REPURCHASE INTENTION FOR LODGING RECOMMENDATION

LUIS RAÚL SÁNCHEZ VÁZQUEZ

REPURCHASE INTENTION FOR LODGING RECOMMENDATION

Dissertation presented to the Graduate Program in of the Universidade Federal de Minas Gerais — Departamento de Ciência da Computação in partial fulfillment of the requirements for the degree of Master in .

Advisor: Rodrygo Luis Teodoro Santos

Belo Horizonte
March 2017

Sánchez Vázquez, Luis Raúl

S211r Repurchase Intention for Lodging Recommendation / Luis Raúl Sánchez Vázquez. — Belo Horizonte, 2017
xvii, 90 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de Minas Gerais — Departamento de Ciência da Computação
Orientador: Rodrygo Luis Teodoro Santos

1. Computação — Teses. 2. Sistemas de recomendação. 3. Indústria hoteleira. I. Orientador. II. Título.

CDU 519.6*73(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Repurchase intention for lodging recommendation

LUIS RAÚL SÁNCHEZ VÁZQUEZ

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

RLSantos

PROF. RODRYGO LUIS TEODORO SANTOS - Orientador
Departamento de Ciência da Computação - UFMG

PROF. DENIS PARRA SANTANDER
Departamento de Ciência da Computação - PUC Chile

Jussara Marques de Almeida Gonçalves

PROFA. JUSSARA MARQUES DE ALMEIDA GONÇALVES
Departamento de Ciência da Computação - UFMG

PROF. PEDRO OLMO STANCIOLI VAZ DE MELO
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 30 de março de 2017.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor, Prof. Rodrygo Luis Teodoro Santos, for his support, guidance, knowledge, and all the time of research and writing of this dissertation. I also would like to thank all the people from Brazilian and Mexican institutions who contributed and helped to undertake this MSc. Especially, my sincere gratitude to CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), COCYTEN (Consejo de Ciencia y Tecnología del Estado de Nayarit), and CONACYT (Consejo Nacional de Ciencia y Tecnología de México) for the funding and trust they provided for my research, which was invaluable to complete this work. Also, I compliment the great administrative and research staff at the graduate program of computer science at Federal University of Minas Gerais. I also thank my fellow labmates at LATIN (Laboratory for Treating Information) and LBD (Database Laboratory), for the stimulating discussions and days we worked together before deadlines and exams. It has been a period of intense learning for me, not only in my academic and scientific career, but also on a personal level. I am extremely grateful to my family for the continued and unconditional support. I express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them. Thank you.

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 travelers with hosts, allowing both to exchange cultural experiences in addition to the economic transaction. Inspired by recent marketing analyses of repurchase intent behavior on Airbnb, we propose a 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 choice, the risk involved in choosing it rather than other available options, the authenticity of the experience it could provide, and its overall perception by other users. Through a comprehensive evaluation using publicly available Airbnb data, we demonstrate the effectiveness of our proposed approach compared to a number of alternative recommendation benchmarks, including a simulation of Airbnb's own recommender.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Recommender systems (RS) are applied in a wide range of domains. They have become crucial for many everyday tasks [Sivapalan et al., 2014] in order to tackle the information overload caused by the vast amount of information that exceeds human capabilities to manually search for relevant items. Recommender systems extensively demonstrated their importance in many industries as a major source of revenue and user satisfaction. They are particularly popular for e-commerce companies (e.g. online retailers, streaming movie sites, music applications, and social networks) [Dabholkar and Sheng, 2012; Knijnenburg et al., 2012], representing a substantial component of their business models [Schafer et al., 1999].

The tourism industry has largely adopted the Internet as one of its main sales' channel, helping customers to find information in order to plan their trip [Law et al., 2014]. Due to the complexity of the travel planning process, specialized e-tourist agencies became popular services, which help travelers with personalized assistance to plan their trip [Bobadilla et al., 2013; Buhalis and O'Connor, 2005]. More recently, recommender systems have been proposed to tackle the problem [Felfernig et al., 2007; Kabassi, 2010; Borras et al., 2014] in order to personalize recommendations for users, allowing them to have a customized interaction with online platforms, dealing with the issue of information overload and simplifying the complexity of the task.

The tourism and travel industry is currently being reshaped by the so-called the sharing economy, which is a broad term embracing peer-to-peer-based activities of giving, sharing, exchanging, or purchasing goods and services, exploiting information and communication technologies [Hamari et al., 2015]. Due to the recent emergence of such phenomena, the theories and understanding of the sharing economy are yet overlooked and unexploited in recommender systems, as in many other domains. Researchers have highlighted the divergence and disparity of customers and items of the sharing economy when compared to their traditional business-to-peer counterparts [Möhlmann, 2015; Hamari et al., 2015], which is mainly explained by customers' motivational drivers towards consumption in an online, collaborative, and shared manner.

