

## I Know What You Did Next Summer: Challenges in Travel Destination Recommendation

DMITRI GOLDENBERG, SARAI MIZRACHI, and ADAM HOROWITZ, Booking.com, Tel Aviv, Israel  
IOANNIS KANGAS, OR LEVKOVICH, MAUD SCHWOERER, ALESSANDRO MOZZATO, MICHELE  
FERRETTI, PANAGIOTIS KORVESIS, and LUCAS BERNARDI, Booking.com, Amsterdam, Netherlands

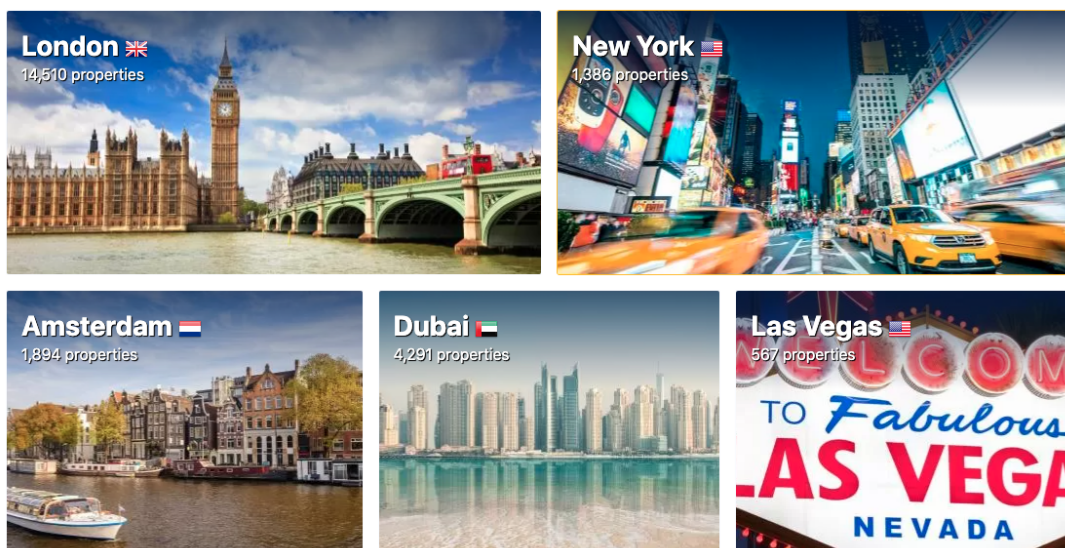


Fig. 1. An example of destination recommendations on Booking.com website

Picking a destination is the first step of any trip planning. While some travelers have an exact destination in mind, others have some degree of open-mindedness and can thus benefit from receiving destination recommendations. Online travel platforms often actively recommend a wide variety of travel destinations to their customers; however, customers at different stages of trip planning may expect different recommendations to serve different needs such as good alternative destinations or destinations for extending a trip. This makes recommending a destination challenging, requiring a complex synthesis of available contextual data along with consideration of a customer's needs and the holistic user experience of receiving recommendations. At a technical level, destinations also represent complex entities with ambiguous geographical properties and naming conventions. In this paper, we discuss the technical and customer-centric challenges of building real-life destination recommendation systems. We supplement this by providing applied solution examples from Booking.com, one of the world's leading online travel platforms.

CCS Concepts: • **Information systems** → **Personalization; Recommender systems.**

Additional Key Words and Phrases: Personalization, Travel, Recommender Systems

### 1 INTRODUCTION

Online travel platforms offer a vast variety of products that serve different needs of travelers. Such platforms often rely on recommender systems [5] to assist various customer decisions, seeking to help narrow down diverse offerings

by suggesting well-fit options in a personalized manner [24]. Picking a destination is a high-impact decision in trip planning, affecting subsequent decisions about transportation, accommodations or things to do while travelling.

Choosing a travel destination is not a trivial task. Availability, information gaps, budget, timing, preferences and even weather constraints introduce non-negligible complexity [19]. While some travelers have an exact destination in mind, others have some degree of open-mindedness. Destination recommendations can help travelers navigate a vast landscape of available options. However, customers are likely to benefit from destination recommendations differently in different contexts. For example, the way in which destination recommendations can help someone who has not begun travel planning (*"I'm not even thinking about traveling yet"*) is not the same as someone who has a specific trip already in mind (*"I want to go to Barcelona"*), and completely different for someone who is somewhere in-between (*"I just need a break on the beach"*). Because customers aren't used to expressing their travel intent explicitly when interacting with online travel platforms, destination recommendation systems need to be able to interpret a full spectrum using implicit cues as inputs, including cues set by user context [22].

From a customer point of view, the choice of a travel destination can be viewed as a decision making problem, where a single or multiple travel destinations are chosen out of a large but finite set of available alternatives [19, 20]. Destination recommendation systems may help addressing such decision-making problems by highlighting relevant travel alternatives.

Recommending a destination thus requires a complex consideration of available context data [33], customer needs, recommendation purpose, and holistic user experience [17]. Further, destination items themselves represent complex entities with ambiguous geographical properties and naming conventions [15].

Recommending travel destinations therefore presents unique challenges including, but not limited to, continuous cold-start [4], seasonality, and long time windows between an initial interaction and trip completion [39]. Moreover, recommending destinations often requires a system to capture unique information about travel trends, trip preferences, and timing [27], generally with limited context data to rely on [37]. These challenges necessitate employing sophisticated modeling approaches, such as sequential learning [25, 26], combining different feature types within model [40, 45], and introducing online adaptive algorithms [3]. Together with domain knowledge on travel behavior and destination geo-properties, these methods can offer unique, tailored-made solutions to the challenge of destination recommendations.

This work provides an overview of the domain-specific challenges in building real-life destination recommendation systems. We discuss the unique characteristics of travel recommendations and, in particular, destination selection. We also explain the special context of customer interaction with such recommender systems and survey the complexity of problems related to destination items and to mining travel patterns. We demonstrate modeling approaches alongside engineering and user experience solutions to overcome these challenges, including by discussing a set of real, applied recommender system production examples from Booking.com.

## 2 CHALLENGES

In this section, we focus on specific challenges that we have encountered while building a destination recommendation systems. We break them down into three categories, according to their relationship to user travel patterns and interactions with the recommender system (or lack thereof), destination characteristics, and customer behavior and needs.

### 2.1 Travel Patterns

Mining and understanding travel behavior patterns of users can surface critical challenges that exist in destination recommendation systems.

