Exploiting Socio-Economic Models for Lodging Recommendation in the Sharing Economy

Raul Sanchez-Vazquez
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil
raulsanchez@dcc.ufmg.br

Jordan Silva
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil
jordan@dcc.ufmg.br

Rodrygo L. T. Santos
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil
rodrygo@dcc.ufmg.br

ABSTRACT

Recent years have witnessed the emergence of sharing economy marketplaces, which enable users to share goods and services in a peer-to-peer fashion. A prominent example in the travel industry is Airbnb, which connects guests with hosts, allowing both to exchange cultural experiences in addition to the economic transaction. Nonetheless, Airbnb guest profiles are typically sparse, which limits the applicability of traditional lodging recommendation approaches. Inspired by recent socio-economic analyses of repurchase intent behavior on Airbnb, we propose a context-aware learning-to-rank approach for lodging recommendation, aimed to infer the user's perception of several dimensions involved in choosing which lodging to book. In particular, we devise features aimed to capture the user's price sensitivity as well as their perceived value of a particular lodging, the risk involved in choosing it rather than other available options, the authenticity of the cultural experience it could provide, and its overall perception by other users through word of mouth. Through a comprehensive evaluation using publicly available Airbnb data, we demonstrate the effectiveness of our proposed approach compared to a number of alternative recommendation baselines, including a simulation of Airbnb's own recommender.

KEYWORDS

Lodging Recommendation; Context Awareness; Learning to Rank; Consumption Behavior; The Sharing Economy

1 INTRODUCTION

The travel industry has largely adopted the Internet as one of its main sales channels, helping customers find information in order to plan their trip [19]. As a result, specialized e-tourist agencies evolved and became popular services that simplified the travel planning process [2, 5], helping travelers with personalized assistance. Naturally, recommender systems (RS) have also been proposed to tackle problems in the travel domain [3, 18], allowing users to have a customized interaction with online platforms. Recently, the travel industry has been reshaped by the so-called sharing economy, which introduced online marketplaces with innovative consumption modalities that stand in sharp contrast to their traditional counterparts [1, 12, 15]. Such consumption modalities are characterized by user-to-user transactions in a peer-to-peer (P2P) fashion, with the online marketplace serving as a mediator [15].

Airbnb[1] is a prominent representative of the sharing economy and the largest P2P lodging provider, allowing users to list, search, and rent lodgings, enabling guests to benefit from the advice of local hosts. Researches have studied P2P lodging and found important peculiarities that substantially distinguish it from a traditional hotel lodging scenario [15, 22, 25, 39]. In particular, P2P lodging has been described as more than just a hotel substitute for three main reasons: (1) P2P lodging provides a much more dynamic ecosystem as lodging supply can rapidly respond to changes in demand, (2) P2P lodging serves a wider range of use cases due to the increased diversity and geographical coverage of lodging supply, (3) P2P lodging customers give importance to cost savings, utility, trust, and familiarity, shaping unique customer preferences. At the same time, as illustrated in Figure 1 for the Airbnb dataset used in our experiments, guest profiles are typically small, with 80% of the users having fewer than 5 bookings in their profile. As a result, the user-lodging matrix is extremely sparse (99.9997%), which hinders an accurate modeling of users' distinctive preferences.

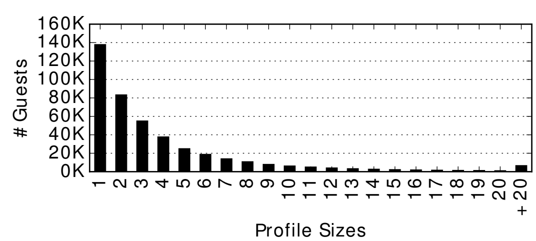

Figure 1: Profile size distribution of Airbnb users.

Inspired by recent socio-economic analyses of consumption behavior on Airbnb, we propose a context-aware learning-to-rank approach for lodging recommendation. Our approach tackles the aforementioned challenges by modeling several contextual dimensions associated with guests' decision of which lodgings to book. In particular, we devise features to capture guests' price sensitivity as well as their perceived value of a particular lodging option, the risk involved in choosing it rather than other available options, the

[1] https://www.airbnb.com/

RecSys '17, August 27–31, 2017, Como, Italy.

DOI: http://dx.doi.org/10.1145/3107411.3107471

authenticity of the experience it could provide, and its overall perception by online guests' word of mouth. Through a comprehensive evaluation using publicly available data from Airbnb, we demonstrate the effectiveness and robustness of our approach in contrast to effective baselines from the literature as well as a simulation of Airbnb's own recommender. Moreover, through an ablation study, we provide further evidence to support or refute existing theories of lodging consumption behavior in the sharing economy. To the best of our knowledge, our approach is the first to address lodging recommendation in a sharing economy marketplace.

Our main contributions are three-fold:

(1) A novel lodging recommendation approach that models the socio-economic context around available lodgings as contextual features for learning to rank.
(2) A test collection and evaluation methodology for lodging recommendation in the sharing economy.
(3) A comprehensive evaluation of the proposed approach in a case study using Airbnb.

In the remainder of this paper, Section 2 discusses related work on lodging recommendation. Section 3 details our proposed context-aware learning-to-rank approach for tackling this problem. Sections 4 and 5 describe the experimental setup and the results of the empirical evaluation of our approach. Lastly, Section 6 provides our concluding remarks and directions for future work.

2 RELATED WORK

According to surveys in RS [3, 11, 18], there are multiple works aiming to assist tourists and travelers in their planning process. Concretely, these different approaches recommend one or more of the following items: travel destinations, touristic attractions, recreational activities, accommodations, and eatery venues. The majority of these works recommends lodgings in conjunction with (or constrained to) other items, leaving it unclear how to apply them to pure lodging recommendation. While no previous work has tackled lodging recommendation in the sharing economy to the best of our knowledge, in the following, we describe related approaches for the more traditional problem of hotel recommendation.

