Received: 9 April 2021 | Accepted: 16 June 2021

DOI: 10.1002/aaai.12057

SPECIAL TOPIC ARTICLE

Building effective recommender systems for tourists

David Massimo | **Francesco Ricci**

Faculty of Computer Science, Free University of Bolzano-Bozen, Piazza Domenicani, 3, Bolzano 39100, Italy

Correspondence
David Massimo, Faculty of Computer Science, Free University of Bolzano-Bozen, Piazza Domenicani, 3, Bolzano 39100, Italy.
Email: davmassimo@unibz.it

Abstract

Recommender systems (RSs) are personalized information search and discovery applications helping users to identify and choose useful items and information. In this paper, we focus on the tourism application scenario and its specific requirements. We discuss a novel RS approach that copes with the specific application constraints of the domain and produces recommendations that better match the true needs of tourists. We illustrate the proposed next POI recommendation approach in a case study and we compare it with a state-of-the-art nearest neighbor-based next item RS. With the analysis of this case study, we aim at illustrating the specific features of the compared approaches also with the goal to raise the discussion on RSs validation methods, with a particular attention to tourism applications. We finally discuss some significant limitations of current evaluation approaches that must be addressed in future studies.

INTRODUCTION

Recommender systems (RSs) are personalized information search and discovery applications helping users to identify and choose useful items and information (Jannach et al. 2016; Ricci, Rokach, and Shapira 2015). RSs are nowadays very popular in streaming platforms (e.g., Netflix and Spotify), and eCommerce websites (Amazon). In this paper, we focus on the tourism application scenario and its specific requirements (Staab et al. 2002; Werthner et al. 2015; Werthner and Ricci 2004; Rabanser and Ricci 2005). In particular, we concentrate on a typical tourist's information search task: finding novel and compelling points of interest (POIs) to visit, and eventually extend an already initiated or planned visit itinerary to a destination, for example, a city (Braunhofer, Elahi, and Ricci 2015).

Tourists often face this sequential decision-making problem, while planning their visit to a destination or when at the destination continuing an already initiated trajectory of visited POIs (Staab et al. 2002). We note that in tourism, quite differently from the above mentioned applications (movies, music, ecommerce), there is no clearly defined catalog of recommendable items. In fact, what is recognized as a point of interest for some tourists may not be seen as a tourism target for others. For instance, while an Italian tourist may be recommended to visit a small town in a nearby region of her residency, this will not be recognized as a compelling target for a Japanese tourist, who will instead consider the whole Italy as a possible destination, maybe in alternative to France (Hwang, Gretzel, and Fesenmaier 2002). So, it might be critical for an RS to help any type of tourist, at the decision/choice point, to find POIs that can be recognized as interesting targets, based on the tourist's culture, knowledge, and personality (Gretzel et al. 2004). Moreover, POIs are worth to be visited because they generate experiences, and the quality of these experiences is hard to be fully estimated beforehand, at planning time. Hence, first, the RS should be able to "persuade" the tourist of the goodness of its recommendations, since, as we said, we cannot expect that such a quality can be fully assessed on the base of the provided information, especially if the POI is not already known by the tourist (Gretzel and Fesenmaier 2006). Second, the recommended POIs, when they are actually visited, must satisfy the tourist, give a "reward," and create a memorable experience (Gretzel et al. 2015).

In this article, we discuss the difficulties that these two goals create to the design of an RS in the tourism domain. We note that RS research has already tackled the problem of next item recommendation (Hariri, Mobasher, and Burke 2012; Hashemi and Kamps 2017; Jannach and Lerche 2017; Ludewig and Jannach 2018; Quadrana, Cremonesi, and Jannach 2018; Shani, Heckerman, and Brafman 2005; Zhang, Chow, and Li 2014; Moling, Baltrunas, and Ricci 2012), but the state-of-the-art solutions, while being generally applicable to a wide range of application domains, have failed to address the specific needs of tourists. In fact, major players of the online tourism market, such as Booking.com or Tripadvisor.com, have not yet adopted these sophisticated solutions and nowadays they offer a recommendation functionality that is not personalized: it is either based on the average opinion of the users or on the items' popularity. However, for business motivations, they do consider, when generating recommendations for tourists, constraints, and goals imposed by the suppliers side (Abdollahpouri et al. 2020). Hence, ultimately, this RS application domain has not grown with the same fast pace that other domains have seen.

One of the causes of the slow development of RS applications in tourism is surely related to the difficulty to acquire information about the true user behavior, that is, the sequence of experiences that travelers perform. So, while their online information search activity is easy to be tracked (Choe, Fesenmaier, and Vogt 2017), their true experiences, that is, the POIs they visit, are only known indirectly, in the form of selected reviews, which only specific travelers (bloggers) usually provide (Marchiori, Cantoni, and Fesenmaier 2013; Zhang and Fesenmaier 2018). This is substantially different from other domains; in Netflix, for instance, the users' watching behavior is easily tracked and users can express their "like" for a movie by just one click (Auksorncherdchoo and Sukstrienwong 2018; Krishnamurthy and Wills 2009; Castelluccia, De Cristofaro, and Perito 2010). So, RSs in the tourism domain suffer from a continuous state of "coldness": they do not have enough users' preference data to generate effective and personalized recommendations (Elahi et al. 2018). From a more technical point of view, we argue that the unsatisfactory results of current tourism RSs reside also on the usage of standard recommendation models, which are optimized to precisely predict the observed tourist behavior, and therefore, they offer suggestions that match, as precisely as possible, what the single tourist is observed to do. But, tourists are rarely experts, especially when visiting new destinations, and their behavior is typically exploratory. So, their, even scarce, observed behavior cannot be directly used as model training data or ground truth for measuring the goodness of the recommendation model. In fact, an important goal of an RS is to support "knowledge discovery," and this is particularly true in the tourism domain: recommendations should indicate novel items that the user is not aware of, but will like (Werthner et al. 2015).

In order to address these requirements and issues, we discuss in this paper a novel RS approach that copes with the specific application constraints of the domain and produces recommendations that better match the true needs of the tourists (Massimo and Ricci 2018a; 2021a). This recommendation approach is implemented in three steps. First, clusters of tourists with a similar observed behavior are created. We note that tourists are normally classified in standard prototypical types (Yiannakis and Gibson 1992). In our approach, a cluster corresponds to a type of tourists, but these clusters are not apriori defined, as in the cited tourism literature. Conversely, clusters are computed by running a clustering algorithm directly on the (scarce) observed behavior data, which consists of the trajectories of successive POI visits in a city that are performed by a collection of observed tourists. Moreover, the obtained clusters of tourists depend on a specific representation of the visit trajectories, which we define, and it comprises features related to the content of the visited POIs (e.g., the historical period of the POI), and the visit context (e.g., the part of the day when the POI was visited).

In a second step, for each identified cluster of tourists, a behavior model of the sequential decision-making process of the tourist is built. The behavioral model determines which POIs a tourist in a cluster will likely choose next, that is, after having chosen other POIs, and how much "reward" the tourist is estimated to obtain by a POI visit. The behavioral model is learnt via Inverse Reinforcement Learning (IRL) (Abbeel and Ng 2004; Babes-Vroman et al. 2011) and it is only based on the observed behavior, that is, tourists are not supposed to give any explicit feedback on their past POI visit experiences. However, the learning procedure implicitly assumes that tourists aim at maximizing an unknown reward function that is actually estimated by the learning algorithm. We note that, by building a behavioral model for each cluster, the model, while not being individually specific, as it is common in RSs, it is not even completely general (one single model for all) as in the above-mentioned industry solutions. We note as well that the main rationale of clustering tourists and building a behavioral model for each cluster is the above-mentioned "coldness" of the available data: rarely there is enough, previously observed, individual behavior data that suffice to build a fully personalized and individualized model.

In the third step of the proposed recommendation approach, the learnt behavioral models, one for each cluster of tourists, are leveraged for building next POI recommendations that have the characteristics of the POIs typically visited by the tourists in the same cluster of the target tourist. We stress that, differently from more traditional approaches used in session-based RSs, which tend to recommend the items more likely to be consumed by the target user, the proposed approach tries to identify the items (POIs) that will be perceived as having, and will actually give, a larger "reward" to the tourist. The reward is a system proxy for the satisfaction of the experience of the POI. This is achieved by implementing alternative heuristics, aimed at balancing these two, possibly conflicting goals: identify POIs that the tourist can recognize as relevant, before experiencing them, but also that will produce satisfying experiences when actually visited. These alternative heuristics are called "recommendation strategies" and prioritize specific characteristics of the generated recommendations, hence they are not limited to maximize recommendation accuracy, as in more traditional approaches. A key ingredient of the proposed recommendation strategies is instead the maximization of the estimated reward that a tourist can obtain from the recommended experiences (POIs), that is, we try to prioritize the quality of the experience of a recommended POI, rather than the accuracy to match the observed behavior. However, the probability that the tourist will recognize, before the visit, that the recommended POI matches her preferences, is an important element to consider, and we, therefore, offer also an hybrid solution aimed at attaining this goal as well.

In this article, we illustrate the proposed next POI recommendation approach in a case study and we compare it with a state-of-the-art nearest neighbor-based next item RS. With the analysis of this case study, we aim at illustrating the specific features of the compared approaches also with the goal to raise the discussion on RSs validation methods, with a particular attention to tourism applications. In particular, we illustrate to what extent results obtained in an offline evaluation study are confirmed in a user study. But, we also discuss some significant limitations of both evaluation approaches that must be addressed in future studies.

This paper is organized as follows. Section "Recommender Systems for Tourism" presents an overview of Tourism RSs developed in industry and academia, summarizing open challenges. Section "Next POI Recommendation" introduces our next-POI recommendation approach and Section "Evaluating Tourist RSs" discusses important issues arising in the evaluation of tourism RSs. Section "Offline and Online Next POI Recommendations" illustrates the experimental results we collected by means of offline and online evaluation studies. Finally, we discuss challenges and future research directions in Section "Open Challenges for Tourism Recommender Systems".