### 2.1.1 Continuous Cold Start.

One of the most discussed problems in industrial recommendation systems is the so-called 'Cold Start' problem [38, 44]. During a cold start, a recommender system has little or no information at all about users or items, posing a high difficulty in making accurate predictions for these users and items. Destination recommendation systems tend to suffer from a Continuous Cold Start (CoCoS) problem [4], wherein sufficient information may be lacking even for known users and/or items. Specifically, since travel tends to be infrequent (with most travellers booking trips once or twice per year), even for already-existing customers, what is contextually relevant at present may be highly distinctive from previous interactions. For customers who do travel more often, different trips still often have different purposes and configurations. For instance, what is contextually relevant will differ for a family beach holiday versus a solo business trip.

**2.1.2 Trip Structure.** Trips are often conceptualized in terms of travel to a single destination; travelers often go to multiple destinations within a single trip, however. Unlike classical e-commerce platforms, a sequence of reservations on an online travel platform often reflects a sequence of physical movement by a traveler. Take, for example, a multi-destination vacation in Italy, as in Figure 2. The trip starts from Milan, then goes to Venice, Florence, Rome, Naples, Reggio Calabria, and finishes in Palermo. The sequential order of destinations follows the geographical location of each of the destination, and this sequential order is therefore more likely than alternative reshuffled sequences of the same cities. This sequence will not necessarily be the proper fit for all travelers, however. The specific destinations selected in building an itinerary can have varying dependencies, such as when a person is traveling, what sites or landmarks they are aware of, and other personal travel preferences. A recommender system can help a customer extend their trip by adding new destinations to their itinerary, but must do so in a way that accounts for these variances.

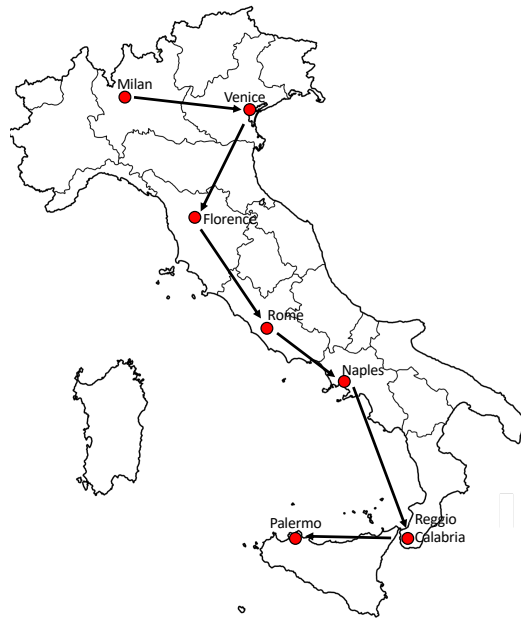


Fig. 2. Example Multi-Destination-Trip in Italy

**2.1.3 Unique needs and preferences.** A successful recommendation system needs to take into consideration the special needs and preferences of its users. In the context of destination recommendation systems, there are several travel requirements that have been traditionally neglected, such as accessibility [9], and that have been recently gaining more importance, such as sustainability [8]. In some instances, these overlooked needs and preferences can actually be strong enough as to act as a filtering mechanism for a customer. Hence, accounting for typically overlooked needs and preferences can increase the quality of a recommendation system's output [23].

## 2.2 Destinations

Destination recommender systems are further challenged by the complexity of the item set.

### 2.2.1 Geospatial Characteristics.

Not all destinations have the same set of characteristics, offering different things to see, do, and experience. Geospatial differences are highly consequential for travel: beach destinations and ski destinations will gain or lose relevance during different times of year, for example. This makes it essential for destination recommendations to be aware of geospatial distinctions. For instance, if seeking to recommend an alternative to a beach destination, those recommendations should be similarly beach-based.

To efficiently arrange geospatial information, traditional approaches rely on *gazetteers*, dictionaries of named locations that are used to link unstructured data to geographic space. [1, 28]. *Gazetteers* pinpoint geo-identifiers, such as locally relevant points-of-interest. While gazetteers provide a straightforward way of organising spatial data, they present unclear boundaries and have no spatial scales. This can cause overlapping spatial units to confound recommendations, with relevant items missing from the appropriate classes while inappropriately surfacing non-relevant ones. "New York City" as a classification is an example of this; this classifier can be defined to include either Manhattan only or to include all five of the city's boroughs.

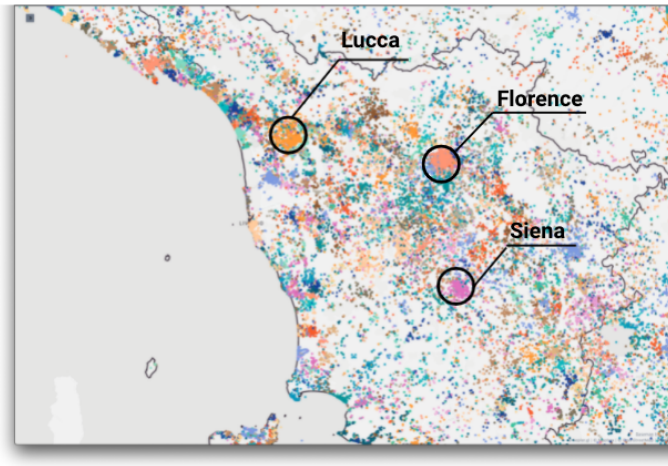


Fig. 3. Property locations in Tuscany (Italy), color-coded according to the gazetteer nomenclature. Beyond such artificial separations, subjective factors influence a user's perception of and preference for a given destination e.g. where Florence actually is?



Further, a mismatch between destination and real location introduces noise in feature values which are used to classify destination attributes (such as distance to certain points-of-interest, nearby destinations, etc). Figure 3 demonstrates the challenge of recommending popular travel destinations in Tuscany, Italy. Given the absence of boundaries, property assignment *within* the recommended destination is no longer just a data-driven decision. Rather, multiple subjective factors such as the partner preference for a certain market, as well as the user's perception of a destination, factor into shaping the perception of where an item *actually* is.

Naming conventions further aggravate the issue. While properties' locations may surface technical details about locality names, customers often are interested in specific regions due to unique travel intentions. As an additional example, while destination names like *Magny-le-Hongre* or *Bussy-Saint-Georges*, may not be familiar for an international traveler, recommending them as the *Disneyland* area in France will set the right context to the customers.

### 2.2.2 Availability.

An interesting challenge arises from the fact that destinations are not the final product being sold from an online travel platform. For example, at Booking.com, a user chooses the destination in order to book an accommodation. Therefore, a recommendation engine is required to ensure that there is sufficient inventory in the proposed destinations. This becomes crucial in a scenario of large-scale batch recommendation such as marketing. For instance, emails with recommendations sent to millions of subscribers through batch jobs can trigger millions of searches to destinations with limited room capacity or poor inventory quality. Consequently, if availability is not taken into account, many subscribers will end up landing on the website on destinations with few properties, or, even worse, with no available accommodations for their dates of interest.