Saga et al. [31] proposed to use an undirected hotel-guest graph obtained from booking transactions. The graph is converted into a preference transition network, i.e., a graph G with hotels as nodes, and with two hotel nodes h_i and h_j having an edge between them if users that booked h_i are likely to book h_j. Recommendation is performed in two stages. First, the user explicitly selects an initial hotel h to produce a list of neighboring nodes as candidates. Then, each candidate hotel h_i is scored by $f(h_i) = in_i - out_i - C^2$, where in_i and out_i are the in- and out-degree of the candidate hotel h_i in G, whereas the factor C quantifies previous recommendations of the same hotel, in order to promote novelty in the ranking.

Levi et al. [21] characterized the hotel domain as a cold-start scenario. They assumed that users give more importance to reviews written by similar users than reviews written by users with different background, where similarity is defined in terms of nationality, travel intention (single, couple, family, group, and business), and preferences on hotel traits (location, service, food, room, etc.). In their work, hotels are modeled as feature vectors using the words (features) contained in their reviews. In addition, they extract features for each nationality, travel intention, and hotel trait. Hotel recommendation is performed by aggregating the importance assigned to each feature by users with similar preferences as the target user. A crowdsourcing experiment using Amazon Mechanical Turk showed evidence of the usefulness of mining reviews to identify typical types of user for any given hotel.

Zhang et al. [40] proposed a hybrid approach to hotel recommendation. By modeling users and hotels in latent topic spaces, they produce user and hotel similarity matrices. These matrices are then leveraged when factorizing the user-hotel rating matrix. Data normalization is performed by taking into account each user's traveling mode (single, couple, group, family, business, and others). The final step produces recommendation using maximal marginal relevance [7], a diversification technique, which reduces redundancy while maintaining relevance. Relatedly, PCA-ANFIS [26] tackles recommendation when hotels possess ratings in various aspects (e.g. value, rooms, location, cleanliness, check in/front desk, service, and business services). Such ratings are used to create a 3-dimension tensor, corresponding to: users, items, and the multiple ratings. The tensor is used to cluster users and for each individual cluster dimensionality reduction is achieved. A second step consists of training multiple ANFIS, a neural network that integrates fuzzy logic principles, to predict overall ratings in each cluster. Their experiments demonstrated that PCA-ANFIS leads to improvements in predictive accuracy of tourism multi-criteria prediction.

3 CONTEXTUAL SOCIO-ECONOMIC MODELS FOR LODGING RECOMMENDATION

Given a user u and a target location l, a lodging recommender must rank the lodgings $\mathcal{I}$ in the surroundings of l, prioritizing those where u would like to sojourn. Recommender systems typically leverage information on users, items, or interactions between users and items. As illustrated in Figure 1, Airbnb guest profiles are severely sparse, which limits the applicability of traditional collaborative or content-based approaches. On the other hand, as discussed in Section 2, existing hotel recommendation approaches ignore the economic drivers that motivate users to consume lodgings in the sharing economy. In order to overcome these challenges, we propose a context-aware learning-to-rank approach for lodging recommendation, aimed to exploit the socio-economic context around available lodgings as multiple ranking features.

Our approach is inspired by recent socio-economic studies in the domain of Airbnb. In particular, Liang [22] analyzed five aspects of users' recurrent consumption of Airbnb lodgings:

(1) **Perceived Value**, the trade-off between the benefits versus the cost of each available lodging;
(2) **Perceived Risk**, the assessment of all possible negative outcomes derived from booking the lodging;
(3) **Price Sensitivity**, the extent to which the price of a lodging affects a guest's booking behavior;
(4) **Perceived Authenticity**, the extent to which a guest feels like natively living at the lodging place; and
(5) **Electronic-Word-of-Mouth**, informal opinions that frame the judgment of other users towards the lodging.

In the following, we model each of these consumption behavior aspects as a preference dimension represented by multiple ranking features. Table 1 lists all 176 features[2] used to represent each candidate lodging in our approach, which capture the socio-economic context of an available lodging in different ways.

Table 1: Lodging recommendation features. The "input" column denotes whether each feature is estimated based upon candidate item *i*, item context *c*, or both.

feature class	input	qty
Perceived Value (PV)		
Pricing	*i*	6
Property type	*i*	21
Room type	*i*	3
Bed type	*i*	5
Equipments	*i*	4
Property capacity	*i*	1
Guests allowed	*i*	1
Amenities	*i*	40
Nearby venues	*c*	3
Nearby venues check-ins (min, max, avg, med)	*c*	12
Nearby venues distance (min, max, avg, med, norm.)	*i*, *c*	24
Perceived Risk (PR)		
Cancellation policy	*i*	5
Ratings	*i*	7
Reviews (std, norm.)	*i*	2
Nearby lodgings	*c*	1
Nearby lodgings reviews (avg, std)	*c*	2
Price Sensitivity (PS)		
Histogram lodgings prices (avg, skw, kur)	*c*	3
Sampled lodgings prices (avg, skw, kur)	*c*	3
Price (normalized)	*i*, *c*	3
Perceived Authenticity (PA)		
Authenticity score (avg, med, min, max, skw, kur)	*i*	6
Electronic Word of Mouth (EWoM)		
Sentiment score (avg, med, min, max, skw, kur)	*i*	24
Grand total		176

3.1 Perceived Value (PV)

Perceived Value (PV) quantifies the cost-benefit trade-off of a lodging. In particular, the cost of a lodging is measured by multiple numeric features conveying different pricing policies (e.g., daily, weekly, monthly price) as well as special fees including a security deposit or a fee for external visitors. On the other hand, the benefit of a lodging is conveyed by categorical features such as property type (e.g., apartment, house, loft), room type (private, shared, entire home), bed type (airbed, couch, futon, real bed, pull-out sofa), numerical features such as property capacity and number of available equipments (e.g., bathrooms, bedrooms, beds), and indicator features, such as whether external guests are allowed or which amenities (e.g., cable tv, doorman, dryer) the lodging provides.

Many studies demonstrate the importance of points of interest for touristic involvement, which has been shown to influence the selection of travel destinations [14, 27, 35]. To further quantify the benefit offered by a lodging, we propose to characterize its surroundings, by leveraging data from a location-based social network. In particular, using the Foursquare API,[3] we obtain information about up to 50 nearby venues from three different Foursquare categories: food, art & entertainment, and travel & transportation. We delimit the sampling radius based on findings on human time and distance perception: 500 m for eatery venues (accessible by a 5 minutes walk [4, 36]) and 1.5 km for entertainment and travel venues (accessible by a 15 minutes car drive[4][5] [20, 33]).

