundles (e.g., journey, hotel, car rental, packages with multiple destinations) instead of single items [10,12,13,16]. Other works describe conversational recommender systems used in the travel domains [7,14]. Few works investigate which factors can potentially influence the decision-making process in tourism recommender systems [12]. Ricci et al. in [12] developed and tested NutKing, an online system that helps the user to construct a travel plan by recommending attractive travel products or by proposing complete itineraries. The NutKing system was empirically evaluated in a preliminary between subject study showing that users receiving recommendations spent more time in examining information rather than moving through pages. Levi et al. in [9] describes a recommender system for online hotel booking, which was evaluated on 150 subjects. Zanker et al. [16] present an interactive travel assistant designed for an Austrian spa resort. Preference and requirement elicitation is explicitly performed via a sequence of question/answer forms. Field analysis confirmed an increase from 3.5% to 9% in the conversion rate (percentage of users requesting availability information at the spa resort). Delgado et al. in [4] describe the application of a collaborative attribute-based recommender system to the Sky-Europe.com web site, specialized in winter sky vacations. Elicitation is performed using a question-and-answer conversation. Recommendations are produced by taking into account both implicit and explicit user feedbacks. According to a field study, the recommender system was able to quadruplicate the conversion rate (percentage of visitors requesting a quotation for a vacation).

3 The Empirical Study

3.1 Research Hypotheses and Variables

Prior studies (see [2] for a survey) in e-commerce pinpoint that recommendations generated by different algorithms have different effects on the UX. We are interested to explore if the same is true in the hotel online booking domain when recommendations are smoothly introduced in an existing system. In addition, we want to investigate if the dynamic characteristics of resources availability interfere with these effects. Our study focuses on two research hypotheses:

H1: The effects on UX of the "conservative introduction" of personalized recommendations in online booking depend on the algorithm

H2: The effects on UX of the "conservative introduction" of personalized recommendations in online booking are influenced by room availability

We operationalize the *effects* on the UX associated to the introduction of recommendations using the a subjective variable - *choice satisfaction* (measured using a questionnaire) and 3 objective variables - execution time, extent of hotel search, and list view interaction (be measured using interaction logs) that are related to the *effort* requested to meet the user's goal (making a reservation) and the efficacy of the decision making process [15]. *Choice satisfaction* is the subjective judgment of quality/value for the user in relationship to the final choice, i.e., the reserved hotel. *Execution time* is the time taken for the user to explore the product offer, search for hotel information, and make a final decision. *Extent of hotel search* is the number of hotels that have been searched and for which detailed information has been acquired. *List view interactions* is the number of times the user *changed the ordering* of hotels in the list view of hotels matching some specified characteristics (e.g., stars, price, accommodation type). Ordering change is a measure of the *efficacy* of recommendations in situations where conversion rate, i.e., the percentage of recommended items that are actually purchased by users, cannot be assessed [16]. This typically happens when a system does not present a *separate* list of recommended items, but recommendations are rendered by sorting items in descending order of relevance as estimated by the recommender algorithm and the "top-N" items represent the "de facto" recommendation list.

The effects of recommendations are explored under *eight* different experimental conditions, defined by the combination of two manipulated variables*: hotel availability* and *recommendation algorithm*. Hotel availability can assumes two values: *high availability* - all hotels have rooms available in the dates specified by the user; *low availability* - some of the hotels that the user tries to reserve have no room available in the selected period. Concerning algorithms, we consider *one* non-personalized algorithm, *HighestRated*, and *three* personalized RSs representative of three different classes of algorithms: *PureSVD* (collaborative) *DirectContent* (content-based), and *Interleave* (hybrid).

HighestRated is the most common approach of online booking systems, which provides the user with non personalized lists of hotels matching some the chosen product attributes, and presents items in decreasing order of average user rating. *PureSVD* is a recently proposed latent factor algorithm based on conventional SVD [3]. Latent factor models, also informally known as Singular Value Decomposition (SVD) models, try to explain ratings by characterizing items and users with factors that are automatically inferred from a user's feedback. *DirectContent* is a simplified version of the content-based LSA algorithm described in [2]. Content-based algorithms recommend items whose content is similar to the content of items the user has positively rated or visited in the past. *Interleave* is a hybrid algorithm and generates a list of recommended hotels alternating the results from *PureSVD* and *DirectContent*. Interleave has been proposed in [1] with the name "mixed hybridization" and, although trivial in its formulation, has been shown to improve diversity of recommendations.