Furthermore, it is even more challenging to recommend destinations which have supply that is actually *relevant* for the customer. To provide the best service for our customers and avoid bad user experience, it is not enough to have sufficient availability of properties in a destination. Instead, there should be availability of "high quality" items, as perceived by a particular guest's preferences; namely, do the recommended destinations include sufficient number of relevant properties, which users would be more interested in booking? The disruption to the travel industry caused by COVID-19 further exacerbated a shift from quantity to quality, as a decrease in supply caused a focus on on quality items [35].

### 2.2.3 Dynamic Environment and Constraints.

Creators of a destination recommender system also need to take into consideration the constantly changing nature of

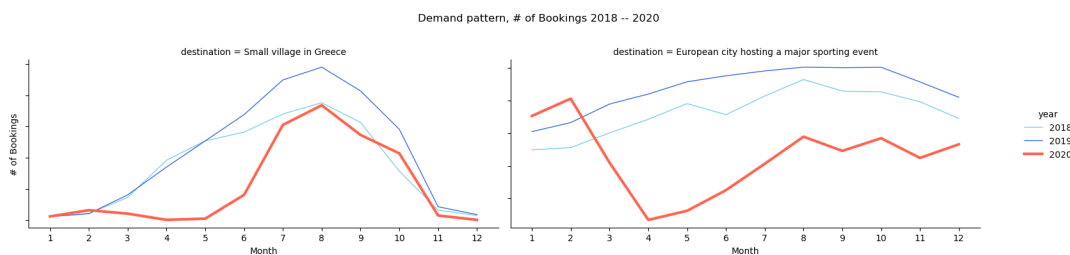


Fig. 4. Dynamicity and constraints due to seasonal and special events: i) Left panel shows shift in demand due to COVID-19 in a small village in Greece. ii) right panel shows demand in a European city hosting a major sporting event (May 2018), alternated to random pattern trends after COVID-19 hit. The data are collected from Booking.com.

the item set, influenced by two main categories of factors: *dynamic factors*, such as seasonality or special events, and *constraints*, such as legal restrictions to travelling (e.g., due to COVID-19). An ideal system should be able to factor both in accordingly, increasing recommendations of in-season items, while filtering out destinations for which travelling is restricted. Among the first set of factors, seasonality plays a major role, leading to extreme high and low peaks of demand [2]. The added complexity comes in understanding and incorporating the causal factors of these demand fluctuations, such as weather conditions and public holidays. An example of a seasonal destination is shown in the left panel of Figure 4, with the typical seasonal booking pattern of a top summer destination in a small Greek village. In addition to seasonality, other special events might periodically or temporarily increase the demand of some destinations, thereby throwing off recommendations. For example, in the right panel of Figure 4 the peak of May 2018 caused by the football's Champions league final in a major European city is plotted.

Moreover, dynamic constraints also play a critical role in destination recommendation. This is highlighted during the pandemic of COVID-19, where traveler choices are restricted and constantly revised. The effect of COVID-19 is shown in Figure 4 where the summer destination had a shift of almost two months for the in-season reservations in 2020 compared to the previous two years and the Champions league final hosting city had an arbitrary demand pattern in 2020 compared to the previous two years.

## 2.3 Customer-centricity

To build an effective recommender system, it's essential to look how users interact with an online travel platform and what they expect from a destination recommendation - not only in terms of which recommendations are offered, but also in how they are presented to them.

### 2.3.1 Matching Goal and Purpose.

Customers arrive at online travel platforms in different stages of their travel planning and booking process. In each of these, customers have different most-relevant goals. The choice of a travel destination can be viewed as a decision making processes in which the customer makes a choice of one or multiple destinations per trip, out of a finite set of alternative destinations [20]. The customers' travel destination decisions can be influenced by various factors such as their own characteristics and travel planning demands or needs, the characteristics of the travel destinations, the available set of destinations and the information the customer has available per destination [19]. The purpose of a destination recommendation - or the meaning it will bring to a customer - will consequently differ per travel planning stage.

We might, for instance, consider the following: A phase of **discovery**, in which customers are still collecting options that might suit their travel. In this phase, a customer may have expressed only minimal intent to travel, for instance by making an initial visiting the online travel platform. Here, the purpose of a recommendation would be to reinforce travel intent and reduce friction to choose a destination.

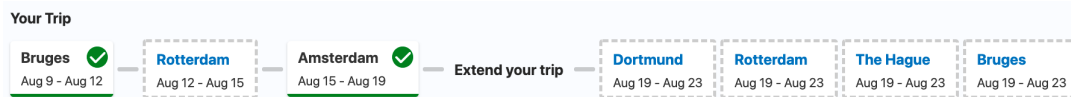


Fig. 5. Multi-Destinations Trip extension recommendation bar

**Alternatives**, is a phase of considerations in which a customer may be amenable to considering alternative destinations that are potentially better-matched to them, perhaps because their first choice of destination was not available or did not fit their budget. Here, the purpose of a recommendation is to surface alternative destinations that the customer is likely to view as acceptable replacements for an original choice.

Another phase of considerations exists when a customer may have completed one booking in one destination but has not yet completed the itinerary planning in full. Here, the purpose of a recommendation can be to help the customer select the next, **complementary** destination to visit before or after the stay that was purchased (such as in Figure 5)

Correspondingly, an important challenge for the author of a destination recommender system becomes navigating between a wide set of different system types, each of which has a purpose tailored to a specific travel planning goal.

### 2.3.2 *The HOW is as Important as the WHAT.*

How recommendations are presented on a page is often more important than the way they are generated. Copywriting, design, and UI can be adapted to and optimized for different user needs. For instance, the fit of a recommendation may only be understandable by the relevant information about it that is displayed (such as availability in Figure 8 or thematic context in Figure 9). Distance from a searched location, ease of public transport connections, the number of available accommodations, the average price of accommodation, or potential discounts may be relevant, but may also provide information overload if all presented. Decisions about what information is relevant and in what context must accordingly be considered.

## 3 SOLUTION APPROACHES

Following the above-mentioned challenges, this section demonstrates uniquely tailored approaches to serve destination recommendation. We present applied examples of utilizing domain-specific knowledge, designing user experience and specific modeling approaches for destination recommendation challenges.

### 3.1 Domain Knowledge

#### 3.1.1 *Cross Learning.*

Given a wide catalogue of travel products (accommodations, flights, cars, taxis, attractions) we face a unique opportunity to provide the users with the most relevant destination recommendations across these various products. Cross-learning between travel product can be applied to mitigate the bias impact introduced by missing data on the user journey as well as the Continuous Cold Start problem (subsubsection 2.1.1). We can utilize data from user interactions and bookings of multiple products to infer user travel patterns on other, non-observed, behaviours and bookings and improve our cross product recommendations. For example, by linking data from car rentals pick up airport agencies locations with hotel bookings, we are able to learn about the potential airport catchment areas and apply these areas as basis for a flight-destination recommendation (see more details in subsubsection 3.1.3).