In addition to the total number of nearby venues in each category, we summarize distributions of two properties of these venues: number of check-ins and distance from the candidate lodging. Each of these distributions is encoded into multiple numeric features using different summary statistics, such as minimum, maximum, average, and median. Finally, just as psychological maps shape geographical boundaries in people's minds [29], the perceived distance from a lodging to popular venues may differ from their actual physical distance. In order to quantify this intuition, we propose to normalize the distance to express that customers are more willing to travel longer distances in order to reach popular venues. We achieve such a normalization by dividing the distance to a venue by the number of check-ins that the venue possesses.

3.2 Perceived Risk (PR)

Previous studies highlighted Perceived Risk (PR) as one of the key factors that affect repurchase intention on Airbnb [22]. Following a classical risk taxonomy [17], users take a *performance risk* as they cannot experience a lodging prior to their arrival, a *financial risk* as they may incur in cancellation fees, and a *physical risk* as they rarely know the host. In many online environments, risk symptoms can be alleviated through different mechanisms. In our scenario, we propose to quantify a lodging's performance risk through numerical features quantifying its received ratings on several aspects (cleanliness, accuracy, value, check-in, location, communication, and overall experience). In turn, financial risk is conveyed by a categorical feature indicating the available booking cancellation policies. Lastly, physical risk is quantified by measuring the popularity of the lodging, as well as the average and standard deviation of the popularity of nearby lodgings. These summary statistics are further used to normalize each lodging's raw popularity.

3.3 Price Sensitivity (PS)

Price Sensitivity (PS) is arguably one of the most important factors, broadly and intuitively accepted as a decisive motivator in consumers' behavior [9, 10, 34]. Airbnb users may present a high degree of PS as choosing a lodging requires a vivid exploration of multiple price options. Indeed, it has been shown that PS increases the more users are aware of the price dispersion of a given product [9]. Therefore, to model this preference dimension, we propose to summarize the price distribution of all candidate lodgings in a target location. To this end, we consider two price distributions: a price histogram displayed on the Airbnb search interface (see

[2]See [32] for further details.

[3]https://developer.foursquare.com/

[4]https://bitre.gov.au/publications/2015/files/is_073

[5]http://www.census.gov/hhes/commuting

Figure 2) and a price distribution sampled from the set of candidate lodgings. To quantify a guest's perception of these price distributions, they are both summarized using multiple statistics, including average, skewness, and kurtosis. In addition, these distributions' average and standard deviation are used to standardize each lodging's raw price. Standardized prices can be interpreted as how cheaper or more expensive a lodging is compared to nearby alternatives.

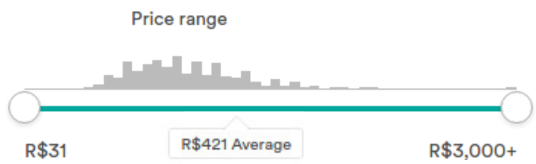

Figure 2: Airbnb price histogram.

3.4 Perceived Authenticity (PA)

Perceived Authenticity (PA) measures the extent to which a lodging offers a truly local living experience [22]. To quantify PA, we propose to mimic the way reviews are used by guests to sense authenticity as perceived (and reported) by other guests. To this end, we hand-build an authenticity lexicon comprising words that convey an authentic experience. The authenticity lexicon was built by reading several reviews and manually selecting words that usually appear when a user explicitly expresses a sense of PA. The lexicon can be seen in Table 2 and is composed of a total of 33 words.

Table 2: Handbuilt authenticity lexicon.

live	experience	authentic	share	truly	unique
real	recommend	welcome	talk	meet	friend
warm	community	explore	advice	connect	people
charm	communicate	discover	help	home	feel
place	neighborhood	genuine	time	chat	cozy
useful	hospitality	nearby			

An authenticity score for a review is estimated as the likelihood that it generates the authenticity lexicon. To this end, we use a standard language modeling approach with Dirichlet smoothing [28] after applying standard text processing techniques (lowercasing, stopword removal, Hunspell stemming). Finally, after producing a distribution of authenticity scores for the reviews associated with a lodging, we compute various authenticity scores for the lodging itself by using multiple summary statistics (average, median, minimum, maximum, skewness, and kurtosis).

3.5 Electronic Word of Mouth (EWoM)

In P2P lodging websites, reviews are the most common manifestation of Electronic Word of Mouth (EWoM) [22]. Indeed, guests are aware that their personal opinion on a lodging may help frame the perception of other users regarding the quality of this lodging [37], ultimately playing an key role in the booking decision process.

To take the pulse of Airbnb users towards individual lodgings, we propose to leverage standard sentiment analysis techniques. In particular, we use Vader [16], a state-of-the-art sentiment analysis tool, to extract four sentiment polarity scores from each review: positiveness, negativeness, neutrality, and a compound score. For a given lodging, EWoM scores are then produced by aggregating the various polarity scores associated with its reviews. Once again, we use multiple summary statistics to this end (average, median, minimum, maximum, skewness, and kurtosis).

3.6 Contextual Learning

Sections 3.1 through 3.5 presented multiple contextual models for lodging recommendation inspired by recently proposed socio-economic theories of consumption behavior on Airbnb. To further leverage these models for tackling the lodging recommendation problem, we resort to learning to rank [23]. In particular, our goal is to learn a ranking model $f : \mathcal{X} \rightarrow \mathcal{Y}$ mapping the input space $\mathcal{X}$ into the output space $\mathcal{Y}$. Our input space includes n learning instances $\{\vec{X}_j\}_{j=1}^{n}$, where $\vec{X}_j = \Phi(u_j, l_j, \mathcal{I}_j)$ is a feature matrix representation (produced by a feature extractor Φ) of a sample of lodgings $\mathcal{I}_j$ retrieved for user u_j near target location l_j. As described in Table 1, we consider a total of 176 features organized into five broad preference dimensions. In turn, our output space $\mathcal{Y}$ comprises n label vectors $\{\vec{Y}_j\}_{j=1}^{n}$, where $\vec{Y}_j$ provides relevance labels for each lodging $i \in \mathcal{I}_j$. To learn an effective ranking model f, we use LambdaMART [38], a gradient boosted regression tree learner, which represents the current state-of-the-art in learning to rank [8].