Table 1. Dataset statistics

Hotels	Total	3,164
	With reviews	2,884
Users (reviewers)	Total	209,704
	Venere	72,347
	TripAdvisor	137,357
Reviews ratings	Total	245,939
	Venere	80,562
	TripAdvisor	165,377
Hotel content	Unique features	481

3.2 Instrument

For the purpose of our study, we have developed PoliVenus, a web-based testing framework for the hotel booking field, which can be easily configured to facilitate the execution of controlled empirical studies in e-tourism services. PoliVenus implements the same layout as Venere.com online portal and simulates all of its functionality (with the exception of payment functions). The Polivenus framework is based on a modular architecture and can be easily customized to different datasets and types of recommendation algorithms. Venere.com provided us with a catalog of more than 3,000 hotels and 80,000 related users' reviews. We have enriched the dataset with additional reviews extracted from the TripAdvisor.com web site using a web crawling tool. Each accommodation is provided with a set of 481 features concerning, among the others: accommodation type (e.g., residence, hotel, hostel, B&B) and service level (number of stars), location (country, region, city, and city area), booking methods, average single-room price, amenities (e.g., cinema), and added values (e.g., in-room dining). User's reviews associated to each accommodation consist of a numeric rating and a free-text. Table 1 reports detailed statistics of the subset of data used in our experiments.

PoliVenus can operate in two different configurations. The *Baseline Configuration* corresponds to the existing Venere.com portal. Users can filter the hotel catalogue according to hotel characteristics (e.g., budget range, stars, accommodation type, city area) and retrieve non-personalized results. The *default* sorting criterion for the list of filtered hotels is "user average rating" but users can also sort the list by popularity, service level, or price. In the *Personalized Configuration***,** the only difference is the default sorting criterion: filtered hotels are ordered on the basis of the personalized recommendations.

The user profile required by the algorithms to provide recommendations contains implicit hotel that are dynamically calculated on the basis of the user's current interaction with the system (*implicit elicitation*) [5, 6, 8, 11, 12]. This choice is motivated by two reasons: (i) we want to support users who have no rating history or who are not interested in logging into the system; (ii) we want to explore a smooth, *conservative* integration of personalized recommendations in Venere.com: to enable explicit elicitation would require the introduction of an intrusive add-on. Lack of space prevents us to provide details of Polivenus implicit elicitation mechanism.

Finally, to control for *resources availability*, PoliVenus can be configured to simulate *two different situations*, achieved by shrinking or widening the set of hotels which have rooms available for the dates selected by the user. In the "high availability" configuration, all hotels always have rooms available. In the "low availability" configuration, the first 4 hotels for which the user checks for room availability are "forced" to result fully booked.

3.3 Participants

The total number of recruited subjects who completed the task and filled the questionnaire was *240*. They were aged between 20 and 40, had some familiarity with the use of the web and had never used Venere.com before the study. To encourage participation and to induce participants to play for real, we used a lottery incentive [17]. Participants had the chance of winning a prize, consisting of a discount coupon of the value of 150€, to be used at the hotel reserved using PoliVenus. All participants were not aware of the goal of the experiment and were given the following instructions when accessing Polivenus: "*Imagine that you are planning a vacation in Rome and are looking for an accommodation for three days during Christmas season; choose an hotel where you would like to stay and make a reservation; dates and accommodation characteristics (stars, room type, services, location) are at your discretion. After confirming the reservation (simulated), please complete the final questionnaire*".

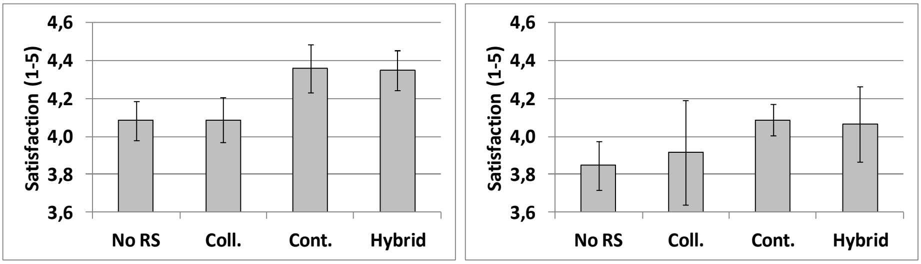
4 Results

N-way ANOVA indicates that algorithm and availability have a significant impact on almost all of the dependent variables ($p < 0.05$) with the exception of the number of explored hotels. Regardless the presence of recommendations and the availability of hotels, the *number of explored hotels* is between 5 and 8 for almost all of the users. In other words, customers wish to compare in details at least 5 and no more than 8 alternative choices before committing to a final decision.

We ran multiple pair-wise comparison post-hoc tests using Tukey's method on all of the remaining variables. The results are shown in Figures 2 and 3, where the mean is represented by a bar and the 95% confidence interval as a line.