In this dissertation we focus on the lodging recommendation problem in sharing economy environment. More specifically, the recommendation task is given a location, where the user wishes to sojourn, retrieve a list of lodgings sorted by relevance, which is composed of accommodations located at the neighboring of the input location. In order to tackle the problem, we proposed CLLR, a recommender which adopts the concepts of a repurchase intention model [Liang, 2015], to build five preference dimensions to be leveraged by a learning-to-rank algorithm, which are sustained in five customer behavioral premises towards consumption in the sharing economy: price value (cost/benefit of a particular lodging choice), perceived risk (uncertainty of unwanted outcomes in the transaction), perceived authenticity (genuineness of experiencing locals' lifestyle), electronic-word-of-mouth (informal opinions of other users), and price sensitivity (awareness of the price dispersion).

Through a comprehensive evaluation we demonstrate the effectiveness of our proposed model using publicly available data from the largest lodging provider of the sharing economy, Airbnb.[1] To the best of our knowledge, this dissertation is the first of its kind that addresses the sharing economy in recommender systems.

1.1 Motivation

According to the U.S. Census Bureau,[2] in 2014, the accommodation industry reported revenues of $221.8 billion dollars. Some studies claim that Airbnb, a booking site for lodging and one of the most notorious icons of the sharing economy, have rapidly grown to encompass up to 10% of the total hotel industry revenue for some travel destinations [Zervas et al., 2016]. With the promising future of the sharing economy, researchers expect the emergence of more peer-to-peer travel and tourism agencies [Zekanovic-Korona and Grzunov, 2014]. The launch of the new service Airbnb Experience (end of 2016) allows users to design experiences/activities that other users can book, this service is evidence of the expansion of the sharing economy in the recreational touristic domain. Having demonstrated their potential, it is natural to expect that over the following years we will be witnessing the sharing economy becoming an important portion of the digital economy. Thus, the recommender systems community should also extend its efforts to tackle the challenges that such domain may pose.

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

[2] http://www.census.gov/newsroom/press-releases/2014/cb14-tps81.html

1.2 Dissertation Statement

The statement of this dissertation is that sharing economy lodging recommendation can be improved by modeling users' repurchase intention. In particular, this dissertation aims to answer the following research questions:

- **RQ1**: How accurate is CLLR for lodging recommendation?
- **RQ2**: How robust is CLLR for lodging recommendation?
- **RQ3**: How do single features contribute to the performance of CLLR?
- **RQ4**: How do our results relate to existing theories of the sharing economy?

In order to answer these research questions, in this dissertation we explored social-economic theories relevant for the sharing economy in the lodging domain, from which we proposed an approach that models such theories. Then, we designed an evaluation methodology that required the construction of an adequate test collection, to finally analyze the results to answer our research questions.

1.3 Contributions

This research has led to the accomplishment of the following contributions:

- **Contribution 1:** A feature-based lodging recommendation approach, inspired by socio-economic theories of repurchase intention for the sharing economy.
- **Contribution 2:** A test collection and evaluation methodology for lodging recommendation.
- **Contribution 3:** A comprehensive evaluation of the proposed approach in a case study using Airbnb.

1.4 Dissertation Overview

Below we present a brief outlook that summarizes each one of the chapters in this dissertation:

- **Chapter 2: Background and Related Work** gives the base theory to understand recommender techniques and sharing economy concepts, with a survey of the research works relevant to this dissertation.

- **Chapter 3: Repurchase Intention Model for Recommendation** defines our proposed model which integrates sharing economy concepts to recommendation systems.
- **Chapter 4: Data Acquisition** describes the procedure for the construction of our test collection and characterizes the dataset we built.
- **Chapter 5: Experimental Setup** explains the design of our evaluation framework with the procedure employed to conduct our experiments.
- **Chapter 6: Experimental Results and Analyses** presents and analyzes the results.
- **Chapter 7: Conclusion and Future Works** concludes the dissertation, summarize findings and introduces a set of future research directions.

Chapter 2

Background and Related Work

This chapter provides an overview of the sharing economy as a technological trend, and discusses implications for modeling users' behavioral attitudes, with a special consideration for repurchase intention. In addition, it provides background on the recommendation problem, reviewing classical recommendation techniques, as well as efforts related to the lodging recommendation domain.