4 EXPERIMENTAL SETUP

This section describes the experimental setup that supports the evaluation of our context-aware learning to rank approach for lodging recommendation introduced in Section 3. In particular, we aim to answer the following research questions:

- *Q1.* How effective is our lodging recommendation approach?
- *Q2.* How do different features contribute to our approach?
- *Q3.* How do our results relate to lodging consumption theories?

In the following, we describe the test collection used in our experiments, the baselines used for comparison and the procedure undertaken to train and test them as well as our own models.

4.1 Test Collection

Our experiments are conducted using publicly available data collected between March and September 2016 from Airbnb for two target cities: New York City, USA (NYC) and London, UK (LON). The crawled dataset includes lodgings geographically spread over each city. For each lodging, we obtain a structured representation of its description, along with the entire history of reviews it received from guests up to the time of crawl. For each guest, we further obtain a sample of up to 10 of the most recent reviews received from hosts all over the world. Because hosts may have multiple lodgings on offer, we further crawl the list of lodgings offered by each host to reconcile the precise lodging in each guest's history. Table 3 provides summary statistics of the produced dataset, which is available upon request. Percentages are relative to Airbnb's claimed numbers of guests and lodgings in the world, except for the number of transactions, which is extrapolated proportionally to the average number of transactions per lodging in the dataset.

Table 3: Salient statistics of the recommendation dataset.

	NYC	LON	World	Airbnb
# Guests	219.9 k	223.1 k	9.25 M (31.9%)	60.0 M
# Lodgings	17.3 k	22.1 k	0.48 M (26.3%)	2.0 M
# Transactions	250.5 k	266.7 k	15.70 M (15.4%)	49.2 M

Our test collection comprises multiple test cases of the form $\langle u, l, t, \mathcal{I}, i^* \rangle$, where u is a target guest, l is the location where the guest wishes to sojourn, t is the time of the recommendation request, $\mathcal{I}$ is a set of candidate items within a radius of 2 km of l, and $i^* \in \mathcal{I}$ is the lodging originally booked by u, which should be promoted by a lodging recommender. In each city, we sample 1,500 test cases geographically distributed across the city. In order to simulate the exploratory nature of a guest seeking lodging in a given location l, we replicate each test case five times, varying the candidate set $\mathcal{I}$ each time within the original radius of 2 km. These five distinct test cases portray the guest u using an interactive map to browse and explore the listing supply around the neighborhood of l, where he or she wishes to sojourn. All five simulations are included in our evaluation as independent test cases, totaling $1,500 \times 5 = 7,500$ test cases per city, 15,000 test cases overall.

4.2 Training & Test Procedure

Sorting the test cases in our test collection by their booking date in ascending order, we split the test collection into 12 disjoint folds with roughly the same number of test cases each. As illustrated in Figure 3, evaluation is performed in a sliding window comprising five folds across eight evaluation rounds. In all rounds, the first four folds are used for training different models in a 4-fold cross validation, whereas the fifth fold is used for testing trained models. Hyperparameter tuning is performed via grid search using the training folds of the first round. Ranking effectiveness is measured using Mean Reciprocal Rank (MRR) on the test folds across all rounds. Statistical significance is verified using a paired t-test with $\alpha < 0.05$ with Bonferroni correction for multiple comparisons.

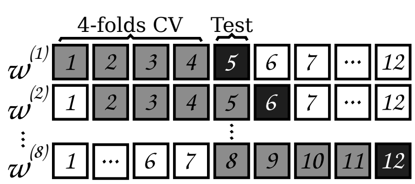

Figure 3: Sliding time windows.

4.3 Recommendation Baselines

We contrast our approach to three effective baselines:

(1) **Popularity** is a bias-centric model based on the lodging's booking count. Ties are solved by randomly sorting conflicting items [6].
(2) **Airbnb** consists of the non-personalized ranking produced by Airbnb's own recommender given the same input as all the other recommenders in our evaluation.
(3) **Bayesian Personalized Ranking Matrix Factorization** (BPRMF) [30] is a state-of-the-art matrix factorization approach for top-n recommendation.

5 RESULTS & DISCUSSIONS

In this section, we empirically evaluate our contextual learning-to-rank approach for lodging recommendation (CLLR, for short) in sharing economy marketplaces. In the following, we address each of the three research questions stated in Section 4 in turn.

5.1 Model Effectiveness

In this section, we address *Q1*, by analyzing the effectiveness of CLLR in contrast to the baselines introduced in Section 4.3: Popularity, BPRMF, and Airbnb. Table 4 shows MRR figures along with 95% confidence intervals for each baseline and our CLLR model. The lowest performance is attained by BPRMF, followed closely by the Popularity recommender, resulting in a statistical tie. These findings highlight the difficulties of collaborative recommenders to deal with the sparsity problem [31, 40]. Airbnb's recommender delivers a competitive performance, attaining the best MRR among the considered baselines. Finally, our CLLR model attains the highest MRR, significantly outperforming all three baselines, as denotes by the ▼ symbol alongside each baseline.

Table 4: MRR of CLLR vs multiple baselines.

	BPRMF	**Popularity**	**Airbnb**	**CLLR**
MRR	0.0215▼	0.0231▼	0.0328▼	**0.0400**
CI (95%)	0.0014	0.0015	0.0018	0.0021

To improve our understanding of the effectiveness of CLLR, Figure 4 shows MRR figures across all eight test folds at different points in time, with error bars denoting 95% confidence intervals for the means at the various points. For all but one test fold, CLLR consistently outperforms all baseline recommenders. Red stars indicate that CLLR performed statistically better ($\alpha = 0.05$) than all baselines, whereas blue stars denote statistical significance for all but one of Airbnb or Popularity ($\alpha = 0.05$). The only exception is the test fold of the fourth round, where Airbnb was the top performer.