RECOMMENDER SYSTEMS FOR TOURISM

Even though tourism applications of RSs have attracted less attention, compared to, for instance, mainstream music and movie applications, the next-POI recommendation problem received some specific recognition (Adamczak et al. 2020). In this application problem, clearly, the sequential nature of the items consumption plays a relevant role (Dellaert, Ettema, and Lindh 1998). Moreover, next-POI recommendation solutions have tried to address an important challenge of the domain, which is the lack of individual data about tourists' POI visits. In fact, tourists do have privacy concerns (Poikela et al. 2015; Perentis, Vescovi, and Lepri 2015) and many tourists are reluctant to share their location with companies. As a consequence of that, for each single tourist, the set of opinions about the visited POIs., for example, booked hotels or attractions, could be very small and even empty (Bin et al. 2019). To partially circumvent this problem, many studies dealing with next-POI recommendation use data derived from social networks (Baraglia et al. 2013; Oppokhonov et al. 2017; Palumbo, Rizzo, and Baralis 2017; Sánchez and Bellogín 2020). It is worth noting that social network users do not represent the full spectrum of tourists, and the core problem of acquiring unbiased and representative behavioral data remains. However, for this population of social networks users, by leveraging check-in data or geo-tagged media content uploaded by users on web platforms, it is possible to reconstruct their (partial) POI visit activities, for example, during a visit to a city (Silva et al. 2019). Hence, nowadays industrial players with their social network platforms, like Google[1], Foursquare[2], and Facebook[3] are in a much better position for implementing next POI recommendation solutions, even compared with players of the tourism market.

In general, we must observe that many state of the art solutions, tackle the next-POI recommendation problem without appropriately considering the typology of the POI, in any tourism-related classification of the POIs, and without considering the context of the visit, for example, with whom the tourist visited the POI (Oppokhonov et al. 2017; Huang and Gartner 2014; Wang et al. 2018). Hence, state-of-the-art solutions do not try to "understand" what conditions and features make a POI worth to be visited by a specific tourist. These solutions reuse trajectory data mining approaches (Zheng 2015) where it is assumed that only spatio-temporal aspects define the similarity of POI-visit trajectories performed by tourists. Understanding the motivations that steer tourists to make specific choices is left apart. A common pitfall of these solutions can be found, for instance, in Torrijos, Bellogín, and Sánchez (2020) where, in order to identify tourists interested to a target POI, important information related to the POI visits, for example, the weather conditions at visit time and the type of visited POI, is neglected, while more easily measurable properties, borrowed from trajectory mining techniques, such as, the distance of the points coordinates of the shape described by the POI visit trajectories, are considered. In general, state-of-the-art solutions, such as nearest neighbor-based RSs, leverage the similarity of POI-visit trajectories, and generate next-POI recommendations by mining frequent patterns in similar trajectories (Hariri, Mobasher, and Burke 2012; Jannach and Lerche 2017; Sánchez and Bellogín 2020).

Another line of research of the state -of-the-art relates to identifying distinguished typologies of tourists by clustering them on the base of features derived from their traits or behavior (Palumbo, Rizzo, and Baralis 2017; Yao et al. 2017). In Palumbo, Rizzo, and Baralis (2017), clusters of tourists are identified by leveraging demographic information acquired from social media platforms. The reconstructed POI-visit trajectories are enriched with features describing the category of each POI. The authors try to identify POI categories relevant for a target tourist but the final step of producing recommendations is not addressed. In another solution based on check-in data (Yao et al. 2017), the authors propose to use a deep neural network to extract behavior features that capture space- and time-invariant characteristics of trajectories collected from social networks.

A more sophisticated clustering approach for next-POI recommendation is presented in McKenzie and Janowicz (2014). Given the user's preferences over places derived from a location-based social network, the model finds similar individuals based on properties of the preferred items and recommends places based on related preferences of these similar individuals. Clustering is applied to users' check-in data to identify individual's daily activities. For each cluster, a POI that best represents each cluster is identified as "typical activity." By considering week-day and weekend activities a user is characterized with specific activities to be performed on those specific days. Recommendations are then generated by user-to-user collaborative filtering.

Interestingly, and somewhat related to the topic of clustering tourists, GroupTourRec (Lim et al. 2016) is a system that includes the functionality to form groups of homogeneous people, by identifying POIs appropriate to each group and assigning a guide to each group. Hence, here clustering is used for forming groups of users to travel together; users are independent travelers and are clustered together according to their behavior. The suggestions of POIs to visit are generated by solving an orienteering problem rather than using predictive techniques.

The sequential nature of the item consumption in tourism plays a relevant role in the itinerary recommendation solution proposed in Herzog, Laß, and Wörndl (2018) and Rani, Kholidah, and Huda (2018). Here the supported task is to advise the tourist while planning the visit activity. In Rani, Kholidah, and Huda (2018), the authors aim at finding optimal itinerary recommendation in terms of distance and travel time. They start from the assumption that the user has already identified the POIs she wants to visit and the number of days she will spend in the region. In this situation, a clustering algorithm distributes the POIs in clusters that correspond to the available days. Then a traveling salesman problem algorithm determines the actual visit order. We observe that this solution assumes that tourists are already knowledgeable about a place and they already know what they want to visit.

In Wörndl, Hefele, and Herzog (2017), the authors present a travel RS that recommends a list of POIs that the tourist does not necessarily know in advance. Given a start and an endpoint, an itinerary is built by using a custom shortest path algorithm that optimizes user preferences over POI categories and time constraints objectives. The estimated suitability of the POIs for the itinerary is based on the tourist stated preferences and the POIs reputation, which is derived from the ratings and the number of votes, collected from a social network.

As we already mentioned at the beginning of the section, the surveyed approaches tend to ignore an important dimension of the tourist POI experience: the context of the visit. Contextual factors, such as, visiting a POI on a "sunny day" with the "family" during the "spring holidays," influence not only the tourists' choices but also their memories (Lamsfus et al. 2014; Matzarakis 2006). In Hong et al. (2019), the authors investigate how the cultural dimension influences the acceptance of the recommendation. In fact, tourist's culture is intertwined with the visit context (Savard and Mizoguchi 2019) and they jointly affect the users' preferences and experiences. The authors propose to use clustering and dimensionality reduction techniques to identify cross-cultural factors that are leveraged in the prediction of POI recommendations. More in general, previous literature on context modeling in tourism has dealt with the temporal context (Sánchez and Bellogín 2020; Zhao et al. 2019) but only few authors considered also the categories of the POIs when dealing with contextual effects (Li et al. 2019; 2020).

We point out that most of the itinerary recommendation approaches that have been proposed in the past tend to reinforce the consumption of POIs that are popular and often already known by the users. Moreover, no past approach have tried to model and leverage the "reward" that tourists obtain by visiting the recommended POIs. While precisely defining such a reward is difficult, our approach tries to capture such as hidden reward by making the assumption that overall tourists make and report visits that are rewarding for them, and only erroneously they visit POIs that have not this property. Hence, by leveraging a specific learning approach, namely IRL, aimed at learning such an hidden reward that motivates the decision maker, we try to generate recommendations that have the characteristics of the items preferred by the tourist.

Moreover, very little attention has been given to the proper, user-based, evaluation of next-POI RSs, which is clearly due to the planning and management costs inherent to these evaluation methods (Gunawardana and Shani 2015). User studies in the travel domain can be found in Braunhofer, Elahi, and Ricci (2014), Nguyen and Ricci (2018), Herzog and Wörndl (2019), but the focus of these works was not on next-POI recommendations.

NEXT POI RECOMMENDATION

We focus on a scenario where the RS is used to assist tourists in sequential decision-making, that is, in facing the next-POI recommendation problem: looking for an additional POI to visit after having visited some other POIs (Massimo and Ricci 2018a; 2021a). We present here the three-step approach that we have sketched in the introduction. In the rest of this section, we assume that there is a data set of observed visit trajectories of a collection of tourists that is used to learn the behavior model. Each visit trajectory is composed by a sequence of POI visits. Each POI visit is described by a visited POI and a set of contextual conditions observed at the visit time, for example, the weather conditions.

Clustering tourists' visit trajectories

At first, we cluster tourists' visit trajectories, into groups of trajectories related to a common topic. These clusters are extracted directly from the analysis of tourists' behavior, after having identified a set of features that can be used to describe the content of a POI and the context of the visit to the POI.

In order to identify such clusters, we represent each observed tourist's POI-visit trajectory in a document-like format, where the terms of a trajectory document are the content and context features describing the POI visits contained in the trajectory. Hence, this representation of the visit trajectory captures different dimensions that characterize the traveler experience: the context of the POI visits, for example, the part of the day and weather when the visit occurred; and what is visited, for example, POI category and historic period (Massimo and Ricci 2020). By doing so, we abstract from the visit order and the identity of the specific visited POIs, and we focus on what may interest the tourist (content features) and in which conditions (context features). It is important to note that in order to succeed in the identification of clusters that can really correspond to meaningful tourist typologies, it is fundamental to leverage the "right" set of descriptive features to represent the visit to a POI. This is an activity that we have performed by leveraging specific domain knowledge (see Massimo and Ricci (2020) for more details).

Then, to form the required clusters of POI-visit trajectories, we used a topic model approach based on non-negative matrix factorization (Massimo and Ricci 2018a). This method allows us to identify a small number of hidden topics in the document-trajectory collection. A topic is described by a collection of terms: those more related to the topic. For instance, by using the data set of visit trajectories described later in the paper we have identified five topics, and one of these (hidden) topics is associated to trajectories that are characterized by the terms: morning, cold, square, palace, 15th century (the full description of these topics can be found in Massimo and Ricci (2020)). Hence, in the observed set of visits, a group of tourists seems to be interested in visiting palaces and squares of the 15th century in cold mornings. The clusters of visit trajectories are then defined by grouping together the trajectories more strongly associated to the identified topics, that is, one topic defines one cluster, and a trajectory can belong to more than one cluster.