#### 3.1.2 *Dealing with Dynamic and Supply Constraints.*

An example of dealing with supply constraints in destination recommendation, given dynamic properties of the destinations set itself (as discussed in subsubsection 2.2.3) can be seen in our featured "domestic trending destinations" recommender. This system is able to provide tailored recommendations by leveraging items from a near-past window to leverage recent and seasonal information. Moreover, this solution is agnostic to a destination type; customers choose from a pool of recommendations ranging from entire regions to small islands, which partially solves supply constraints.

Finally, to adapt with disrupted travels due to COVID-19 restrictions we retrained our model on destinations within a fixed maximum distance and introduced hard constraints online to retrieve only destinations within the same country.

Another aspect of supply-constraint challenges, related to hotel properties availability (see subsection 2.2.2), can be addressed by factoring in such constraints directly in the modelling process. This can be done by first identifying and measuring the accommodation supply "quality" of destinations. A proposed potential metric for supply quality considers the proportion of availability of top properties, out of the total availability of properties within an area. This allows measuring the optimal level and mix of the supply, and its relevance to the user, instead of just the quantity.<sup>1</sup> Integrating this metric as part of the recommender model can better reflect supply-limited nature of travel destinations, and set a hard constraint for avoiding recommending destinations that have poor quality availability or are sold out.

### 3.1.3 Exploiting Geographical Knowledge.

To address geo-data related challenges we experimented with several solutions. As a general solution we leveraged industry-adopted standards such as spatial indexes. These are a class of space-partitioning algorithms that allow us to efficiently query items at a global scale, but also serve as basis (e.g. aggregational units) for analytical tasks [7]. On top of this framework we built custom solutions. A prime example of it is our destination recommendation in flights, which can be divided roughly into three main cases - recommendation for booking flight destinations, recommendation for flights relevant origin airport given users' geo-location,<sup>2</sup> and a recommendation for travel destinations for users who have already booked a flight. In all cases, a main challenge is recommending a travel airport and destinations in the area served by the airport, which requires understanding the geographical extent of the recommended class.

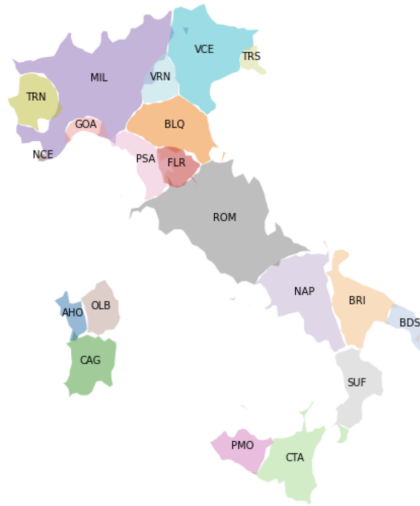


Fig. 6. Delineation of main city-level IATA code catchment areas for destination recommendation classes in Italy

<sup>1</sup>Calculating and evaluating this metric could consider additional dimensions such as the supply quality within specific travel dates, and over stated user preferences such as hotel or room attributes filtering.

<sup>2</sup>Recommending origin airports is important for air travellers. The importance of recommending the relevant origin airport is crucial given seasonal activity of airports and carrier-routes, and particularly given fast changes in travel restrictions (which create a barrier for crossing borders to get to an airport, which is common in many parts of Europe).

Flights users choose airports as destinations, but airports themselves are only a stop in the traveller's journey before reaching their destinations. For this reason, it's important to understand how travel destinations are linked with airport by delineating catchment areas.<sup>3</sup> To classify destinations to service airports area we followed several modeling approaches. The obtained catchment areas (see an example of main Italy airports in Figure 6) are then used to group destinations into classes for recommending destinations for flights users:

**Flights travel patterns data and Cross-learning between travel products.** Airport catchment areas can be identified by linking flight user-searches data to identified geo-location. Additionally, service areas can also be delineated by applying cross-learning from other travel products. One example for this is by making use of data on users' rental cars airport pick-up location, and link it to data on their chosen travel destination (hotel bookings).

**Gravity model approach.** Given low data availability, a simple non-constrained gravity model approach for modeling spatial interaction [21, 29, 32] can link travel destinations to the most relevant service airport. The links' "strength" is positively related to airport traffic, and negatively related to travel costs to destinations (proxied by travel distance), national border crossing and geo-barriers.

### 3.2 Design and User Experience

UX and UI choices are key for the success of recommender systems. Rapid and iterative experimentation for both ML and UX are crucial for reaching impactful solutions.

#### 3.2.1 Adapting UI according to Customer Needs.

When a customer first comes to a travel platform, there is little context or user history to use for modeling. The subsequently generic recommendations can be made more tailored as user interactions progress. At Booking.com, for instance, a searched or clicked recommendation on the main landing page becomes a seed to adapt the next recommendations shown, even within the same session, as illustrated on Figure 7.

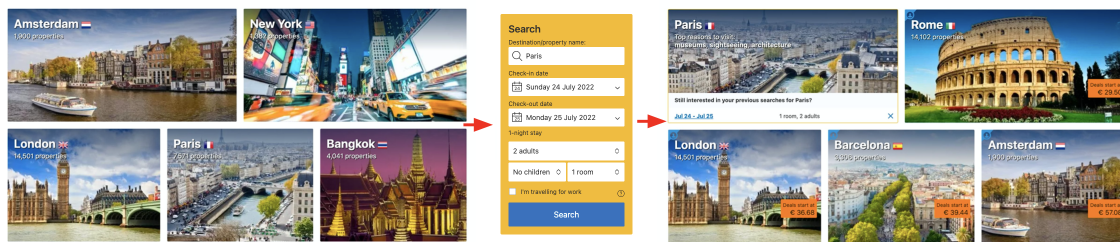


Fig. 7. An example flow of recommendations refinement after a single customer search interaction with the website

UI optimizations, moreover, can more seamlessly integrate recommendations with other useful tools to assist with where the customer is in their planning process. As an example, Booking.com ties destination recommendations to recent searches, allowing users to easily either revisit a previously search or explore another recommended destination.

Booking.com also modifies the displayed destination recommendation in search results, depending on the level of a search and the number of results available for the search. For customers predicted to still be making destination-level decisions, as indicated by searching at the region level, the layout of the page gives higher priority to the destination

<sup>3</sup> Airport coverage areas can often overlap (a single travel destination can be served by multiple "competing" airports). This means that in this context a travel destination can be included in more than one recommendation classes.