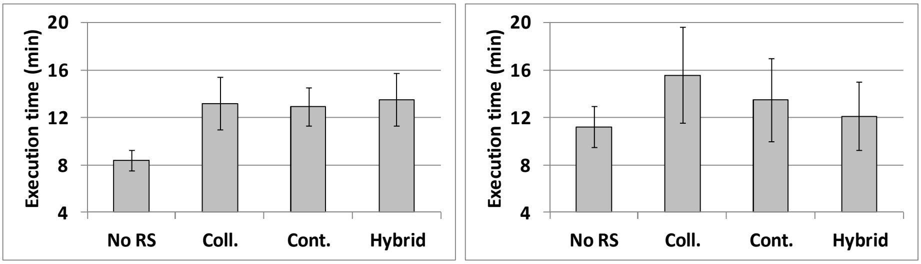
Figure 2 highlights the combined effects of *algorithm* and *resources availability* on choice *satisfaction*. It shows that the usage of *content and hybrid algorithms* in the online booking system *increases user's satisfaction*. Not surprisingly, in the *low-availability* scenario the user's satisfaction is on average lower (because of the difficulties in booking a hotel and the potential disappointment due to resources scarcity). Still, both *content and hybrid algorithms are able to increase user's satisfaction to the same level* of users not using recommendations in conditions of low-season/high availability. The improvement does not happen with the collaborative algorithm: users receiving collaborative recommendations have, on average, the same level of satisfaction of users who did not received personalized recommendations.

The increased user satisfaction when using RSs is not always correlated with the user effort. Figure 3 plots the execution time by algorithm in the 2 different conditions



(a) High Availability (b) Low Availability

Fig. 1. User satisfaction



(a) High Availability (b) Low Availability

Fig. 2. Task execution time

of resources availability. Users with low availability intuitively require a longer execution time with respect to users with high availability. Less intuitively, users receiving recommendations require more time than users not receiving recommendations..

Behavioral data emerging from users' interaction log files provides some interesting results. As mentioned in sect. 3.1., we have estimated the *efficacy* of the RS by measuring how many users changed the default sorting options (recommended items first) by setting other parameters (e.g., price, popularity, stars). The results show that only *37%* of the users with personalized recommendations changed the sorting of hotels, as compared to *54%* of the users in the baseline configuration, but no significant difference in the efficacy variable exists between different conditions of resources availability.

5 Discussion and Conclusions

Our findings show that *hypothesis 1 is confirmed*, with different algorithms inducing different effects on UX quality (in line with prior results, e.g., [1]). With respect to the baseline algorithm (no personalized recommendations) the hybrid and content algorithms *improve* choice satisfaction, while the collaborative one has no significant perceived effect. Our interpretation is that collaborative algorithms, although able to provide relevant recommendations, are not always able to provide novel recommendations – e.g., collaborative algorithms are biased toward "obvious" recommendations

[2]. In contrast, there is *no significant variation among personalized algorithms with respect to objective effort*: none of them statistically differ from the baseline in terms of execution time and extent of product search, i.e., number of explored hotel pages. Our results show a *mismatch between satisfaction* and *effort*: users exposed to hybrid and content recommendations perceived the decision activity process as more satisfying than those without personalized recommendations, although they spent the more time in the process. Our interpretation is that, thanks to the personalized recommendations, the presented hotels are more interesting for the users who spend more time in evaluating the different alternatives. This result is partially in line with previous work hypothesizing that, thanks to RSs, users spend less time in searching for items and more time in the more satisfactory activity of exploring information related to the choice process [12].

Hypothesis 2 is also confirmed. Hotel availability influences the effects of personalized recommendations. More precisely, the introduction of the content and hybrid personalized algorithms produces a significant *increment* of satisfaction in situations of *full* availability of resources. Still, personalized recommendations generated by both algorithms bring *no significant benefit* in situations of scarcity of resources. This result is not fully surprising: RSs perform well when information overload is the prevailing trait, but their benefits decrease when the reduction of the search space size, such as in the experimental condition characterized by low availability of resources. As resources availability is a dynamic characteristic of several domains outside the control of service providers, online operators are looking for ways to mitigate the negative effects induced by scarcity of resources in specific periods. The role of recommender systems in these situations deserves further investigation in order to design recommenders that can work well also in conditions of both abundance and scarcity of resources.