The main benefit of this approach resides in the fact that we can identify groups of related visit trajectories, even when dealing with small sized datasets of observed tourists' choices. Besides, even if we had at our disposal many POI-visit trajectories for each tourist, they will still reveal a restricted set of preferences, which are biased by the tourist limited knowledge of the destination. So, these trajectories may also contain suboptimal choices. Clustering is a first step to overcome these problems: suboptimal choices made by one tourist may be compensated by better choices of other tourists in the same cluster (by assuming that not all tourists make the same errors). Learning a behavioral model for each cluster is the second step to extract from the observed visit trajectories a useful model of the true preferences of the tourist.

Tourists behavior learning

We want to learn the user behavior models that characterize the tourists' typologies captured by the generated clusters. This means that we want to estimate the unknown reward that tourists in a cluster seem to optimize in their behavior, that is, by performing the observed POI visits in that order. The proposed approach does not assume that tourists are completely aware of what makes a POI visit rewarding, but tries to extract the rationale just by observing the characteristics and the context of the visited POIs. For instance, the proposed approach seeks to estimate the reward that a tourist who visits, for example, the *Colosseum* in Rome, obtains by visiting as next POI, *Fontana di Trevi* and thereafter *Villa Borghese*. Moreover, the proposed approach tries to determine which next POIs, after *Villa Borghese* the tourist should, step by step, choose.

We use a standard Markov Decision Problem (MDP) model to frame the tourist's POI-visit decision making task (Abbeel and Ng 2004). A MDP is a tuple (S, A, T, r, γ). S is a finite set of states, and, in our case, a state represents a visit to a POI under specific contextual conditions, for example, visiting the *Colosseum* during a sunny day. A is a finite set of actions: moving to one of the available POIs. T is a finite set of probabilities: $T(s'|s, a)$ is the observed probability to make a transition from state s to s' when action a is performed. These probabilities account for the possibility that when the tourist decides to make the action to visit a next POI, for example, *Fontana di Trevi*, contextual conditions, such as the weather, may change in an unexpected way, hence the reached state is not univocally determined by the performed visit action. The function $r : S \rightarrow \mathbb{R}$ models the reward the decision maker obtains from acting in a certain way, that is, by being in a state, that is, by visiting a specific POI in a particular context. This function is *unknown* in our application scenario, because we do not assume that the tourist gives an explicit feedback (e.g., a rating or a like), and therefore, the reward function must be learnt by using only the observed POI visits. Finally, $\gamma \in [0, 1]$ is a parameter measuring how much rewards from visits performed later in a visit trajectory are discounted with respect to the immediate ones: a reward received k visits after the current visit is worth only γ^{k-1} times what is would be worth if it were received immediately. The lower the value of γ the more myopic is the decision maker, that is, he is just trying to optimize the immediate reward and less the reward that can be obtained by the subsequent visits.

Given the MDP associated to a cluster, which models the common decision problem faced by the tourists whose visit trajectories are contained in the cluster, the behavioral model for this cluster is a decision policy $\pi^* : S \rightarrow A$ that maximizes the cumulative reward that the decision maker obtains by acting according to π^* (optimal policy). The value of taking a specific action a in state s under a policy π, is indicated with $Q_\pi(s, a)$, and it is the (expected) discounted cumulative reward obtained by making the next POI visit a in state s and then continuing to make successive visits by following the policy π. The optimal policy π^* dictates to the decision maker in state s to perform the action that maximizes the value function Q_{π^*}.

Since, as we said, the reward function is unknown, the optimal policy π^*, that is, the optimal behavior of the decision maker, cannot be determined with standard reinforcement learning algorithms (Sutton and Barto 1998). Conversely, in this case, IRL can be used (Abbeel and Ng 2004; Ermon et al. 2015). IRL enables to identify both the reward function, which the decision maker seems to optimize, and the optimal policy for that reward function. In other words, by using IRL one can estimate how tourists in a clusters behave, what reward seem to obtain from visits to different POIs, and, in any possible state, the next best visit that they should make.

The reward function r and the associated optimal action selection policy π^* that are computed by IRL strictly depend on the observed (clustered) POI-visit trajectories but also on the selected state feature function $\phi : S \rightarrow \mathbb{R}^n$ that assigns to each state a vector of feature values (n is the number of features). We also observe that when IRL is used, an apriori defined constraint on the form of the reward function must be imposed, so that the problem can actually be solved. Hence, as in Abbeel and Ng (2004), we assume that r is a linear function, $r(s) = \theta^T \phi(s)$, of the state s feature vector $\phi(s)$. The vector of parameters θ model the unknown decision maker's preference for the state features. Hence, we make a simplifying assumption on the structure of the tourists preferences: the reward grows when the visit to the POI is described by the features (content and context) that the user prefers.

Morever, by using IRL we implicitly assume that a tourist is a rational decision-maker, seeking to optimize a (unknown) reward determined by the visited POIs. Such an agent is typically referred to as an "expert," because the observed behavior is assumed to be dictated by knowledge. However, it is difficult to believe that tourists are true "experts," that is, the observed behavior surely contains suboptimal choices: for instance, tourists may repeatedly visit a few popular POIs. Learning the user behavior from a cluster of POI-visit trajectories of tourists is actually aimed to tame the problems related the presence of suboptimal choices: suboptimal choices, if not correlated, will not jeopardize the learned behavior model.

Recommendation strategies

Having learned a behavioral model for each cluster of tourists, we propose to use it to suggest next-POI visits to the tourists in that cluster (Massimo and Ricci 2018a; 2020; 2021a). We recall here the important assumptions on a suitable next-POI RS that we discussed in the introduction. We do not want to generate recommendations equal to those actually consumed by the target tourist; the recommended POIs must be perceived as valuable and must offer rewarding experiences to the tourist. In order to accomplish this goal, we consider alternative heuristics, aimed at balancing these two, possibly conflicting goals. These heuristics are called "recommendation strategies" and prioritize specific characteristics of the generated recommendations, hence, should not be limited to maximize recommendation precision, as in traditional approaches. Several strategies may be implemented and we hope to see further developments in this direction. In this paper, we exemplify this analysis by considering two of them.

The first one is called `Q-BASE` and it directly exploits the learnt user behavior model of the cluster the tourist belongs to. `Q-BASE` recommends as next POI action visit, the optimal one, according to the optimal decision policy learned in the tourist's cluster. The optimal visit action has the largest Q value in the current user state. Hence, if the tourist will make this choice and will continue to make successive POI visits by choosing the actions with the largest Q value, which are recommended by `Q-BASE`, then the obtained cumulative reward will be maximized. `Q-BASE` is therefore a recommendation strategy that not only tries to suggest the most satisfying immediate next POI visit, but also the visits that the tourist will be able to make after that immediate next. Moreover, since the reward is estimated on the base of the POI characteristics and visit contextual conditions, `Q-BASE` can even recommend novel POIs, not yet visited by tourists, provided that they have the characteristics of the POIs visited by the tourists in the same cluster, and are visited in the contextual condition typically preferred by the tourist in the same cluster.

The second strategy acknowledges that tourists often follow trends, being influenced by POIs popularity and fashion (García 2004), which are easily communicated by websites like TripAdvisor. While these aspects may not influence the experience that the tourist will have by visiting a POI, even though, visiting popular POIs may be considered a target for some tourists, they will certainly influence the decisions of the tourist. It is well known that "familiarity breeds liking." For instance, in experiments made with music, it has been found that people do not select what they think they like but what are more familiar with Madison and Schiölde (2017). Tourists are not different, and they often visit what are considered to be the top attractions and frequently mentioned POIs (Moutinho 1987; Swarbrooke and Horner 2006). Hence, in the second recommendation strategy, which is called `Q-POP PUSH`, we take that aspect into account and we generate recommendations by averaging two criteria: the first is the cumulative reward that can be obtained by making the next-POI visit, as for `Q-BASE`, and the second is the popularity of the POIs, which is estimated on the available visit trajectories.

EVALUATING TOURIST RSs

The effectiveness of RSs has been assessed via offline analysis, user studies and online testing (Gunawardana and Shani 2015). An offline analysis offers a quick and inexpensive tool for evaluating the RS performance by using existing datasets of user–item interactions, and computing predefined metrics, which are mostly estimating the precision of the RS (Karypis 2001; Cremonesi, Koren, and Turrin 2010). Precision relates to the ability of a recommendation approach to predict either the observed user choices (e.g., the visited POIs) or the recorded evaluations for items (e.g., ratings for POIs). The prediction of user choices is typically assessed by computing information retrieval metrics: precision and recall. Precision is computed as the fraction of the relevant items among the recommendations, whereas recall is the fraction of the relevant items that are recommended. The precision of the predicted item evaluations is instead measured by regression type error metrics, such as mean absolute error (MAE) or root mean square error (RMSE) (Gunawardana and Shani 2015; Herlocker et al. 2004; Powers 2008).

Researchers have pointed out that optimizing an RS for precision can even negatively affect the overall user experience (McNee, Riedl, and Konstan 2006). In fact, striving for precision can lead to the recommendations of items that are often uninteresting, as they too closely match what the user already typically consumes and knows. Moreover, these recommendations self-reinforce the consumption of blockbuster items (Zhou et al. 2010; Ball 2010; McNee, Riedl, and Konstan 2006; Vargas and Castells 2011). Hence, it is argued that a proper assessment of an RS should be based on a broader set of indicators of recommendation quality (Ball 2010; McNee, Riedl, and Konstan 2006; Gunawardana and Shani 2015), and the indicators must be properly selected in relation to the application goal of the RS. In Vargas and Castells (2011; 2014), the authors have proposed specific metrics that complements precision in order to measure the "novelty" of the recommendations. Furthermore, in Kumar et al. (2017), an evaluation metric is proposed that assesses how similar the properties of the suggested items are to those in the test set. This enables to understand if the RS can suggest items different from those previously consumed by the user but still similar to them.