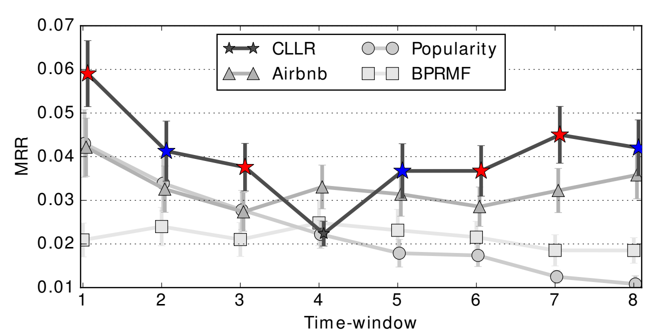

Figure 4: MRR by time window.

To better understand the drop in performance at the fourth round in Figure 4, we perform an ablation study, by discarding information that potentially misled CLLR into making wrong predictions, and hence causing the observed decay in performance. To this end, Figure 5 shows MRR figures attained by CLLR and five variants, each omitting one preference dimension at a time, namely, PV, PR, PS, PA, EWoM. Once again, error bars denotes 95% confidence intervals for the mean performance of each model at each point in time. As observed from the figure, at the fourth test round, removing the preference dimensions PS, PA, and PR (red markers) considerably improves the performance of the model. In particular, the fourth time window includes training data comprising bookings made between 10/2015-03/2016 and testing data comprising bookings made in 04/2016. This observation suggests that these dimensions may have hindered the model's generalization capabilities on this particular time window, perhaps influenced by a seasonal behavior (e.g., most of the training data are in the winter, whereas test data resides in spring) or an anomalous event (e.g., currency fluctuations). Nonetheless, further analysis is required to narrow down which subsets of features in the affected preference dimensions underperformed and what caused them to do so.

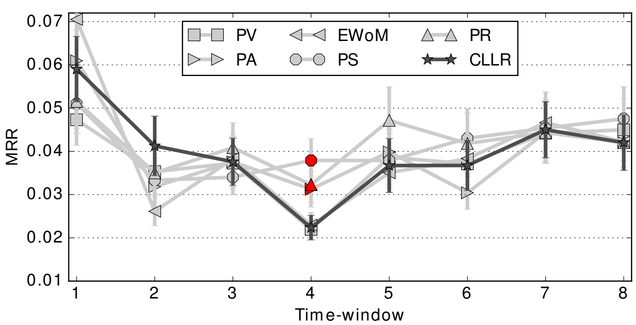

Figure 5: MRR by time window.

To further assess the robustness of CLLR in terms of the information it leverages, Table 5 shows consolidated MRR figures of the CLLR variants previously shown in Figure 5, however averaged across all eight test rounds. Intuitively, we would expect that the removal of certain preference dimensions would result in a significant loss in effectiveness. Nevertheless, none of the CLLR variants significantly differ from the complete model using all preference dimensions according to a two-tailed paired t-test with $\alpha < 0.05$. Recalling *Q1*, this result attests the effectiveness of CLLR and its robustness to noise due to eventual abnormalities in training.

Table 5: MRR after removing individual dimensions.

	PV	PR	PS	PA	EWoM
MRR	0.0383	0.0418	0.0412	0.0397	0.0400
CI (95%)	0.0020	0.0022	0.0021	0.0020	0.0020

5.2 Feature Efficiency

The previous section demonstrated the effectiveness of our proposed CLLR model in contrast to competitive lodging recommendation baselines as well as its robustness to the removal of entire preference dimensions. In this section, we address *Q2*, by assessing the contribution of each individual feature to the effectiveness of the model. To quantify feature importance, we use two complementary metrics. *Least Square Improvement* (LSI) [13, 24] is a feature importance score commonly used to measure the discriminative power of the features in boosting regression tree learners, such as LambdaMART. LSI employs the training instances that led to the creation of the model to quantify the extent to which each feature is trusted to distinguish between relevant and non-relevant instances at different nodes in the tree. While LSI measures the discriminative capabilities of each feature, we propose *Empirical Feature Efficiency* (EFE), a simple metric aimed to measure the impact in effectiveness caused by removing the feature in an ablation study. Precisely, EFE measures the relative gain (or loss) in effectiveness (in our case, given by MRR) brought by a given feature when added to a model that lacked that very same feature. A positive EFE denotes an important feature, whereas a negative EFE denotes a noisy feature. An EFE close to zero denotes an unimportant feature.

Table 6 shows EFE/LSI scores averaged across the time windows for features which obtained a statistically significance EFE (∘ indicates a significance level $\alpha \leq 0.05$, whereas • indicates $\alpha \leq 0.01$). Whenever the EFE is statistically significant and both EFE and LSI lead to the same conclusion, we consider the corresponding feature as being important. In the following, we discuss our observations for features in different preference dimensions.

Perceived Risk. This set of features models users' perception of risk, and is composed of lodging attributes, contextual features, and lodging attributes normalized by contextual features. The review count normalized by the review counts of surrounding lodgings $\widehat{r}$ (8.07/.018) has the best contribution to the model, making it the the most informative feature. Surprisingly, the unnormalized review count r (−9.8/.010) has a negative EFE and almost 50% smaller LSI than $\widehat{r}$. Although both features contribute positively to the model (according to LSI), EFE suggests that $\widehat{r}$ is a powerful substitute of r, yet they are closely related (Pearson coefficient discards trivial correlation). These observations demonstrate the potential gains of expressing a lodging's reputation in the context of its local neighborhood. Related to this observation, we highlight the importance scores of the standard deviation of the review counts of local lodgings σ_r (5.4/.001). The location rating allows guests to endorse the convenience of the lodgings' location. The location rating is one of the few ratings that obtained positive importance scores (3.0/.001) of all rating features, alongside with the overall star-rating score (4.0/.006).