Overall, our findings extend our understanding of the potential of introducing recommendations to improve the UX quality with commercial e-tourism services, and differs from previous work in this domain for a number of aspects. We adopt a *conservative* approach that can promote the acceptance and adoption of recommenders in the e-tourism business. Differently from most previous works in this field where the evaluated RSs create new “ad hoc” user experiences, we assess the effect of recommendations on UX quality by *smoothly extending* a *commercial* online booking system (Venere.com) with personalized recommendations, without creating any major modifications to the “standard” interaction flow and overall user experience. In particular, we consider recommenders involving *implicit* elicitation, while most of existing studies in the e-tourism domain address recommenders with (more intrusive) explicit elicitation. In addition, we compare *three different algorithms* against the baseline scenario without recommendations and against each other. Previous works limit their analysis to a single recommendation algorithm evaluated against a non personalized baseline. Furthermore, it is worth noticing that an implicit assumption of most existing studies on Recommender System (RS) evaluation is that all items are always available, regardless the number of users who have “consumed” (bought, used) them. Still, in many domains, e.g., online services related to travel, clothing, or events, items are or involve physical resources that have constrained availability, i.e., the same product can be consumed by a limited number of users. To our knowledge our work is the first one that investigates this concept for recommender evaluation.

Finally, the *research design* of our empirical study is per se a strength of our work: for the number of variables measured in a single experiment – larger than most existing studies in e-tourism and other domains; for the sophisticated technological instrument used (the PoliVenus framework); for the vast size of the involved subjects (240); and for the lottery based incentive mechanisms adopted to motivate users and commit them to realistic and sound task execution.

References

1. Burke, R.: Hybrid web recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 377–408. Springer, Heidelberg (2007)
2. Cremonesi, P., Garzotto, F., Turrin, R.: Investigating the Persuasion Potential of Recommender Systems from a Quality Perspective: an Empirical Study. ACM Transactions on Interactive Intelligent Systems 2(2), 1–41 (2012)
3. Cremonesi, P., Koren, Y., Turrin, R.: Performance of recommendation algorithms on top-n recommendation tasks. In: Proc. RecSys 2010, pp. 39–46. ACM, New York (2010)
4. Delgado, J., Davidson, R.: Knowledge bases and user profiling in travel and hospitality recommender systems, 2002. In: Proc. ENTER 2002, pp. 1–16. Springer, Wien (2002)
5. Drenner, S., Sen, S., Terveen, L.: Crafting the initial user experience to achieve community goals. In: Proc. of RecSys 2008, pp. 187–194. ACM, New York (2008)
6. Golbandi, N., Koren, Y., Lempel, R.: On bootstrapping recommender systems. In: Proc. CIKM 2010, pp. 1805–1808. ACM, New York (2010)
7. Jannach, D., Zanker, M., Jessenitschnig, M., Seidler, O.: Developing a conversational travel advisor with advisor suite. In: Proc. ENTER 2007, pp. 43–52. Springer, Wien (2007)
8. Jones, N., Pu, P.: User technology adoption issues in recommender systems. In: Proc. NAEC 2007, pp. 379–394 (2007)
9. Levi, A., Mokryn, O., Diot, C., Taft, N.: Finding a needle in a haystack of reviews: cold start context-based hotel recommender system. In: Proc. RecSys 2012, pp. 115–122. ACM, New York (2012)
10. Xie, M., Lakshmanan, L.V.S., Wood, P.T.: Comprec-trip: A composite recommendation system for travel planning. In: Proc. ICDE, pp. 1352–1355. IEEE (2011)
11. Rashid, A.M., Karypis, G., Riedl, J.: Learning preferences of new users in recommender systems: an information theoretic approach. In: SIGKDD Explorer Newsletter, vol. 10, pp. 90–100 (2008)
12. Ricci, F., Missier, F.: Supporting travel decision making through personalized recommendation. In: Karat, C., Blom, J., Karat, J. (eds.) Designing Personalized User Experiences in eCommerce. HCI Interaction Series, vol. 5, pp. 231–251. Springer, Netherlands (2004)
13. Schumacher, M., Rey, J.-P.: Recommender systems for dynamic packaging of tourism services. In: Proc. ENTER 2011, pp. 13–23. Springer, Wien (2011)
14. Mahmood, T., Ricci, F., Venturini, A., Höpken, W.: Adaptive recommender systems for travel planning. In: Proc. ENTER 2008, pp. 1–11. Springer, Wien (2008)
15. Xiao, B., Benbasat, I.: E-commerce product recommendation agents: use, characteristics, and impact. Management Information Systems Quarterly 31(1), 137–209 (2007)
16. Zanker, M., Fuchs, M.W., Tuta, M.H., Müller, N.: Evaluating recommender systems in tourism - a case study from Austria. In: Proc. ENTER 2008, pp. 24–34. Springer, Wien (2008)
17. Porter, S.R., Whitcomb, M.E.: The Impact of Lottery Incentives on Survey Response Rates. Research in Higher Education 44(4), 389–407 (2003)