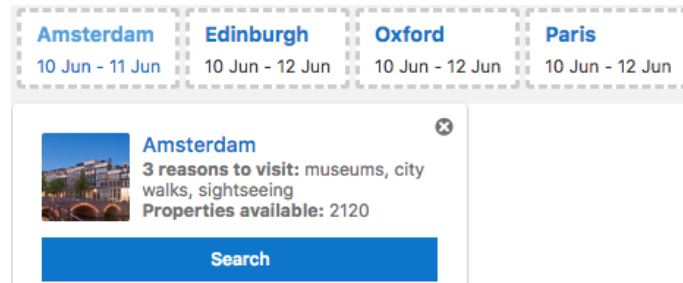


Fig. 8. Additional destination information within a recommendation element

recommendations by putting them on top. On the other hand, customers looking for a defined city will see destination recommendations at the bottom of the page.

As mentioned in subsubsection 2.3.2, what complementary information displayed with the destination recommendation is important, especially for customers looking for suitable alternatives or for next destination to extend their trip. Figure 8 shows an enhanced version of destination recommendation. Adding relevant information about the recommended destination (such as the main reasons to visit or number of accommodations available) and providing a direct link to view accommodations in a recommended destination brought significant improvement on business metrics. Destination recommendations also change over time in their relevance, and so recommender systems also need to consider how and when to renew and to discard recommendations for instance after the customer searched or booked the recommended destination. Here, there is also rationale for providing users a direct interaction with recommendations, such as by allowing them to discard inadequate recommendations.

### 3.2.2 Leveraging Themes to Recommend Destinations.

In some instances, an initial user input can help to provide a better destination recommendation. At Booking.com, we've found this to be particularly useful for customers who know a region they'd like to visit, but not specific towns or cities within that region. By utilizing reviews and other user generated data, together with accommodations and room amenities to score and match destinations and accommodations, we enabled customers to narrow their search

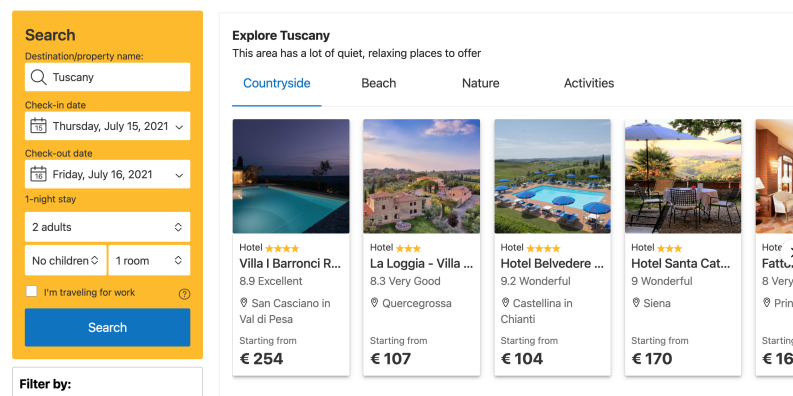


Fig. 9. Example of thematic trip recommendations for Tuscany.



according to themes, such as beach, nature, spa or relaxation, as displayed in Figure 9. The applied themes were selected by relying on user research and data analysis to understand growing trends. Available themes can grow and/or adapt as customers behavior and preferences continue to be analytically observed - as was relevant during the COVID-19 period. Personalizing the themes and/or the destinations within, generates a unique recommendation per each customer. This can also consider each traveller’s unique needs, as discussed in subsubsection 2.1.3.

### 3.3 Modeling Approaches

While many of the challenges in destination recommendation are addressed by designing the right user experience and better understanding of the recommended items, some of the patterns in the travel data can be addressed with unique modeling techniques. Here we present how adaptive online models and sequential learning models help to provide more relevant recommendations.

**3.3.1 Multi-Armed Bandits Modeling.** Recommendations can be naturally formulated as a multi-armed bandits problem [10, 46]. These algorithms balance between exploration of items with low confidence estimates and exploitation of items with high confidence, with user feedback constantly fed into the model and updating the model parameters. In destination recommendation, we select an arm (destination) and observe its specified reward. A main advantage of this approach over models built on historical data, is that the latter tend to recommend very popular destinations and thus suffer from limited diversity. Moreover, these models are trained on historical reservation data which are not directly caused by our interventions and may introduce bias. Setting up the framework for a successful bandit is not trivial and at Booking.com we have deployed bandits in several parts of our business [3]. Here we list key challenges that we identified in the context of destination recommendation in email marketing campaigns:

- **Reward definition.** Our goal is to help subscribers complete their booking and therefore, the obvious reward for this bandit is the booking itself. However, attributing which bookings are a result of our intervention is not trivial. The reason for this is that the customer journey is not always linear, i.e. *[Search → Open email → Click on recommendation → Book recommendation]*. In fact, our subscribers might open the email, see the recommendations and complete their booking without landing on the website by clicking in any of them. Furthermore, they might click on a recommendation that they like, start browsing and then book another destination.
- **Delayed feedback.** In many cases the feedback/reward is not (or cannot be) reported immediately [39]. The subscribers don’t always interact with our emails immediately after receiving them.
- **Off-policy updates.** As a result of the delayed feedback, online learning is infeasible and thus, off-policy learning techniques have to be employed.
- **Multiple recommendations - Slates.** Recommendations usually come as slates [18] instead of single items. We recommend hundreds of thousands destinations and consequently, the set of possible slate is enormous.
- **Productionization.** Sending millions of emails means that we have to get millions of predictions from our models, which makes efficiency a critical factor. In stochastic policies, we have to add another layer of complexity which comes from sampling (without replacement) from the output of our models.

Our email recommendations are built using an internal feedback-loop bandits platform [3], supporting large scale, distributed online learning. The architecture of our recommendation platform is shown in Figure 10. This platform (predictions and feedback tracking box in Figure 10) keeps track of all user interaction data, such as the destinations in the emails (recommendation model’s predictions), the features used to get the predictions, and the interactions with each recommended item/destination. Our model is based on the top-k off-policy approach presented in [11], tailored to

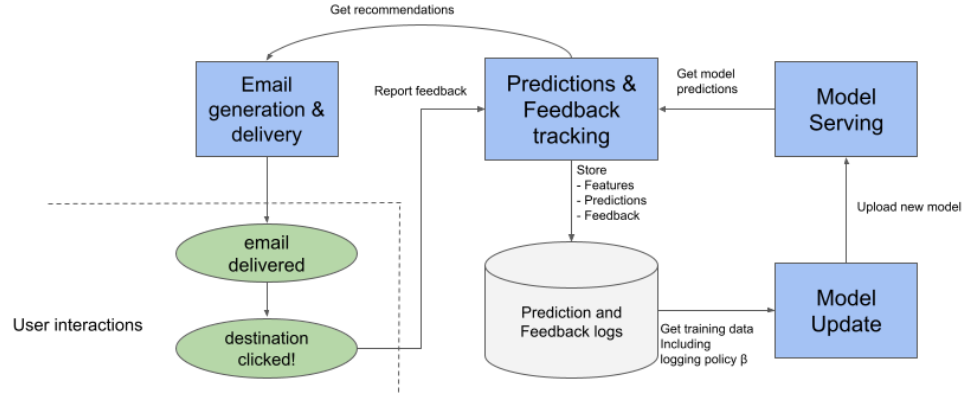


Fig. 10. Recommendation platform architecture

our business needs. We selected clicks to be the reward which, although a proxy to our original goal, is more direct and less sparse than bookings. Clicks are collected over a period of three days after sending each email.