Hence, as the literature suggests, also in the evaluation of tourism RSs, it is fundamental to consider their practical usage and how tourists consume products, that is, POIs in our scenario. In fact, while searching for a POI to visit or a hotel to book, tourists rarely seek suggestions for items that they can autonomously find. Conversely, they are looking for relevant and rewarding discoveries. Specifically, they expect to find items that they do not know yet, hence they are novel, but also aligned with their preferences/needs, and capable to generate memorable experiences and satisfaction. Therefore, precision cannot be the sole metric used to assess the quality of a tourist RS. The novelty of the recommendations and the estimated reward, the user can obtain by consuming them, are important qualities of the recommendations that have to be assessed.

Evaluating the novelty of recommendations in offline studies is hard and is only accomplished by measuring other properties of the recommendations that are associated with novelty, for example, the unpopularity: an item that is not popular in the observed choices of users, should also be novel when recommended (Gunawardana and Shani 2015). Moreover, a major drawback of offline studies lies in the fact that one must make the restrictive assumption that only the evaluations (or the choices) present in the test set can be used to judge the quality of the recommendations. Hence, the user's previously observed behavior is considered as ground truth and novel behaviors, which could be proposed by the recommendations, cannot be judged. Clearly, preferences for novel items that an RS could suggest to the user, are not present in that set, that is, items not yet "evaluated" by the user are all considered as bad recommendations and decrease the estimated system's precision. Moreover, the interaction context cannot be considered in an offline study (Adomavicius et al. 2011; Braunhofer and Ricci 2017). This means that, in offline evaluations, implicitly it is assumed that when the user evaluates an item, already evaluated in another context, the same evaluation would be given; which is rare. However, offline analysis of RSs performance allows comparing different RS variants at once, on a broad set of metrics, and by utilizing various datasets. These properties makes offline evaluations powerful and indispensable tools.

User and online studies do not have that flexibility: only a few alternative RSs can be compared, by letting the users to try them, either in a controlled situation or in the wild (Bellogín and Said 2018; Gunawardana and Shani 2015; Knijnenburg et al. 2012; Pu, Chen, and Hu 2011). Conversely, in user and online studies, the collected user/system interactions can be analzsed and the users' reactions even to "novel" recommendations can be observed. The situation hence is very different from a simulated offline evaluation where there is a single and static reference set of assumed good recommendations, which are the items in the user's test set (Gunawardana and Shani 2015). In user and online studies, the tester is able to analyze the recommended items, and to decide which one is relevant or not, by using her specific idea of what a relevant recommendation is. In other words, online there are no stored "preferences" or choices that the RS must "predict," as offline, but preferences and choices are "constructed" while the user is interacting with the RS (Bettman, Luce, and Payne 1998). These studies are more expensive in terms of invested time and planning. Users have to be found, instructed and posed in an ecologically valid setting: the usage situation, the task to be performed, and the interaction should closely match the real setting in which the user actually interacts with the RS. Hence, such studies must be planned with care, and because of their high cost, they are often avoided in academic research.

In the next section, we will exemplify and further specify these general problems in the comparative analysis of the proposed next POI recommendation approach that we have illustrated in the previous section.

OFFLINE AND ONLINE EVALUATION OF NEXT POI RECOMMENDATIONS

In our experiments, we have used a dataset of POI-visit trajectories derived from tourists' activities on a social network. Specifically, individual POI-visit trajectories in the city of Florence (Italy) are reconstructed from geo-tagged pictures uploaded on the Flickr[4] photo sharing platform (Muntean et al. 2015) and have been augmented with the context of the visit, for example, weather summary or part of the day, and POI features, for example, POI-categories and reputation (Massimo and Ricci 2018b; 2021b). We mentioned in Section 2 that social network users do not represent the full spectrum of tourists, hence the results of the presented experiments should be considered as more indicative of the effect of RS on this particular segment of users. The total number of POI-visit trajectories that we have considered is 1663. A trajectory contains on average 11.7 POI-visits, and the number of unique POIs is 532. We note that the trajectories/users ratio is 1.43. In practice, the majority of the users in this dataset have just one visit trajectory. This makes clearly almost impossible to learn a user-specific user behavior, that is, a distinct reward function and optimal policy for each user, and it justifies the proposed clustering-based approach.

Offline experiment

In Table 1, we show the results of an offline experiment. Here we compare the performance of the two IRL-based recommendation strategies `Q-BASE` and `Q-POP PUSH` (see Section 1) with a popular next item RS baseline, `SKNN`. `SKNN` is a nearest neighbor next-item RS, not specifically tailored to the considered tourism application. `SKNN`, given the POI-visit trajectory of a target user, seeks other users who performed similar visits and recommends the most frequent next POI-visit performed by these similar users (Ludewig and Jannach 2018). Hence, while `SKNN`

TABLE 1 Offline analysis of next-POI recommendation performance (Top-1)

Model	Model description	Reward	Precision	Novelty
`Q-BASE`	Maximal reward	0.073	0.043	0.061
`Q-POP PUSH`	Balance reward and popularity	−0.002	0.099	0.000
`SKNN`	Popular among similar visitors	−0.007	0.109	0.000

aims to predict the next POI the tourist will visit, `Q-BASE`, as discussed in Section 1, tries to identify the POIs that have the characteristics usually liked by similar users (in the same cluster) and give to the tourist the largest (cumulative) reward.

The RS performance metrics shown in Table 1 are meant to address the requirements of a tourist next-POI RS and are: *Reward*, *Precision*, and *Novelty*.

Reward is the average increase of the system estimated reward a tourist obtains if she acts as recommended rather than as she did (test set). We note that the reward function is estimated on the base of the observed tourist behavior, as for the novelty of the users' visited POIs. Importantly, a tourist can obtain even a larger reward by deviating from the observed behavior: hence a less precise recommendation can give a larger reward. This reflects the fact that the observed behavior is not necessarily optimal, and the proposed IRL-based recommendation strategies can detect that, and suggest even better options than those observed in the data. Hence, the reward metric measures a recommendation quality quite different from precision.

Precision is the proportion of the recommendations found in the test set. Finally, *Novelty* is the percentage of the recommendations that covers the less popular items in the data (see Massimo and Ricci (2018a; 2020)). We have to resort to a proxy for measuring novelty since it is impossible in an offline study to appropriately measure the true novelty of a recommendation (Gunawardana and Shani 2015). True novelty of the recommendations will be instead measured in the user study discussed after.

It is clear, by observing the results in Table 1, that `Q-BASE` recommends next POI-visits that have higher reward and are also more novel, at the cost of a lower precision. Interestingly, we note that `Q-POP PUSH`, by trying to optimie both the reward and the popularity of the recommended next-POIs loses the capability of `Q-BASE` to suggest high reward POIs, and it performs substantially equal to `SKNN`. It is worth noting, not shown here for lack of space, that with a better tuning of the weighted combination of the reward and popularity criteria, `Q-POP PUSH` can achieve the precision performance of `SKNN` while offering much of the reward obtained by `Q-BASE`. These results point out the difficult choice for the designer of a tourist RS; the RS should be precise, but the implication is that it will then often suggest popular items that are likely to be already known by the tourist. Hence, in this way, the actual utility of the RS will be limited. `Q-BASE` tries to recommend novel (not popular) items that are estimated to be "rewarding" for the user based on the fact that tourists in the same cluster visit similar items to those recommended.

Clearly, the fact that the recommended next POIs are actually relevant and useful can only be assessed by a user or online study, which, however, presents other challenges: is the user capable to assess the satisfaction (reward) that the true visit experience to the recommended POIs will generate?

Online user study

In order to better understand the users' perceived novelty and expected satisfaction of the next-POI visit recommendations generated by `Q-BASE`, `Q-POP PUSH`, and `SKNN`, we have implemented a web-based application accessible from desktop and mobile browsers to simulate a visit to Florence (Italy) (Massimo and Ricci 2020). We recruited, via social media and mailing lists, 158 subjects who have actually visited Florence before the study. We wanted to address tourists that are somewhat familiar with the destination (and its POIs), so that they can better estimate the quality of next POI recommendations in this city: they should have visited already some POIs to evaluate the RS's suggestions about what to do next. We have designed a user/system interaction that enables the subjects to reflect and make choices as similarly as possible as for a real next-POI visit decision. The experimental system tries to generate the specific context of a true visit. During the interaction with the system, the subject is helped to imagine the real context and make decisions that will be likely to be taken when facing that decision task.

The application first profiles the subjects by asking them to list some of the previously visited POIs in that destination. This process is facilitated by the presence of pictures and descriptions of the POIs (Figure 1).

Then, a small number of POIs (5 items), among those declared to have already been visited, are used to build a personalized itinerary that each subject is supposed to have completed at the time point when she requests a next POI recommendation. Besides, in order to allow `Q-BASE` and `Q-POP PUSH` to generate recommendations, subjects are assigned to one of the pre-existing clusters, which are computed on the previously acquired tourists' visit trajectories.