2.1 Sharing Economy

The sharing economy is a concept that may have different nuances according to the context where it is employed, with authors having to delimit the extension of its meaning. In this dissertation, the sharing economy is defined as new online marketplaces that create innovative consumption modalities that stand in sharp contrast to their traditional or conservative counterparts [Belk, 2014; Hamari et al., 2015; Fraiberger and Sundararajan, 2015]. According to the last definition, sharing economy companies can sell/provide items that are not new. Nevertheless, such companies must adopt information technologies as their main mediator with their customers in addition to offer innovative variants that would separate them from their competitors (not necessarily involving sharing). Collaborative consumption refers to transactions between contributors and end-users in a Peer-to-Peer (P2P) manner, involving sharing resources or conjoint consumption (multiple buyers collaborating to consume) [Hamari et al., 2015]. Online collaborative consumption is in most cases a subclass of the sharing economy [Hamari et al., 2015], where the transaction in a P2P manner is the distinctive factor that defines such companies apart from their traditional counterpart. Prominent examples of P2P sharing economy companies are Uber,[1] the private trans-

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

portation company, and Airbnb,[2] the booking lodging site.

Economic and social researches have studied the sharing economy and found inherent peculiarities of the peer-to-peer lodging domain, highlighting important characteristics such as utility, trust, cost savings, and familiarity, to be particularly relevant for their customers [Möhlmann, 2015; Hamari et al., 2015; Liang, 2015]. This translates to users having different consumption preferences, suggesting that customers portrait their behavior differently in the lodging sharing economy than in traditional accommodation markets. Another important characteristic of the sharing economy in the lodging domain is that in contrast to hotel supply it can rapidly adapt (add and remove lodging supply) in response to periods of high demand. In addition, lodging supply shows a wider geographical coverage, making the lodging sharing economy more than just a substitute for hotels [Zervas et al., 2016]. The latter observation suggests that the lodging sharing economy ecosystem substantially diverges from the hotel recommendation scenario.

2.1.1 SEM Repurchase Intention Model

According to the Theory of Reasoned Action (TRA) there exists a straight and tight relationship between attitude and behavior [Ajzen and Fishbein, 1977]. Precisely, TRA states that intention is the best predictor that precedes human action. Therefore, it is appropriate to predict consumers' purchase behavior by measuring their repurchase intention (RI).

Airbnb[3] is a sharing economy web application that permits users to list, search, and rent lodgings enabling guests to benefit from locals' advice. Liang [2015] studied Airbnb customer's repurchase intention, and validated a model that was constructed using structural equation modeling (SEM), a multivariate statistical analysis technique used to analyze structural relationships. In such model, RI is shown to be dependent on two main variables: perceived value (PV) and perceived risk (PR), and that in turn, PV and PR are influenced by other variables named perceived authenticity (PA), electronic-word-of-mouth (EWoM), and price sensitivity (PS). The SEM model constructed by Liang [2015] (SEMRI) is presented in Figure 2.1. The following sections define the concepts that led to the creation of Liang [2015]'s model, and discuss how they operate in the sharing economy lodging setup.

[2] `https://www.airbnb.com/`
[3] `https://www.airbnb.com/`

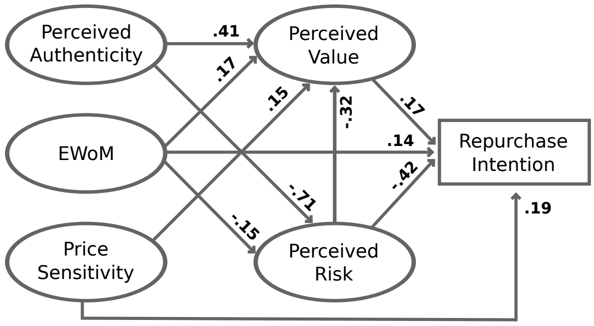


Figure 2.1: SEMRI, edge weights denote correlations.

2.1.2 Perceived Value

Zeithaml [1988] define perceived value (PV), as the overall consumers' judgment with respect to the utility of a service/product, based on their perceptions of cost effectiveness (gain versus given).

PV is defined by many authors as a trade-off between benefits (something the consumer receives) and sacrifices (something the consumer gives up) [Kashyap and Bojanic, 2000]. Both are customer's perceptions of a mixture of the attributes of the item involved (quality, utility, benefits) upon perceived sacrifices (commonly monetary). In our lodging context, this translates to how valuable the amenities, location, comfort, and other features of the accommodation are with respect to the price of the lodging.

2.1.3 Perceived Risk

As defined by Kim et al. [2008], perceived risk (PR) is the risk sensed or the probability of negative consequences (unwanted outcome) associated with a transaction, which increases with higher levels of uncertainty. Researchers have investigated various aspects of customers' perceived risk to understand attitudes towards consumption [Jacoby and Kaplan, 1972]. Typically, purchasing is part of a decision process, where the outcome and consequences of choosing are uncertain, and can only be known in the future [Taylor, 1974]. It has been shown that in the presence of high risk, intention to repurchase is negatively influenced [Chang and Tseng, 2013]. For online setups, PR is clearly a concept that influences customers' behavior, as for most tangible goods and some services found online, it is not possible for customers to experience the actual item prior to purchase.