Price Sensitivity. In this preference dimension, we exclusively use information of the price dispersion of the location of a lodging, to approximate users' exposure to price variation. The lodging price normalized by the Airbnb mean-price of nearby lodgings $\overline{p}_A$ (1.6/.009) obtained greater importance scores than the unnormalized price p (−3.1/.000). Once again, these results show the discriminative power of expressing lodging attributes (less/more

Table 6: Feature Importance Scores (EFE/LSI)

Feature	(EFE/LSI)
Perceived Value (PV)	
Dist. Mean (Travel)	$7.1^{\bullet}/0.001$
Dist. Max (Food)	$5.7^{\bullet}/0.0$
Dist. Mean (Food)	$4.3^{\bullet}/0.0$
Security Dep.	$3.9^{\bullet}/0.003$
Check-ins Mean (Arts)	$3.5^{\bullet}/0.0$
Check-ins Min. (Arts)	$2.9^{\bullet}/0.004$
Check-ins Med. (Travel)	$2.8^{\bullet}/0.0$
Check-ins Min. (Food)	$2.7^{\circ}/0.017$
Bedrooms	$0.6^{\circ}/0.002$
Pets Allowed	$-2.0^{\circ}/0.0$
Hair Dryer	$-2.1^{\bullet}/0.0$
Dist. Norm Max (Food)	$-2.2^{\bullet}/0.017$
Airbed (Bed)	$-2.2^{\bullet}/0.0$
Pool	$-2.9^{\bullet}/0.0$
Price	$-3.1^{\circ}/0.0$
Dist. Med. (Arts)	$-3.3^{\circ}/0.003$
Dist. Min (Food)	$-4.5^{\circ}/0.0$
Perceived Risk (PR)	
Rev. Cnt. Norm. ($\widehat{r}$)	$8.1^{\bullet}/0.018$
Context Rev. Cnt. Std. (σ_r)	$5.4^{\bullet}/0.001$
Star Rating	$4.0^{\bullet}/0.006$
Location (Rating)	$3.0^{\bullet}/0.001$
Rev. Cnt. (r)	$-9.8^{\bullet}/0.01$
Perceived Sensitivity (PS)	
Price Kurtosis	$4.5^{\bullet}/0.008$
Airbnb Skewness	$3.9^{\circ}/0.0$
Price Norm. Airbnb ($\overline{p}_{\mathcal{A}}$)	$2.6^{\circ}/0.009$
Electronic Word of Mouth (EWoM)	
Comp. Skewness	$6.5^{\bullet}/0.0$
Pos. Skewness	$6.2^{\bullet}/0.0$
Neg. Med.	$6.1^{\bullet}/0.009$
Neg. Mean	$5.1^{\bullet}/0.059$
Neu. Med.	$4.6^{\circ}/0.0$
Comp. Med.	$4.1^{\circ}/0.0$
Comp. Max	$-4.1^{\circ}/0.0$
Pos. Min	$-5.8^{\circ}/0.0$
Perceived Authenticity (PA)	
Auth Min.	$6.3^{\bullet}/0.0$
Auth Kurtosis	$4.3^{\bullet}/0.0$

expensive than) relative to their neighbor lodgings. Furthermore, the feature price kurtosis $\mathcal{P}$ Kurtosis (4.5/.008) (symmetry score of the curve shaped by the prices around the accommodation) has a positive contribution to the model. Such results empirically demonstrate the utility of considering the price dispersion around an accommodation. We also observe that the skewness of the Airbnb price histogram (long-tailed score) of the area where a lodging is located obtained diverging importance scores, but EFE denoting a positive impact (3.9/0.0).

Perceived Value. This preference dimension aims to model the trade-off between the monetary sacrifice against the obtained benefits. Firstly, we discuss feature importance of PV's features that overlook the location of the lodging, followed by the results of features that leverage context information. Property type, bed type, room type, cancellation policies, and amenities obtained poor feature importance scores (such attributes are categorical and treated as indicator variables in our model). Also, lodging's traits that one would expect to be important, such as number of beds, bedrooms, and bathrooms, weekly price, and monthly price, surprisingly obtained scores indicating to be noisy or with poor discriminative power. Further experiments are needed to investigate their true usefulness. The following results refer to the usage of features obtained from information of nearby venues around a lodging (leveraging context). Contrary to previous findings where EFE and LSI converged to the same conclusions, features in this category obtained diverging EFE and LSI. A conservative interpretation would not safely attribute unconditional importance to such features. However, our findings are interesting in a way that they present promising results suggesting that it may be feasible to enhance these features' discriminative capabilities to obtain converging EFE and LSI. Some distance features were highlighted by EFE to be impactful. On the other hand, features obtained from the normalized distances were systematically highlighted by the LSI as having high importance scores (6 features for arts & entertainment venues, 3 features for food, 6 features for travel & transportation).

Electronic Word of Mouth & Perceived Authenticity. The following sets of features were obtained using the text contained in the lodging's reviews. Precisely, EWoM includes statistics from sentiment polarity scores (positive, negative, neutrality, and compound) and PA includes features computed using perceived authenticity scores that positively denote a sense of PA. The mean negative sentiment scores (5.1/.059) remarkably obtained the greatest LSI score out of all the features in our model and a congruent EFE. Excluding the latter finding, the rest of the features in both categories did not obtained converging LSI and EFE scores. Nonetheless, their EFE are interesting: compound scores' skewness (6.5/.000) and median (4.1/.000); positive scores' skewness (6.2/.000); negative scores' skewness (6.1/.000); neutrality scores' median (4.6/.000); PA scores' min (6.3/.000) and PA scores' kurtosis (4.3/.000); indicating that they enhance the accuracy of the model. Once again, these feature importance results have to be taken with caution, in order to avoid misleading conclusions. However, these results are encouraging to pursue further investigations of their true usefulness.

Recalling *Q2*, this feature importance study attests the discriminative power of using contextual information to normalize features, as demonstrated for the normalized review count (PR) and the normalized lodging price (PS). Furthermore, characterizing lodgings' surrounding area evidenced the importance of contextual information to enhance the models' accuracy (e.g. standard deviation of the review counts and price kurtosis). Finally, results suggest there is room for improving the EWoM and PA preference dimensions, which may result in a further increase in effectiveness.

5.3 Model Validation

In this section, we further extend the understanding of the preference dimensions in our CLLR model, by analyzing their interactions with one another and observing their accuracy at predicting users' purchase intention. This evaluation addresses *Q3* by drawing a parallel between our data-driven findings and the socio-economic theories of the sharing economy that inspired our study. As previously stated, the preference dimensions in our model were inspired by Liang [22]'s model of Airbnb customers' repurchase intention (Figure 6).