Given a set of features  $x \in R^d$  for a subscriber, the model and policy produce a probability distribution over the  $N$  possible destinations (arms)  $a_i$ ,  $p(a_1, a_2, \dots, a_N|x)$ , which in our case is DNN with a softmax output. Similar to [11], we used Boltzmann exploration in order to maintain good training data for our policy. In order to generate  $k$  recommendation, one has to sample  $K \gg k$  items from a categorical distribution parametrized by the model's output  $p(\cdot|x)$ , and then filter out duplicate destinations. In order to avoid sampling with replacement, we employed the Topk-Gumbel-Trick [34]. Using this trick we add Gumbel noise to the log probabilities and select the top-k outputs, i.e.  $\text{argtop}_k \{\log(p(\cdot|x)) + g \sim \mathcal{G}(0, 1)\}$ . This modeling led us to improve the efficiency of our recommendation engine. Our model is periodically updated using the reward received over the last few days. As mentioned before, the feedback is delayed, which means that the policy which is optimised in every iteration might deviate from the *logging* policy, which produced the feedback used for training in the iteration in hand. Similar to the original paper [11], we use inverse propensity scoring, where the propensities are stored along with the feedback by the system.

### 3.3.2 Sequence Modeling.

Many types of recommendations problems can be naturally viewed as sequence extension problems. A classic example is a language model offering real-time next word recommendations while typing. Similarly, when recommending a travel destination, it is reasonable to expect that not only the set of previously booked destinations, but also their exact order would be a useful input for making the next destination prediction (see subsection 2.1.2). One of the recommendation problems that we try to solve, called the multi-destination trips problem, is when we have information about the beginning of the trip and we want to suggest a trip extension (Figure 5). Recurrent neural networks (RNNs) including their gated variants (e.g. LSTM, GRU) are commonly used for such sequence completion tasks. In general, sequential models have rapidly adopted in the field of recommendation systems [12, 36, 42, 47].

The problem becomes more complex if we want to personalize our recommendations. For example, if we want to recommend where to go next after visiting Amsterdam and London, the recommendation for a user from Japan might differ from a user from Germany. Recent studies show how to use contextual data in a Recurrent Neural Network (RNN) based recommender system [6, 40] and combine the recurrent part within a complex architecture. Going back to our

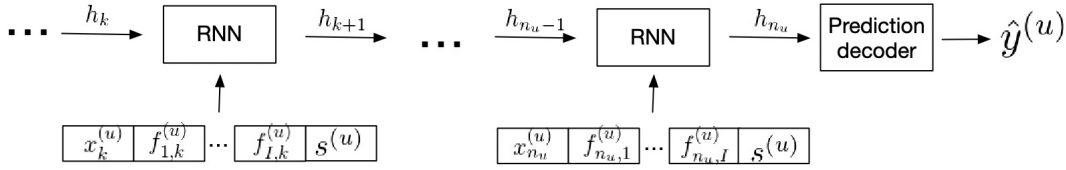


Fig. 11. Concatenating features embedding to the main token embedding

multi-destination problem, in order to get a better recommendation, in addition to past bookings, we can use features such as user country, lengths of stay, season or accommodation type. A popular and straightforward approach for merging item and features embeddings is simply by concatenating them (as shown in Figure 11).

In a general, "where to go next" recommendation, which is not necessarily within a trip, we may use any user's historical interactions, including searches and his bookings. To distinguish between searches and bookings, we concatenate an identifier feature to each destination embedding before passing it to the network. Another important feature we add, similar to *Latent Cross* [6], is the days difference between each of the items that a user has searched or booked. We observed that, a recommendation for a user that lastly booked only a few months ago would be more generic and "popular", while a user with an up-to-date booking would be recommended a destination nearby the last booking.

Another useful approach is the *Wide and Deep* framework [13]. It allows a joint training of wide linear models and deep neural networks. The framework combines the benefits of memorisation and generalisation for recommender

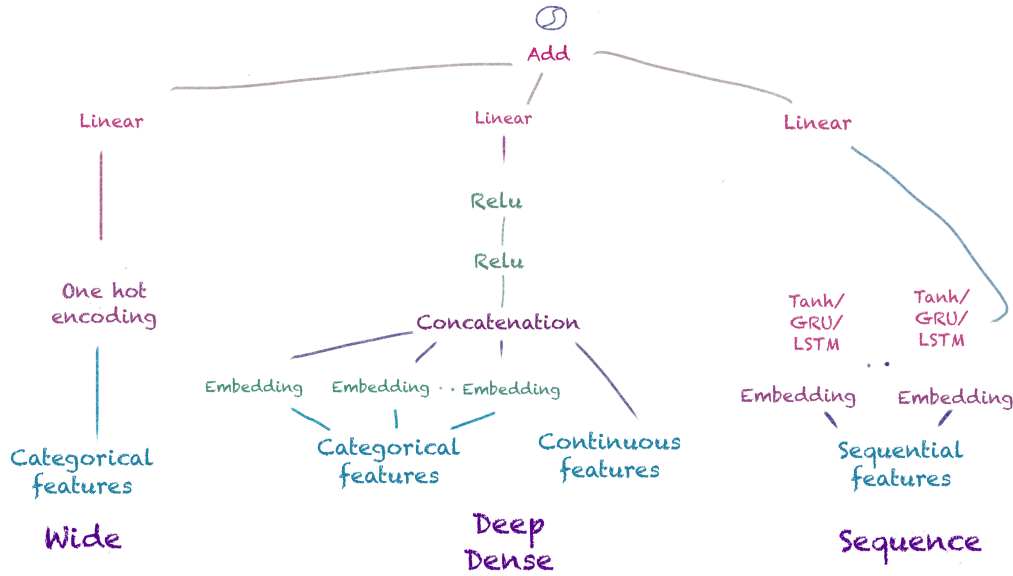


Fig. 12. Wide and Deep Architecture

systems. Our architecture (Figure 12) includes three main components: wide, deep and sequential (contextual RNN). However, modeling the problem with a generic, deep-learning solution, bring additional challenges:

- **High cardinality.** In Booking.com we recommend hundreds of thousands of destinations. To deal with the high cardinality space of the destinations, we map the long tail low-frequency destinations to an *unknown token* during the training. Then during prediction, the unknown destinations are replaced in an offline manner, based on geographical proximity.
- **Different space of recommendations.** Some situations require predictions in an output space which is different from the input space, we may want to recommend a country to visit next based on the user's past history of accommodation bookings. We may instead want to recommend other products of connected trip (i.e attractions, flights, cars rental and taxis) based on past accommodation bookings. To address this, we set an RNN Encoder→Decoder architecture, where the decoder can be set as simple as softmax regression in case of a categorical prediction [6, 40].
- **Not chronological sequences.** Some customers perform reservations in an order that differs from the chronological order of the stays, which opens up a new challenge in the form of "gaps-filling". The missing part of the trip is surrounded by two existing bookings, such as in Figure 5. We use RNN Encoder→Decoder, where the sequence encoded with a mask on the gap, and the decoder filling-in the missing destination.