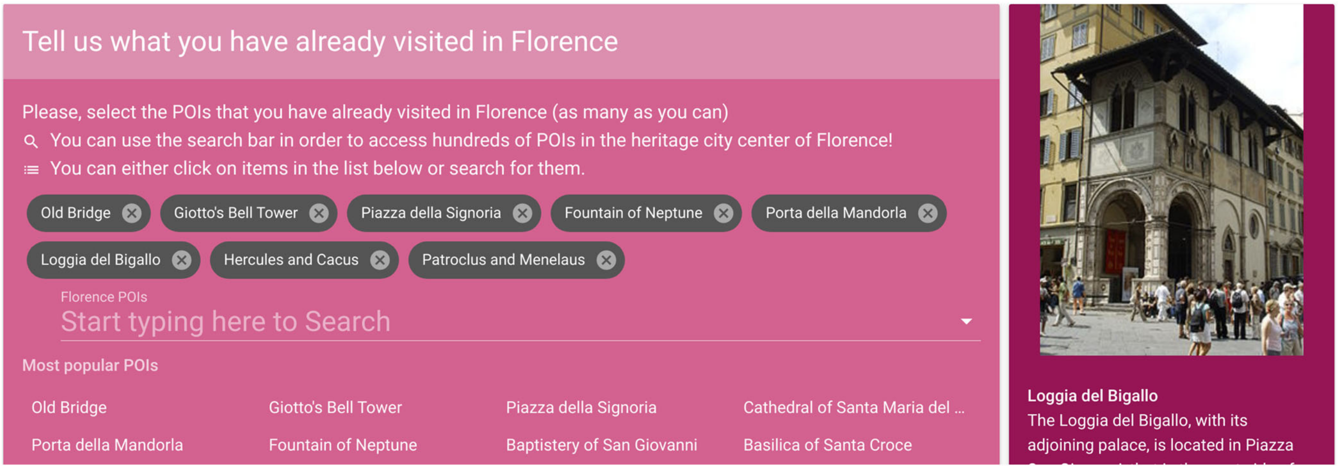

FIGURE 1 User profiling by POI selection UI

The cluster selected for a target subject is the one that best matches the POIs that the subject has declared to have already visited. Then, finally, at recommendation time the subjects are asked to evaluate a list of next-POI recommendations generated by mixing the recommendations computed by the three evaluated RSs: three recommendations for each RS. The subjects were not informed which RS recommends what. Hence, a small number of recommendations are generated, ranging from three, if all the three RSs suggest the same POIs, to nine, if they are all different. By using a designed GUI control, the subjects are then requested to judge if the recommended POIs have been previously "visited," are "liked" or are "novel." We aim at eliciting behavioral responses as close as possible as in a real condition. The user interface designed for the evaluation of the recommendations is shown in Figure 2.

An important aspect to consider, when discussing the results of an online study like this, is surely related to the question whether a subject/tourist could express a reliable "like" judgment on a POI that she does not know, that is, a "novel" POI, by simply relying on the system's presentation of the POI. In fact, while the other types of feedback ("visited" and "novel") are very likely to be correctly formulated, unless the tourist has forgotten some of the previous visit experiences, the "like" judgment is only a subjective signal that the tourist expects to have a rewarding (future) experience when visiting the recommended POI. Clearly, a liked POI may or may not result in a satisfying visit (rewarding), and, even more importantly, not liked POIs can still produce satisfying visits, when they are actually visited.

The obtained results are shown in Table 2. We measured the probability that a subject marks as "visited," "novel," "liked," or both "liked" and "novel" a POI recommended by a specific RS. Probabilities are estimated by dividing the total number of items marked as visited (liked, novel, and both liked and novel), for each RS, by the total number of recommendations offered by the RS.

It is clear that `Q-BASE` recommends POIs that are less likely to have been already visited by the subject, and more likely to be novel, compared to those suggested by `Q-POP PUSH` and `SKNN`. Interestingly, `Q-POP PUSH` and `SKNN` perform similarly, which seems to be connected to the popularity bias of both methods. It is evident that these results are matching the offline study results. This is not always true, as in many cases, offline results diverge from online ones, because different properties of the RS are measured in the two testing scenarios (Chen et al. 2017; Gunawardana and Shani 2015). But, we must also note that `Q-BASE` offers fewer POIs that are liked, compared to the other two recommendation strategies. Hence, apparently, `Q-BASE`, by trying to optimie the reward, is not equally able to produce recommendations that the subject likes. The rationale is that most of `Q-BASE` recommendations are actually novel, that is, the subject does not have an opinion about these items when they are presented. Therefore, the subject must understand whether she likes them or not, solely on the base of the provided information and explanation. This is complex and makes it difficult for the subject to formulate an assessment of the expected satisfaction for the future visit experience, which is supposed to determine the "like" evaluation. Despite this fact, it is interesting to note that `Q-BASE` generates more recommendations that are both liked and novel ("Liked and Novel" feedback), so, when a recommended POI is equally novel for all the three RSs, if it is suggested by `Q-BASE`, then it is more often liked. This matches well the main goal of a tourist RS: letting tourists to discover novel POIs that when visited will produce a satisfying experience. Still, we stress that the evaluation is based only on the subject's estimation of the true value of the recommendation, since POIs are here evaluated before they are experienced.

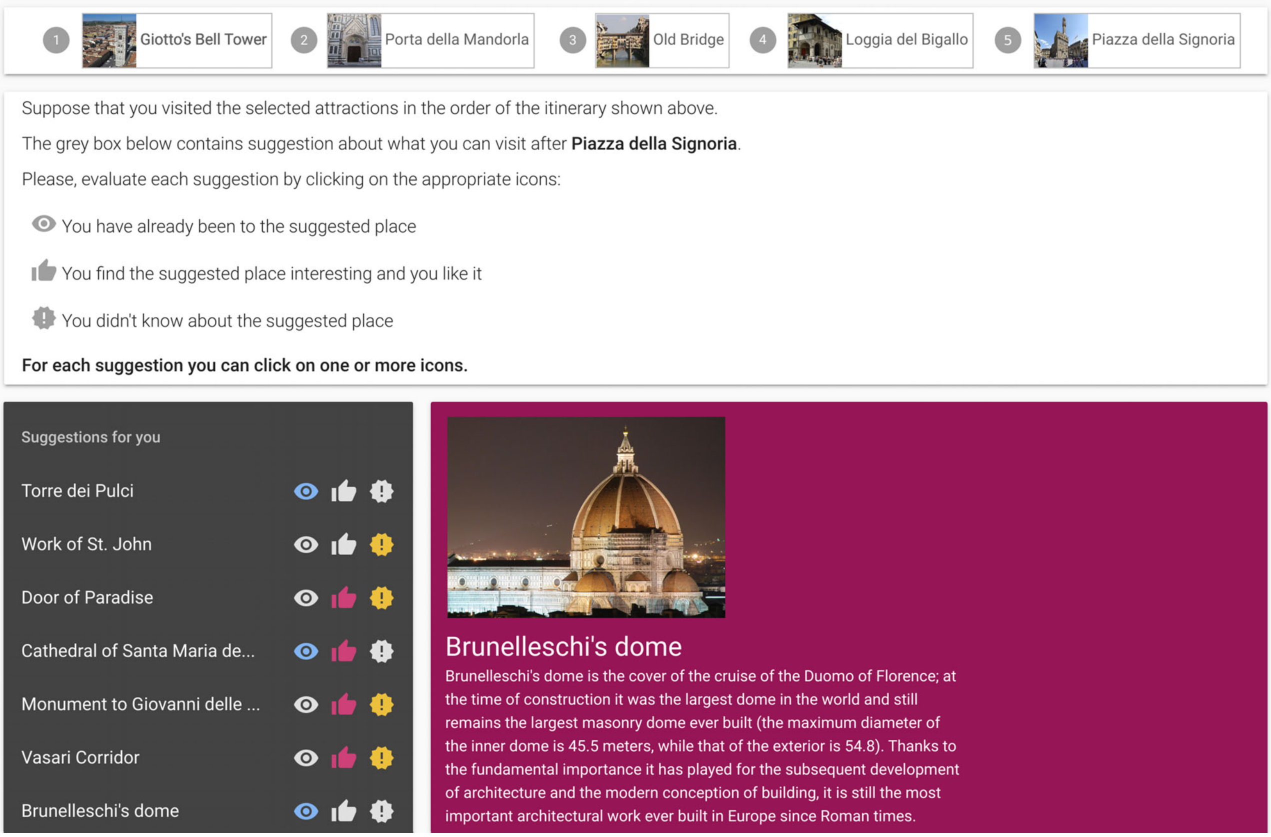

FIGURE 2 Evaluation GUI. From top to bottom: itinerary detail; info box; recommendations and item details

TABLE 2 Probability to evaluate a POI recommendation as visited, novel, liked, and both novel and liked

Recommender system	Visited	Novel	Liked	Liked and novel
`Q-BASE`	0.165*	0.517*	0.361*	0.091
`Q-POP PUSH`	0.245	0.376	0.464	0.076
`SKNN`	0.238	0.371	0.466	0.082

*indicates significant difference from the other two RSs perf. (two proportion z-test, $p < 0.05$).

By summarizing the results of the study, we derive the following conclusions. The POI-visit suggestions generated by `SKNN` and `Q-POP PUSH` are liked more than those produced by `Q-BASE`, because both RSs tend to recommend items that are less novel than those recommended by `Q-BASE`. Moreover, `Q-BASE`, in the attempt to optimize the reward function and suggesting items that have the properties typically liked by the user, does not care for the item popularity and often recommends novel POIs, which are hard to be appreciated. In fact, when the popularity bias is added to `Q-BASE`, that is, by using the hybrid model `Q-POP PUSH`, this IRL-based RS can produce results similar to that of `SKNN`.

Hence, this study illustrates a common "dilemma" in tourist RSs: tourists tend to like more the items they are familiar with, even POI that have been previously visited, but, useful recommendations are for items that are novel, which tend to be liked less. In fact, by analyzing the experimental data, we discovered that for all the three RSs, the probability that a user likes a recommended POI that she has already visited tends to be much larger than the probability to like a novel one. This is confirmed by the outcome of a post-survey in which participants declared that it is difficult to like something that is novel and unknown. This points out two main issues to be considered in the online evaluation of an RS. At first, it is unclear how users can judge items that they have not yet experienced. Then, it is unclear how an evaluation based on the user-perceived (expected) utility for an item can measure the actual utility that the user will gain in the real experience with the recommended item.