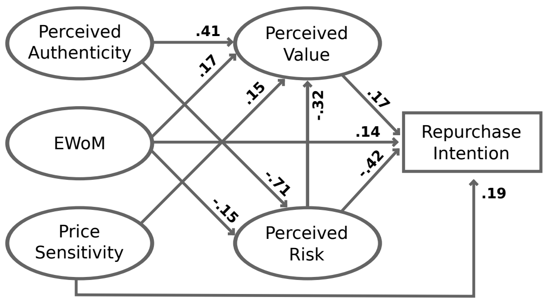

Figure 6: SEM of Airbnb Repurchase Intention (RI) [22].

Liang [22] investigated customers' intentions to reuse Airbnb as an alternative to actually predicting consumers' behavior. She relied on Structural Equation Models (SEM), a technique commonly used in socio-economic studies, which allows to analyze structural relationships based on covariance. In our work, rather than studying *repurchase* intention (RI), we are able to empirically observe Airbnb customers' *purchase* intention (PI). By measuring MRR, we can quantify how accurately each preference dimension can model PI. In particular, let $\mathcal{F} = \{PV, PR, PS, EWoM, PA\}$ denote the preference dimensions in our CLLR. When users consume an item, they implicitly state their preference towards the item's attributes. Therefore, we indirectly observe the guests' preferences for lodging's traits directly from their bookings. Given a preference dimension $\mathcal{F}_i \in \mathcal{F}$, if $\mathcal{F}_i$ shows to be discriminative when predicting users preferences, we assume that it truly implies that users give importance to $\mathcal{F}_i$. In Table 7, $\tau(\mathcal{F}_i)$ shows the MRR obtained from the evaluation of a model using solely $\mathcal{F}_i$. Additionally, $\tau(\mathcal{F}_i \cup \mathcal{F}_j)$ shows the MRR obtained from a model that combines two preference dimensions, $\mathcal{F}_i \cup \mathcal{F}_j$ (• indicates a significant difference between $\tau(\mathcal{F}_i)$ and $\tau(\mathcal{F}_i \cup \mathcal{F}_j)$ according to a two-tailed t-test with $\alpha \leq 0.05$). Finally, the column **SEM RI** indicates correlations in Liang [22]'s SEM.

The first of our findings is that PR is the factor that best explains users' PI and RI in the respective models. The risk involved in booking a lodging is the main driver for consumption and repeated consumption. Also, PV is greatly influenced by PR in both models. Furthermore, both models agree that PV does not influence PR. Our model does not support Liang [22]'s finding that there is a real interaction between PA and PR. Moreover, EWoM has a greater effect in PR than the observed in the SEM. For the rest of the edges, our model often indicates that there is a two-sided effect, not considered in Liang [22]'s research. Recalling question *Q3*, while our model supports many of the observations drawn from the socio-economic analyses conducted by Liang [22], we also find previously unseen effects, further contributing to the understanding of the drivers behind users' lodging consumption behavior in the sharing economy.

Table 7: Preference dimensions, MRR Gain/Loss

$\mathcal{F}_i$	$\tau(\mathcal{F}_i)$	$\mathcal{F}_j$	$\tau(\mathcal{F}_i \cup \mathcal{F}_j)$	SEM of RI
PR	0.038	EWoM	0.036	N/A
		PA	0.037	N/A
		PS	0.037	N/A
		PV	**0.041**	**-0.32**
PS	0.02	EWoM	0.032 •	N/A
		PA	0.037 •	N/A
		PR	0.037 •	N/A
		PV	**0.023 •**	**0.15**
EWoM	0.035	PA	0.039 •	N/A
		PR	**0.036**	**-0.15**
		PS	0.032 •	N/A
		PV	**0.036**	**0.17**
PA	0.034	EWoM	0.039 •	N/A
		PR	**0.037**	**-0.71**
		PS	0.037 •	N/A
		PV	**0.039 •**	**0.41**
PV	0.021	EWoM	0.036 •	N/A
		PA	0.039 •	N/A
		PR	0.041 •	N/A
		PS	0.023 •	N/A

6 CONCLUSIONS & FUTURE WORKS

In this work, we filled a knowledge gap in the recommender systems literature towards the sharing economy environment, precisely, in the lodging domain. We proposed a context-aware learning-to-rank model for lodging recommendation, dubbed CLLR, which demonstrated greater performance than well known state-of-the-art CF techniques and showed significant improvements compared to a real-world lodging recommender. Furthermore, the analysis of the discriminative power of the features in our model permitted the characterization of their effectiveness and demonstrated the importance that contextual information has in the lodging domain. Also, our results suggested that normalizing lodgings' attributes by leveraging context information, such as nearby lodgings' characteristics, improves the discriminative power of the devised features. In addition, we performed a characterization of the preference dimensions in our model, which highlighted the applicability of socio-economic theories of the sharing economy in a practical RS scenario. This characterization provided further evidence to state that customers' perceived risk is one of the most important factors that impacts consumption behavior, which also affects customers' perceived value of a lodging. Furthermore, our findings suggested that electronic word of mouth has a greater effect in perceived risk than the one observed in previous studies. Finally, we observed two-sided effects that were not considered by previous studies of consumption behavior on Airbnb. As future work, we intend to investigate alternative sources of contextual information in order to improve the discriminative power of the features in our model. A promising direction includes normalizing price-related features according to each individual user's perception, whenever historical data is available.

ACKNOWLEDGMENTS

This work was partially funded by projects InWeb (MCT/CNPq 573871/2008-6) and MASWeb (FAPEMIG/PRONEX APQ-01400-14), and by the authors' individual grants from CNPq, FAPEMIG, CONACYT, and COCYTEN.