A recent Booking.com contest [25] conducted on the dataset of the multi-destination trip problem [26] provided benchmark results for the recommendation problem and opened up an opportunity to explore various state-of-the-art Machine Learning techniques. The best-performing team [45] achieved a 0.5939 score, providing a 30% improvement compared to the median result. The solution put to use a blend of three different neural network architectures, using XLNet Transformers [48], Gated Recurrent Units [14], and feed-forward networks with a shared Session-based Matrix Factorization head (SMF). The authors discuss the superior performance of deep-learning based algorithms in the challenge, compared to classical recommender approaches relying on their past experience [31]. The runner-up solution is based on an Efficient Manifold Density Estimator (EMDE) for the recommendation task [16]. Other leading solutions used various deep-learning and recommender techniques, including Transformers, LSTM networks [30, 49], LambdaRank [41, 43], and further state-of-the-art recommender methods.

#### 4 CONCLUSION

In this work, we presented a set of unique challenges in destination recommender systems, displaying the underlying need for domain-specific approaches to resolve these challenges. These challenges are in the areas of both customer context and unique data, modeling, and measurement properties. While the destination recommender systems can benefit from the state-of-the-art recommender methods, it further requires unique and creative approaches for tailor-made solutions. Resolving these challenges therefore calls for adaptation of classical recommender methods. The lessons and solutions described here provide a first layer of applied insights for implementing destination recommendations in real-life systems and lay a groundwork for further research in the field.

#### ACKNOWLEDGMENTS

We thank Thomas Bosman, Klajdi Qirko and Jake Mooney for their help in editing and reviewing this paper; and to the UX Research, Product, Development and Design teams at Booking.com who contributed to bringing those ideas to life.

## REFERENCES

- [1] Elise Acheson, Michele Volpi, and Ross S. Purves. 2020. Machine learning for cross-gazetteer matching of natural features. *International Journal of Geographical Information Science* 34, 4 (2020), 708–734. <https://doi.org/10.1080/13658816.2019.1599123>
- [2] Ahmad Alshuqaiqi and Shida Irwana Omar. 2019. Causes and Implication of Seasonality in Tourism. *Journal of Advanced Research in Dynamical and Control Systems* 11 (03 2019), 1480–1486.
- [3] Lucas Bernardi, Pablo Estevez, Matias Eidis, and Eqlal Osama. 2020. Recommending Accommodation Filters with Online Learning. In *Proceedings of the Workshop on Online Recommender Systems and User Modeling (ORSUM @ RecSys 2021)*.
- [4] Lucas Bernardi, Jaap Kamps, Julia Kiseleva, and Melanie JI Müller. 2015. The Continuous Cold Start Problem in e-Commerce Recommender Systems.
- [5] Lucas Bernardi, Themistoklis Mavridis, and Pablo Estevez. 2019. 150 Successful Machine Learning Models: 6 Lessons Learned at Booking.com. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1743–1751.
- [6] Alex Beutel, Paul Covington, Sagar Jain, Can Xu, Jia Li, Vince Gatto, and Ed H Chi. 2018. Latent cross: Making use of context in recurrent recommender systems. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 46–54.
- [7] B Bondaruk, SA Roberts, and C Robertson. 2019. Discrete Global Grid Systems: Operational Capability of the Current State of the Art. In *Proceedings of the 7th Conference on Spatial Knowledge and Information Canada (SKI2019)*, Vol. 2323.
- [8] Booking.com. 2021. *New research reveals an increased desire to travel more sustainably*. <https://partner.booking.com/en-us/click-magazine/new-research-reveals-increased-desire-travel-more-sustainably> Accessed: 2021-06-30.
- [9] Luchiana C. Brodeala. 2020. Online Recommender System for Accessible Tourism Destinations. In *Fourteenth ACM Conference on Recommender Systems (Virtual Event, Brazil) (RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 787–791. <https://doi.org/10.1145/3383313.3411450>
- [10] Olivier Chapelle and Lihong Li. 2011. An Empirical Evaluation of Thompson Sampling. In *Proceedings of the 24th International Conference on Neural Information Processing Systems (Granada, Spain) (NIPS'11)*. Curran Associates Inc., USA, 2249–2257. <http://dl.acm.org/citation.cfm?id=2986459.2986710>
- [11] Minmin Chen, Alex Beutel, Paul Covington, Sagar Jain, Francois Belletti, and Ed H. Chi. 2019. Top-K Off-Policy Correction for a REINFORCE Recommender System. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (Melbourne VIC, Australia) (WSDM '19)*. Association for Computing Machinery, New York, NY, USA, 456–464. <https://doi.org/10.1145/3289600.3290999>
- [12] Qiwei Chen, Huan Zhao, Wei Li, Pipei Huang, and Wenwu Ou. 2019. Behavior sequence transformer for e-commerce recommendation in alibaba. In *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data*. 1–4.
- [13] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishikesh Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*. 7–10.
- [14] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
- [15] Amine Dadoun, Raphaël Troncy, Olivier Ratier, and Riccardo Petitti. 2019. Location Embeddings for Next Trip Recommendation. In *Companion Proceedings of The 2019 World Wide Web Conference (San Francisco, USA) (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 896–903. <https://doi.org/10.1145/3308560.3316535>
- [16] Michał Daniluk, Barbara Rychalska, Konrad Goluchowski, and Jacek Dąbrowski. 2021. Modeling Multi-Destination Trips with Sketch-Based Model. In *Proceedings of the ACM WSDM Workshop on Web Tourism (WSDM WebTour'21)*.
- [17] Manoj Reddy Dareddy. 2016. Challenges in Recommender Systems for Tourism. In *Proceedings of ACM RecSys Workshop on Recommenders in Tourism (RecTour 2016)* (Boston, MA, USA).
- [18] Maria Dimakopoulou, Nikos Vlassis, and Tony Jebara. 2019. Marginal Posterior Sampling for Slate Bandits.. In *IJCAI*. 2223–2229.
- [19] Thi Hai Ninh Do and Wurong Shih. 2016. Destination Decision-Making Process Based on a Hybrid MCDM Model Combining DEMATEL and ANP: The Case of Vietnam as a Destination. *Modern Economy* 7 (2016), 966–983.
- [20] Ercan Ezin, Iván Palomares, and James Neve. 2019. Group Decision Making with Collaborative-Filtering ‘in the loop’: interaction-based preference and trust elicitation. (2019), 4044–4049. <https://doi.org/10.1109/SMC.2019.8914224>
- [21] A Stewart Fotheringham and Morton E O’Kelly. 1989. *Spatial interaction models: formulations and applications*. Vol. 1. Kluwer Academic Publishers Dordrecht.
- [22] Dmitri Goldenberg. 2021. Putting the Role of Personalization into Context. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*. <https://doi.org/10.1145/3404835.3464929>
- [23] Dmitri Goldenberg. 2021. Putting the Role of Personalization into Context. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2631–2632.
- [24] Dmitri Goldenberg, Kostia Kofman, Javier Albert, Sarai Mizrahi, Adam Horowitz, and Irene Teinemaa. 2021. Personalization in Practice: Methods and Applications. In *Proceedings of the 14th International Conference on Web Search and Data Mining*.
- [25] Dmitri Goldenberg, Kostia Kofman, Pavel Levin, Sarai Mizrahi, Maayan Kafry, and Guy Nadav. 2021. Booking.com WSDM WebTour 2021 Challenge. <http://www.bookingchallenge.com>. In *Proceedings of the ACM WSDM Workshop on Web Tourism (WSDM WebTour '21)*.
- [26] Dmitri Goldenberg and Pavel Levin. 2021. Booking.com Multi-Destination Trips Dataset. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*. <https://doi.org/10.1145/3404835.3463240>