OPEN CHALLENGES FOR TOURISM RECOMMENDER SYSTEMS

We argue that in order to build effective tourism RSs, it urges to focus on the true needs of the users. We must develop models that are able to conceptualize what makes a POI worth to be visited, which implies that they must properly structure the available knowledge. This can enable the RS to learn *what* and *how* tourists consume POIs. The field has not yet achieved the level of development of other types of RSs because the research has not yet addressed its specific requirements and constraints (Werthner et al. 2015). Tourists that seek recommendations should be able to discover new POIs to visit: these are the POIs that they cannot easily find by themselves, for example, by using existing travel portals/guides, which generically suggest popular and highly rated items. We argue that tourism RSs should avoid to recommend blockbuster POIs, or at least accompany such POI recommendations with others that are novel, are perceived as worth trying and will actually produce rewarding experiences. To identify these items, we need to further study methods that are able to correctly estimate the quality of the experience that the tourist can gain by visiting a POI. As it emerged from our research, tourists struggle to judge POIs that are new to them even when they have a high estimated reward, that is, they fit the preferences learned by mining their observed behavior. This clearly suggests that there is a need to identify solutions to give users the ability to better assess the value of those items. We believe that it is important to focus even more on explanation methods for recommendations (Zhang and Chen 2020), especially approaches that can leverage the structural properties of IRL models (Ermon et al. 2015). For instance, we believe that by utilizing a proper knowledge to represent the observed POI-visit trajectories and then by learning the reward function for each POI visit and the associated POI-visit selection policy, we can then employ this information to devise explanation styles (Kouki et al. 2019) that can point out how and why the tourist should make the recommended visit choices (Jameson et al. 2014). In this way, it could be possible to build a more "persuasive" (conversational) system that nudges the user to accept and understand the proposed recommendations, and better help the user to evaluate the expected satisfaction of a visit to a possibly unknown POI.

A second aspect that the research on tourism RSs has to better discuss and consolidate is a proper evaluation approach. First of all, it is important to employ datasets of users' behavioral data (e.g., ratings or choices) and item descriptions (domain knowledge) that are representative of the real behaviors and interests of the tourists. Many existing data sets, including the ones that we have used, offer a partial description of user behavior, and they focus on special users in restricted group sets (e.g., location-based social networks). Moreover, what are the distinguished POIs to be considered and recommended is not obvious: new tourism services are continuously generated (Werthner and Klein 1999) and what is understood as a target POI by certain tourists is not even recognized as a POI by others. For instance, in our post user study survey, it emerged that the database of POIs that we employed was presenting items that were not easily identified as clear touristic landmarks by most of the subjects. For instance, many relatively small POIs (e.g., the door of a church) should be better collapsed into a unique broader POI (the church itself). This highlights the importance of a better definition of what is an item to be recommended, that is, an item that the tourist can judge as a worthy choice.

Furthermore, we would like to note that a promising way to overcome the obstacles in designing and running live user-studies is offered by counterfactual learning methods (Agarwal et al. 2017; Gilotte et al. 2018; Swaminathan and Joachims 2015). These techniques allow to assess offline a new recommendation strategy as if it would have been deployed and tested online, by means of a user study. This is implemented by using an existing data set, as in normal offline studies, but after having debiased the observations actually present in the data set. This avoids to overestimate in the observed users' behavior, choices that are not proper signals of users' preferences but are rather influenced by the recommendations the subject was exposed to (while the logged data were collected). Hence, one can re-weight the relevancy of certain observed POI-visit actions and eventually mitigate their importance, so that a more precise estimate of the true reward brought by different POIs can be computed. This "debiased" reward can then be used to assess offline the performance of a novel RS strategy without the burden of deploying it in an online system. This is termed counterfactual learning and it brings a specific benefit: it allows to bypass the difficulties related to set up a proper user study by allowing researchers to quantify the same objective in offline experiments as if they would have involved real users.

Besides, we believe that it is still an open research question how to correlate offline metrics, not only precision, to the perceived qualities and experience of the recommended item. By better understanding the factors that make a recommendation satisfactory for a user, we could operationalize offline metrics that quantify those factors. Hence, this can help to link real perceptions to quantifiable offline properties of the recommendations.

CONFLICT OF INTEREST

The authors declare that there is no conflict.

ETHICS STATEMENT

The institution approved the research and the conducted study. The user study was conducted anonymously, and users were provided with informed consent to participate in the study.

ORCID

David Massimo https://orcid.org/0000-0001-8853-8881
Francesco Ricci https://orcid.org/0000-0001-5931-5566

ENDNOTES

[1] www.google.com/maps
[2] https://foursquare.com/
[3] https://facebook.com
[4] www.flickr.com

REFERENCES

Abbeel, P., and A. Y. Ng. 2004. "Apprenticeship Learning via Inverse Reinforcement Learning." In *Proceedings of the Twenty-first International Conference on Machine Learning (ICML 2004)*, 1–8 Banff, Alberta, Canada. July 4–8.

Abdollahpouri, H., G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, and L. A. Pizzato. 2020. "Multistakeholder Recommendation: Survey and Research Directions." *User Modeling and User-Adapted Interaction* 30(1): 127–58.

Adamczak, J., Y. Deldjoo, F. B. Moghaddam, P. Knees, G.-P. Leyson, and P. Monreal. 2020. "Session-Based Hotel Recommendations Dataset: As Part of the ACM Recommender System Challenge 2019." *ACM Transactions on Intelligent Systems and Technology* 12(1): 1–20. November.

Adomavicius, G., B. Mobasher, F. Ricci, and A. Tuzhilin. 2011. "Context-Aware Recommender Systems." *AI Magazine* 32(3): 67–80.

Agarwal, A., S. Basu, T. Schnabel, and T. Joachims. 2017. "Effective Evaluation Using Logged Bandit Feedback from Multiple Loggers." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17*, 687–96 New York, NY, USA: Association for Computing Machinery.

Auksorncherdchoo, S., and A. Sukstrienwong. 2018. "Software Development for Student Behavior Tracking System." In *Proceedings of the 9th International Conference on E-Education, E-Business, E-Management and E-Learning, IC4E '18*, 55–9 New York, NY, USA: Association for Computing Machinery.

Babes-Vroman, M., V. Marivate, K. Subramanian, and M. Littman. 2011. "Apprenticeship Learning About Multiple Intentions." In *Proceedings of the 28th International Conference on Machine Learning —ICML'11*, 897–904.

Ball, P. 2010. "Wisdom of the Fool's Choice." *Nature*. https://doi.org/10.1038/news.2010.86

Baraglia, R., C. I. Muntean, F. M. Nardini, and F. Silvestri. 2013. "Learnext: Learning to Predict Tourists Movements." In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, CIKM'13*, eds. Q. He, A. Iyengar, W. Nejdl, J. Pei, and R. Rastogi, 751–6 San Francisco, CA, USA: ACM. October 27–November 1.

Bellogín, A., and A. Said. 2018. "Recommender Systems Evaluation." In *Encyclopedia of Social Network Analysis and Mining*, 2nd ed., edited by R. Alhajj and J. G. Rokne, 2095–112. New York, NY: Springer.

Bettman, J. R., M. F. Luce, and J. W. Payne. 1998. "Constructive Consumer Choice Processes." *Journal of Consumer Research: An Interdisciplinary Quarterly* 25(3): 187–217.

Bin, C., T. Gu, Y. Sun, and L. Chang. 2019. "A Personalized POI Route Recommendation System Based on Heterogeneous Tourism Data and Sequential Pattern Mining." *Multimedia Tools and Applications* 78(24): 35135–56.

Braunhofer, M., M. Elahi, and F. Ricci. 2014. "Usability Assessment of a Context-Aware and Personality-Based Mobile Recommender System." In *E-Commerce and Web Technologies*, edited by M. Hepp and Y. Hoffner, 77–88. Cham: Springer International Publishing.

Braunhofer, M., M. Elahi, and F. Ricci. 2015. "User Personality and the New User Problem in a Context-Aware Point of Interest Recommender System." In *Information and Communication Technologies in Tourism 2015, ENTER 2015, Proceedings of the International Conference in Lugano, Switzerland, February 3–6, 2015*, edited by I. Tussyadiah and A. Inversini, 537–49. Lugano, Switzerland: Springer.

Braunhofer, M., and F. Ricci. 2017. "Selective Contextual Information Acquisition in Travel Recommender Systems." *Journal of Information Technology & Tourism* 17(1): 5–29.

Castelluccia, C., E. De Cristofaro, and D. Perito. 2010. "Private Information Disclosure from Web Searches." In *Privacy Enhancing Technologies*, edited by M. J. Atallah and N. J. Hopper, 38–55. Berlin, Heidelberg: Springer.

Chen, H., C. Chung, H. Huang, and W. Tsui. 2017. "Common Pitfalls in Training and Evaluating Recommender Systems." *SIGKDD Explorations* 19(1): 37–45.

Choe, Y., D. R. Fesenmaier, and C. Vogt. 2017. "Twenty-Five Years Past Vogt: Assessing the Changing Information Needs of American Travellers." In *Information and Communication Technologies in Tourism 2017, ENTER 2017, Proceedings of the International Conference in Rome, Italy, January 24–*, edited by R. Schegg and B. Stangl, 489–502. Rome, Italy: Springer.

Cremonesi, P., Y. Koren, and R. Turrin. 2010, 01. "Performance of Recommender Algorithms on Top-N Recommendation Tasks." In *RecSys'10 —Proceedings of the 4th ACM Conference on Recommender Systems*, 39–46.