REFERENCES

[1] Russell Belk. 2014. You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research* 67, 8 (2014), 1595–1600.
[2] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. 2013. Recommender systems survey. *Knowledge-Based Systems* 46 (2013), 109–132.
[3] Joan Borras, Antonio Moreno, and Aida Valls. 2014. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications* 41, 16 (2014), 7370–7389.
[4] Raymond C Browning, Emily A Baker, Jessica A Herron, and Rodger Kram. 2006. Effects of obesity and sex on the energetic cost and preferred speed of walking. *Journal of Applied Physiology* 100, 2 (2006), 390–398.
[5] Dimitrios Buhalis and Peter O'Connor. 2005. Information communication technology revolutionizing tourism. *Tourism recreation research* 30, 3 (2005), 7–16.
[6] Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12, 4 (2002), 331–370.
[7] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval.* ACM, 335–336.
[8] Olivier Chapelle and Yi Chang. 2011. Yahoo! learning to rank challenge overview. In *Proceedings of the Yahoo! Learning to Rank Challenge.* 1–24.
[9] Alexandru M Degeratu, Arvind Rangaswamy, and Jianan Wu. 2000. Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of research in Marketing* 17, 1 (2000), 55–78.
[10] Naveen Donthu and Adriana Garcia. 1999. The internet shopper. *Journal of advertising research* 39, 3 (1999), 52–52.
[11] A Felfernig, S Gordea, D Jannach, E Teppan, and M Zanker. 2007. A short survey of recommendation technologies in travel and tourism. *OEGAI Journal* 25, 7 (2007), 17–22.
[12] Samuel P Fraiberger and Arun Sundararajan. 2015. Peer-to-peer rental markets in the sharing economy. *NYU Stern School of Business Research Paper* (2015).
[13] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.
[14] Michael J Gross and Graham Brown. 2008. An empirical structural model of tourists and places: Progressing involvement and place attachment into tourism. *Tourism management* 29, 6 (2008), 1141–1151.
[15] Juho Hamari, Mimmi Sjöklint, and Antti Ukkonen. 2015. The sharing economy: Why people participate in collaborative consumption. *Journal of the Association for Information Science and Technology* (2015).
[16] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International AAAI Conference on Weblogs and Social Media.*
[17] Jacob Jacoby and Leon B Kaplan. 1972. The components of perceived risk. In *SV-Proceedings of the third annual conference of the association for consumer research.*
[18] Katerina Kabassi. 2010. Personalizing recommendations for tourists. *Telematics and Informatics* 27, 1 (2010), 51–66.
[19] Rob Law, Dimitrios Buhalis, and Cihan Cobanoglu. 2014. Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management* 26, 5 (2014), 727–750.
[20] Bun Song Lee and John F McDonald. 2003. Determinants of commuting time and distance for Seoul residents: The impact of family status on the commuting of women. *Urban Studies* 40, 7 (2003), 1283–1302.
[21] Asher Levi, Osnat Mokryn, Christophe Diot, and Nina Taft. 2012. Finding a needle in a haystack of reviews: cold start context-based hotel recommender system. In *Proceedings of the sixth ACM conference on Recommender systems.* ACM, 115–122.
[22] Lena Jingen Liang. 2015. *Understanding repurchase intention of Airbnb consumers: perceived authenticity, EWoM and price sensitivity.* Ph.D. Dissertation. University of Guelph.
[23] Tie-Yan Liu and others. 2009. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval* 3, 3 (2009), 225–331.
[24] Claudio Lucchese, Franco Maria Nardini, Salvatore Orlando, Raffaele Perego, and Nicola Tonellotto. 2015. Speeding up Document Ranking with Rank-based Features. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 895–898.
[25] Mareike Möhlmann. 2015. Collaborative consumption: determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of Consumer Behaviour* 14, 3 (2015), 193–207.
[26] Mehrbakhsh Nilashi, Othman bin Ibrahim, Norafida Ithnin, and Nor Haniza Sarmin. 2015. A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS. *Electronic Commerce Research and Applications* 14, 6 (2015), 542–562.
[27] Andreas Papatheodorou. 2001. Why people travel to different places. *Annals of tourism research* 28, 1 (2001), 164–179.
[28] Jay M Ponte and W Bruce Croft. 1998. A language modeling approach to information retrieval. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval.* ACM, 275–281.
[29] Daniele Quercia, Joao Paulo Pesce, Virgilio Almeida, and Jon Crowcroft. 2013. Psychological maps 2.0: a web engagement enterprise starting in London. In *Proceedings of the 22nd international conference on World Wide Web.* ACM, 1065–1076.
[30] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence.* AUAI Press, 452–461.
[31] Ryosuke Saga, Yoshihiro Hayashi, and Hiroshi Tsuji. 2008. Hotel recommender system based on user's preference transition. In *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on.* IEEE, 2437–2442.
[32] Raul Sanchez-Vazquez. 2017. *Repurchase Intention for Lodging Recommendation.* Master's thesis. Universidade Federal de Minas Gerais.
[33] Erika Sandow and Kerstin Westin. 2010. The persevering commuter–Duration of long-distance commuting. *Transportation Research Part A: Policy and Practice* 44, 6 (2010), 433–445.
[34] Venkatesh Shankar, Arvind Rangaswamy, and Michael Pusateri. 1999. The online medium and customer price sensitivity. *University Park: E-business Research Center* (1999).
[35] Beverley Sparks, John Bowen, and Stefanie Klag. 2003. Restaurants and the tourist market. *International Journal of Contemporary Hospitality Management* 15, 1 (2003), 6–13.
[36] Inc TranSafety. 1997. Study compares older and younger pedestrian walking speeds. *Road Management & Engineering Journal* (1997).
[37] Wen-Chin Tsao, Ming-Tsang Hsieh, Li-Wen Shih, and Tom MY Lin. 2015. Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management* 46 (2015), 99–111.
[38] Qiang Wu, Chris J. C. Burges, Krysta M. Svore, and Jianfeng Gao. 2008. *Ranking, boosting, and model adaptation.* Technical Report. Microsoft Research.
[39] Georgios Zervas, Davide Proserpio, and John Byers. 2016. The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Boston U. School of Management Research Paper* 2013-16 (2016).
[40] Kai Zhang, Keqiang Wang, Xiaoling Wang, Cheqing Jin, and Aoying Zhou. 2015. Hotel recommendation based on user preference analysis. In *Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference on.* IEEE, 134–138.