- [27] Dmitri Goldenberg, Guy Tsyse, Igor Spivak, Javier Albert, and Amir Tzur. 2021. Learning to Persist: Exploring the Tradeoff Between Model Optimization and Experience Consistency. In *Companion Proceedings of the Web Conference 2021*. Association for Computing Machinery, New York, NY, USA, 527–529. <https://doi.org/10.1145/3442442.3452051>
- [28] M. F. Goodchild and L. L. Hill. 2008. Introduction to digital gazetteer research. *International Journal of Geographical Information Science* 22, 10 (2008), 1039–1044. <https://doi.org/10.1080/13658810701850497>
- [29] Kingsley E Haynes and A Stewart Fotheringham. 1985. *Gravity and spatial interaction models*. Regional Research Institute, West Virginia University. Reprint. Edited by Grant Ian Thrall. WVU Research Repository, 2020.
- [30] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [31] Dietmar Jannach, Gabriel de Souza P. Moreira, and Even Oldridge. 2020. Why Are Deep Learning Models Not Consistently Winning Recommender Systems Competitions Yet? A Position Paper. In *Proceedings of the Recommender Systems Challenge 2020*. 44–49.
- [32] Jameel Khadaroo and Boopen Seetanah. 2008. The role of transport infrastructure in international tourism development: A gravity model approach. *Tourism management* 29, 5 (2008), 831–840.
- [33] Julia Kiseleva, Melanie JI Mueller, Lucas Bernardi, Chad Davis, Ivan Kovacek, Mats Stafseng Einarsen, Jaap Kamps, Alexander Tuzhilin, and Djoerd Hiemstra. 2015. Where to go on your next trip? Optimizing travel destinations based on user preferences. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1097–1100.
- [34] Wouter Kool, Herke van Hoof, and Max Welling. 2020. Estimating Gradients for Discrete Random Variables by Sampling without Replacement. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020*. OpenReview.net. <https://openreview.net/forum?id=rkJ2EFvB>
- [35] Dewi Ayu Kusumaningrum and Suci Sandi Wachyuni. 2020. The shifting trends in travelling after the COVID 19 pandemic. *International Journal of Tourism & Hospitality Review* (2020).
- [36] Tobias Lang and Matthias Rettenmeier. 2017. Understanding consumer behavior with recurrent neural networks. In *Workshop on Machine Learning Methods for Recommender Systems*.
- [37] Neal Lathia. 2017. Bootstrapping a Destination Recommender System. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (Como, Italy) (RecSys '17)*. Association for Computing Machinery, New York, NY, USA, 341. <https://doi.org/10.1145/3109859.3109924>
- [38] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4, Part 2 (2014), 2065–2073. <https://doi.org/10.1016/j.eswa.2013.09.005>
- [39] Themis Mavridis, Soraya Hausl, Andrew Mende, and Roberto Pagano. 2020. Beyond algorithms: Ranking at scale at Booking. com. In *Proceedings of the Fourth Workshop on Recommendation in Complex Scenarios. CEUR-WS (Virtual Event, Brazil)*.
- [40] Sarai Mizrahi and Pavel Levin. 2019. Combining Context Features in Sequence-Aware Recommender Systems.. In *Late-Breaking Results of the 13th ACM Conference on Recommender Systems (Copenhagen, Denmark) (RecSys '19)*.
- [41] Aleksandr Petrov and Yuriy Makarov. 2021. Attention-based neural re-ranking approach for next city in trip recommendations. In *Proceedings of the ACM WSDM Workshop on Web Tourism (WSDM Webtour '21)*.
- [42] Massimo Quadrana, Paolo Cremonesi, and Dietmar Jannach. 2018. Sequence-aware recommender systems. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 1–36.
- [43] C Quoc and Viet Le. 2007. Learning to rank with nonsmooth cost functions. *Proceedings of the Advances in Neural Information Processing Systems 19* (2007), 193–200.
- [44] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. 2002. Methods and Metrics for Cold-Start Recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (Tampere, Finland) (SIGIR '02)*. Association for Computing Machinery, New York, NY, USA, 253–260. <https://doi.org/10.1145/564376.564421>
- [45] Benedikt Schifferer, Chris Deotte, Jean-Francois Puget, Gabriel de Souza Pereira Moreira, Gilberto Titericz, Jiwei Liu, and Ronay Ak. 2021. Using Deep Learning to Win the Booking.com WSDM WebTour21 Challenge on Sequential Recommendations. In *Proceedings of the ACM WSDM Workshop on Web Tourism (WSDM Webtour '21)*.
- [46] Aleksandrs Slivkins. 2019. Introduction to Multi-Armed Bandits. *CoRR abs/1904.07272* (2019). arXiv:1904.07272 <http://arxiv.org/abs/1904.07272>
- [47] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1441–1450.
- [48] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. <https://arxiv.org/pdf/1706.03762.pdf>
- [49] Yuanzhe Zhou, Shikang Wu, and Chenyang Zheng. 2021. Explore next destination prediction. In *Proceedings of the ACM WSDM Workshop on Web Tourism (WSDM Webtour '21)*.