Dellaert, B. G., D. F. Ettema, and C. Lindh. 1998. "Multi-Faceted Tourist Travel Decisions: A Constraint-Based Conceptual Framework to Describe Tourists' Sequential Choices of Travel Components." *Tourism Management* 19(4): 313–20.

Elahi, M., M. Braunhofer, T. Gurbanov, and F. Ricci. 2018. "User Preference Elicitation, Rating Sparsity and Cold Start." In *Collaborative Recommendations—Algorithms, Practical Challenges and Applications*, edited by S. Berkovsky, I. Cantador, and D. Tikk, 253–94. Singapore: WorldScientific.

Ermon, S., Y. Xue, R. Toth, B. Dilkina, R. Bernstein, T. Damoulas, P. Clark, et al. 2015. "Learning Large Scale Dynamic Discrete Choice Models of Spatio-Temporal Preferences with Application to Migratory Pastoralism in East Africa." In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence Pattern*, 644–50.

García, B. 2004. “Urban Regeneration, Arts Programming and Mojor Events Glasgow 1990, Sydney 2000 and Barcelona 2004.” *International Journal of Cultural Policy* 10(1): 103–18.

Gilotte, A., C. Calauzènes, T. Nedelec, A. Abraham, and S. Dollé. 2018. “Offline a/b Testing for Recommender Systems.” In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18*, 198–206 New York, NY, USA: Association for Computing Machinery.

Gretzel, U., and D. R. Fesenmaier. 2006. “Persuasion in Recommender Systems.” *International Journal of Electronic Commerce* 11(2): 81–100.

Gretzel, U., N. Mitsche, Y. Hwang, and D. R. Fesenmaier. 2004. “Tell Me Who You are and I Will Tell you Where to go: Use of Travel Personalities in Destination Recommendation Systems.” *Journal of Information Technology & Tourism* 7(1): 3–12.

Gretzel, U., H. Werthner, C. Koo, and C. Lamsfus. 2015. “Conceptual Foundations for Understanding Smart Tourism Ecosystems.” *Computers in Human Behavior* 50: 558–63.

Gunawardana, A., and G. Shani. 2015. “Evaluating Recommender Systems.” In *Recommender Systems Handbook*, edited by F. Ricci, L. Rokach, and B. Shapira, 265–308. New York, NY: Springer.

Hariri, N., B. Mobasher, and R. Burke. 2012. “Context-Aware Music Recommendation Based on Latenttopic Sequential Patterns.” In *Proceedings of the 6th ACM Conference on Recommender Systems—RecSys'12*, 131.

Hashemi, S. H., and J. Kamps. 2017. “Where to Go Next?: Exploiting Behavioral User Models in Smart Environments.” In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP '17*, 50–8 New York, NY, USA: ACM.

Herlocker, J. L., J. A. Konstan, L. G. Terveen, and J. T. Riedl. 2004. “Evaluating Collaborative Filtering Recommender Systems.” *ACM Transactions on Information Systems (TOIS)* 22(1): 5–53.

Herzog, D., C. Laß, and W. Wörndl. 2018. “Tourrec: A Tourist Trip Recommender System for Individuals and Groups.” In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys '18*, 496–7 New York, NY, USA: Association for Computing Machinery.

Herzog, D., and W. Wörndl. 2019. “User-Centered Evaluation of Strategies for Recommending Sequences of Points of Interest to Groups.” In *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys '19*, 96–100 New York, NY, USA: Association for Computing Machinery.

Hong, M., S. An, R. Akerkar, D. Camacho, and J. J. Jung. 2019. “Cross-Cultural Contextualisation for Recommender Systems.” *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-019-01479-9

Huang, H., and G. Gartner. 2014. “Using Trajectories for Collaborative Filtering-Based POI Recommendation.” *International Journal of Data Mining Modelling and Management* 6: 333–46.

Hwang, Y., U. Gretzel, and D. R. Fesenmaier. 2002. “Behavioral Foundations for Human-Centric Travel Decision-Aid Systems.” In *Proceedings of the Information and Communication Technologies in Tourism, ENTER 2002*, eds. K. Wöber, A. J. Frew, and M. Hitz, 356–65. Innsbruck, Austria: Springer.

Jameson, A., B. Berendt, S. Gabrielli, C. Gena, F. Cena, F. Vernero, and K. Reinecke. 2014. “Choice Architecture for Human-Computer Interaction.” *Foundations and Trends in Human–Computer Interaction* 7(1–2): 1–235.

Jannach, D., and L. Lerche. 2017. “Leveraging Multi-Dimensional User Models for Personalized Next-Track Music Recommendation.” In *Proceedings of the Symposium on Applied Computing—SAC'17*, 1635–42.

Jannach, D., P. Resnick, A. Tuzhilin, and M. Zanker. 2016. “Recommender Systems: Beyond Matrix Completion.” *Communications of the ACM* 59(11): 94–102.

Karypis, G. 2001. “Evaluation of Item-Based Top-N Recommendation Algorithms.” In *Proceedings of the Tenth International Conference on Information and Knowledge Management, CIKM'01*, 247–54 New York, NY, USA: Association for Computing Machinery.

Knijnenburg, B. P., M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell. 2012. “Explaining the User Experience of Recommender Systems.” *User Modeling and User-Adapted Interaction* 22(4): 441–504.

Kouki, P., J. Schaffer, J. Pujara, J. O'Donovan, and L. Getoor. 2019. “Personalized Explanations for Hybrid Recommender Systems.” In *Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI'19*, 379–90 New York, NY, USA: Association for Computing Machinery.

Krishnamurthy, B., and C. Wills. 2009. “Privacy Diffusion on the Web: A Longitudinal Perspective.” In *Proceedings of the 18th International Conference on World Wide Web, WWW'09*, 541–50 New York, NY, USA: Association for Computing Machinery.

Kumar, G., H. Jerbi, and M. P. O'Mahony. 2017. “Towards the Recommendation of Personalised Activity Sequences in the Tourism Domain.” In *Proceedings of the 2nd Workshop on Recommenders in Tourism co-located with 11th ACM Conference on Recommender Systems (RecSys 2017), Volume* 1906 *of CEUR Workshop Proceedings*, eds. J. Neidhardt, D. R. Fesenmaier, T. Kuflik, and W. Wörndl, 26–30. Como, Italy. CEUR-WS.org. August 27.

Lamsfus, C., D. Wang, A. Alzua-Sorzabal, and Z. Xiang. 2014, 01. “Going Mobile: Defining Context for on-the-go Travelers.” *Journal of Travel Research* 54: 691–701.

Li, G., Q. Chen, B. Zheng, H. Yin, Q. V. H. Nguyen, and X. Zhou. 2020, March. “Group-Based Recurrent Neural Networks for POI Recommendation.” *ACM/IMS Transactions on Data Science* 1(1): 1–18.

Li, X., D. Han, J. He, L. Liao, and M. Wang. 2019, September. “Next and Next New POI Recommendation via Latent Behavior Pattern Inference.” *ACM Transactions on Information Systems* 37(4): 1–28.

Lim, K. H., J. Chan, C. Leckie, and S. Karunasekera. 2016. “Towards Next Generation Touring: Personalized Group Tours.” In *Proceedings of the Twenty-Sixth International Conference on Automated Planning and Scheduling, ICAPS 2016*, eds. A. J. Coles, A. Coles, S. Edelkamp, D. Magazzeni, and S. Sanner, 412–20 London, UK: AAAI Press. June 12–17.

Ludewig, M., and D. Jannach. 2018. “Evaluation of Session-Based Recommendation Algorithms.” *User Modeling and User-Adapted Interaction* 28(4-5): 331–90.

Madison, G., and G. Schiölde. 2017. “Repeated Listening Increases the Liking for Music Regardless of Its Complexity: Implications for the Appreciation and Aesthetics of Music.” *Frontiers in Neuroscience* 11: 147.

Marchiori, E., L. Cantoni, and D. R. Fesenmaier. 2013. “What Did They Say About Us? Message Cues and Destination Reputation in Social Media.” In *Information and Communication Technologies in Tourism 2013, ENTER 3013, Proceedings of the International*

Conference in Innsbruck, Austria, January 22–25, 2013, edited by L. Cantoni and Z. P. Xiang, 170–82. Innsbruck, Austria: Springer.

Massimo, D., and F. Ricci. 2018a. "Harnessing a Generalised User Behaviour Model for Next-POI Recommendation." In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018*, eds. S. Pera, M. D. Ekstrand, X. Amatriain, and J. O'Donovan, 402–6 Vancouver, BC, Canada: ACM. October 2–7.

Massimo, D., and F. Ricci. 2018b. "Harnessing a Generalised User Behaviour Model for Next-POI Recommendation." In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018*, eds. S. Pera, M. D. Ekstrand, X. Amatriain, and J. O'Donovan, 402–6 Vancouver, BC, Canada: ACM. October 2–7.

Massimo, D., and F. Ricci. 2020. "Next-POI Recommendations for the Smart Destination Era." In *Information and Communication Technologies in Tourism 2020*, edited by J. Neidhardt and W. Wörndl, 130–41. Cham: Springer International Publishing.

Massimo, D., and F. Ricci. 2021a. "Next-POI Recommendations Matching User's Visit Behaviour." In *Information and Communication Technologies in Tourism 2021*, edited by W. Wörndl, C. Koo, and J. L. Stienmetz, 45–57. Cham: Springer International Publishing.

Massimo, D., and F. Ricci. 2021b. "Next-POI Recommendations Matching User's Visit Behaviour." In *Information and Communication Technologies in Tourism 2021*, edited by W. Wörndl, C. Koo, and J. L. Stienmetz, 45–57. Cham: Springer International Publishing.

Matzarakis, A. 2006, 08. "Weather-and Climate-Related Information for Tourism." *Tourism and Hospitality Planning & Development* 3: 99–115.

McKenzie, G., and K. Janowicz. 2014. "Activities in a New City: Itinerary Recommendation Based on User Similarity."

McNee, S. M., J. Riedl, and J. A. Konstan. 2006. "Being Accurate is not Enough: How Accuracy Metrics have Hurt Recommender Systems." In *Extended Abstracts Proceedings of the 2006 Conference on Human Factors in Computing Systems, CHI 2006*, eds. G. M. Olson and R. Jeffries, 1097–101 Montréal, Québec, Canada: ACM. April 22–27.

Moling, O., L. Baltrunas, and F. Ricci. 2012. "Optimal Radio Channel Recommendations with Explicit and Implicit Feedback." In *Proceedings of the 6th ACM Conference on Recommender Systems—RecSys'12*, 75.

Moutinho, L. 1987. "Consumer Behavior in Tourism." *European Journal of Marketing* 21: 2–44.

Muntean, C. I., F. M. Nardini, F. Silvestri, and R. Baraglia. 2015. "On Learning Prediction Models for Tourists Paths." *ACM Transactions on Intelligent Systems and Technology* 7(1): 8:1–8:34. October.

Nguyen, T. N., and F. Ricci. 2018. "A Chat-Based Group Recommender System for Tourism." *Information Technology and Tourism* 18(1): 5–28.

Oppokhonov, S., S. Park, and I. K. E. Ampomah. 2017. "Current Location-Based Next POI Recommendation." In *Proceedings of the International Conference on Web Intelligence, WI'17*, 831–6 New York, NY, USA: ACM.

Palumbo, E., G. Rizzo, and E. Baralis. 2017. "Predicting Your Next Stop-over from Location-Based Social Network Data with Recurrent Neural Networks." In *Proceedings of the 2nd ACM International Workshop on Recommenders in Tourism (RecTour'17)*, CEUR Proceedings Vol. 1906, 1–8.

Perentis, C., M. Vescovi, and B. Lepri. 2015. "Investigating Factors Affecting Personal Data Disclosure." In *Proceedings of the 24th International Conference on World Wide Web, WWW'15 Companion*, 89–90 New York, NY, USA: ACM.

Poikela, M., R. Schmidt, I. Wechsung, and S. Möller. 2015. "About Your Smartphone Usage" - Privacy in Location-Based Mobile Participation." In *Proceedings of the 2015 IEEE International Symposium on Technology and Society, ISTAS 2015*, 1–6 Dublin, Ireland: IEEE. November 11–12.

Powers, D. 2008. "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness and Correlation." *Machine Learning Technologies* 2: 37–63.

Pu, P., L. Chen, and R. Hu. 2011. "A User-Centric Evaluation Framework for Recommender Systems." In *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys'11*, 157–64 New York, NY, USA: Association for Computing Machinery.

Quadrana, M., P. Cremonesi, and D. Jannach. 2018. "Sequence-Aware Recommender Systems." *ACM Computing Surveys* 51(4): 1–36.

Rabanser, U., and F. Ricci. 2005. "Recommender Systems: Do They Have a Viable Business Model in e-Tourism?." In *Information and Communication Technologies in Tourism, ENTER 2005, Proceedings of the International Conference in Innsbruck, Austria, 2005*, edited by A. J. Frew, 160–71. Innsbruck, Austria: Springer.

Rani, S., K. N. Kholidah, and S. N. Huda. 2018. "A Development of Travel Itinerary Planning Application Using Traveling Salesman Problem and K-Means Clustering Approach." In *Proceedings of the 7th International Conference on Software and Computer Applications, ICSCA 2018*, eds. K. Z. Zamli, V. Mezhuyev, and L. Benedicenti, 327–31 Kuantan, Malaysia: ACM. February 08–10.

Ricci, F., L. Rokach, and B. Shapira, eds. 2015. *Recommender Systems Handbook*. New York, NY: Springer.

Sánchez, P., and A. Bellogín. 2020. "Applying Reranking Strategies to Route Recommendation Using Sequence-Aware Evaluation." *User Modeling and User-Adapted Interaction* 30(4): 659–725.

Savard, I., and R. Mizoguchi. 2019. "Context or Culture: What is the Difference?." *Research and Practice in Technology Enhanced Learning* 14(1): 23.

Shani, G., D. Heckerman, and R. I. Brafman. 2005. "An MDP-Based Recommender System." *Journal of Machine Learning Research* 6: 1265–95.

Silva, T. H., A. C. Viana, F. Benevenuto, L. Villas, J. Salles, A. Loureiro, and D. Quercia. 2019, February. "Urban Computing Leveraging Location-Based Social Network Data: A Survey." *ACM Computing Surveys* 52(1): 1–39.

Sánchez, P., and A. Bellogín. 2020. "Time and Sequence Awareness in Similarity Metrics for Recommendation." *Information Processing & Management* 57(3): 102228.

Staab, S., H. Werthner, F. Ricci, A. Zipf, U. Gretzel, D. R. Fesenmaier, C. Paris, and C. A. Knoblock. 2002. "Intelligent Systems for Tourism." *IEEE Intelligent Systems* 17(6): 53–64.

Sutton, R. S., and A. G. Barto. 1998. *Reinforcement Learning—An Introduction. Adaptive Computation and Machine Learning*. Cambridge, Massachusets, USA: MIT Press.

Swaminathan, A., and T. Joachims. 2015. "Counterfactual Risk Minimization: Learning from Logged Bandit Feedback." In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Volume 37 of JMLR Workshop and Conference Proceedings*, eds. F. R. Bach and D. M. Blei, pp. 814–23 Lille, France. JMLR.org. 6–11 July.

Swarbrooke, J., and S. Horner. 2006. *Consumer Behaviour in Tourism*, 2nd ed. London, UK: Butterworth-Heinemann.

Torrijos, S., A. Bellogín, and P. Sánchez. 2020. "Discovering Related Users in Location-Based Social Networks." In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP'20*, 353–7 New York, NY, USA: Association for Computing Machinery.

Vargas, S., and P. Castells. 2011. "Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems." In *Proceedings of the 5th ACM conference on Recommender systems —RecSys'11*, 109.

Vargas, S., and P. Castells. 2014. "Improving Sales Diversity by Recommending Users to Items." In *Proceedings of the 8th ACM Conference on Recommender Systems, RecSys'14*, 145–52. New York, NY, USA: Association for Computing Machinery.

Wang, H., H. Shen, W. Ouyang, and X. Cheng. 2018. "Exploiting POI-Specific Geographical Influence for Point-of-Interest Recommendation." In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018*, ed. J. Lang, 3877–83. Stockholm, Sweden. ijcai.org. July 13–19.

Werthner, H., A. Alzua-Sorzabal, L. Cantoni, A. Dickinger, U. Gretzel, D. Jannach, J. Neidhardt, et al. 2015. "Future Research Issues in IT and Tourism." *Journal of Information Technology & Tourism* 15(1): 1–15.

Werthner, H., and S. Klein. 1999. "ICT and the Changing Landscape of Global Tourism Distribution." *Electronic Markets* 9(4): 256–62.

Werthner, H., and F. Ricci. 2004. "E-Commerce and Tourism." *Communications of the ACM* 47(12): 101–5.

Wörndl, W., A. Hefele, and D. Herzog. 2017. "Recommending a Sequence of Interesting Places for Tourist Trips." *Information Technology and Tourism* 17(1): 31–54.

Yao, D., C. Zhang, Z. Zhu, J. Huang, and J. Bi. 2017. "Trajectory Clustering via Deep Representation Learning." In *Proceedings of the 2017 International Joint Conference on Neural Networks, IJCNN 2017*, 3880–7 Anchorage, AK, USA: IEEE. May 14–19.

Yiannakis, A., and H. Gibson. 1992. "Roles Tourists Play." *Annals of Tourism Research* 19(2): 287–303.

Zhang, J.-D., C.-Y. Chow, and Y. Li. 2014. "Lore: Exploiting Sequential Influence for Location Recommendations." In *Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL'14*, 103–12 New York, NY, USA: ACM.

Zhang, W., and D. R. Fesenmaier. 2018. "Assessing Emotions in Online Stories: Comparing Self-Report and Text-Based Approaches." *Journal of Information Technology & Tourism* 20(1-4): 83–95.

Zhang, Y., and X. Chen. 2020, 01. "Explainable Recommendation: A Survey and New Perspectives." *Foundations and Trends®in Information Retrieval* 14: 1–101.

Zhao, P., H. Zhu, Y. Liu, J. Xu, Z. Li, F. Zhuang, V. S. Sheng, and X. Zhou. 2019. "Where to Go Next: A Spatio-Temporal Gated Network for Next POI Recommendation." In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, 5877–84. Honolulu, Hawaii, USA: AAAI Press. January 27–February 1.

Zheng, Y. 2015. "Trajectory Data Mining: An Overview." *ACM Transactions on Intelligent Systems and Technology* 6(3): 1–41. May.

Zhou, T., Z. Kuscsik, J. Liu, M. Medo, J. R. Wakeling, and Y. Zhang. 2010. "Solving the Apparent Diversity-Accuracy Dilemma of Recommender Systems." *Proceedings of the National Academy of Sciences* 107: 4511–5.

AUTHOR BIOGRAPHIES

David Massimo is a Post-doctoral Researcher at the Free University of Bozen-Bolzano, Italy. His research focuses on recommender systems and, in particular, on their application in the tourism domain. The outcome of his research has been industrialized and actively used in real scenarios.

Francesco Ricci is a Professor of Computer Science at the Free University of Bozen-Bolzano, Italy. His research focuses on recommender systems and decision support, in context-aware and group recommendation scenarios, with a particular attention to the tourism application domain. He co-edited the Handbook of Recommender Systems (2022, Springer).

How to cite this article: Massimo, D., and F. Ricci. 2022. "Building effective recommender systems for tourists." *AI Magazine* 43: 209–24.
https://doi.org/10.1002/aaai.12057