2.1.4 Price Sensitivity

Price sensitivity in the touristic domain has been subject of multiple studies that investigate how prices affect tourist' purchasing behavior [Masiero and Nicolau, 2012a,b; Nicolau and Masiero, 2013]. PS is a concept that refers to the extent to which pricing dictates consumption. Also, it is directly linked to the level of importance consumers place on price relative to other purchasing criteria [Masiero and Nicolau, 2012b]. Price in online contexts has been subject of many research studies [Donthu and Garcia, 1999; Shankar et al., 1999; Degeratu et al., 2000] and is probably one of the most important decision factors, broadly accepted in influencing consumers purchase behavior. It has been shown that the price sensitivity of a customer increases the more he or she is aware of the price dispersion of a given product [Degeratu et al., 2000]. As the trip planning process is an activity that requires a vivid information search exercise [Fodness and Murray, 1997], it is intuitive to assume that travelers become aware of the lodging supply's price dispersion, while comparing different lodging provider alternatives (e.g. hotels, hostels, campsites), thus, making this scenario intrinsically price sensitive.

2.1.5 Electronic-Word-of-Mouth

Electronic-Word-of-Mouth (EWoM) is defined as any form of informal Internet-based information directed to consumers that is related to the usage/attributes of particular goods, services, or their sellers (e.g. social media posts, blogs, forums, reviews) [Litvin et al., 2008]. Reviews are one of the most frequent forms of EWoM that can be found in most online vendors' websites [Pang and Lee, 2008]. Most tourist and travel online agencies have also adopted the usage of reviews as a standard practice. Indeed, many studies evidence the strong influence that reviews have on travelers' purchase decisions [O'Connor, 2008; Gretzel and Yoo, 2008], demonstrating their function as a catalyst of users' perception regarding a particular lodging [Cantallops and Salvi, 2014].

2.1.6 Perceived Authenticity

In our scenario, Perceived Authenticity (PA) is the extent to which a guest feels like natively living at the lodging place. In the sharing economy lodging domain, guests have the chance to experiment local's lifestyle, due to the close interaction with their hosts' and their living space. PA is a concept that characterizes P2P lodging companies and a distinctive factor with respect to hotels, their most conventional counterpart. Grayson and Martinec [2004] state that PA is defined as something or someone perceived to be real and consequently authentic. More specifically, in the scope of this work, the concept of PA is defined to be the genuineness of the foreigner traveler experiencing locals' lifestyle [Liang, 2015]. Such familiarity trait is attractive for sharing economy lodging customers [Guttentag, 2015] and is positively correlated with user's satisfaction, which in turn encourages consumption.

2.2 Recommender Systems

Recommender systems (RS) are a branch of information systems that are widely used in many real-world setups and can be particularly common in e-commerce websites. Recommender systems aim to filter items from a large catalog that is intractable for humans to explore. The filtering criterion employed is usually intended to fit users' preferences, interests, tastes, or needs. In contrast to other information filtering systems, recommender techniques are aimed to be proactive, enhancing discovering mechanisms, as users do not have to explicitly state their information needs. According to Jannach and Adomavicius [2016], the main purpose of recommenders is to suggest *good items* for a user, which in practice translates to two scenarios: (1) predicting user's rating scores (rating prediction) or (2) ranking items according to the estimated user preferences (top-N recommendation). Balabanović and Shoham [1997] proposed one of the first categorizations of RS in the literature, which is composed of three classes, named collaborative filtering (CF), content based (CB), and hybrid recommendation techniques.

Collaborative filtering are techniques that make recommendations based on users' opinions [Resnick et al., 1994; Shardanand and Maes, 1995] and are the most traditional approaches. The intuition behind CF, is that given a user u, CF objective is to identify other users with similar tastes than u to exploit the items they have liked in the past, which are employed to make recommendations for the target user u. CF techniques are advantageous as they are completely agnostic of the representation of the items being recommended. Indeed, they use what users consumed (generally a user-item rating matrix) without explicitly having to model complex objects such as music and movies [Burke, 2002].

Content based are techniques that recommend items that are similar to the set of items that user u has liked in the past. In contrast to CF, a content based recommender needs users'/items' representation of their interests, based on the items that he or she has rated in the past, by leveraging the features that those items possess [Schafer et al., 1999], in order to compute item-to-item similarity [Balabanović and Shoham, 1997].

Finally, **hybrid techniques** are defined as methods that combine two or more recommendation techniques, with the purpose to outperform the usage of a single one [Burke, 2002, 2007]. Hybridization aims to combine the strengths of multiple methods while alleviating the drawbacks and weaknesses of individual recommendation techniques. The feature-based model proposed in this dissertation falls into this category. In the next section, we review the classification of hybrid recommendation proposed by Burke [2007], which is composed of seven different hybridization techniques: weighted, switching, mixed, feature combination, feature augmentation, cascade, and meta-level.

2.3 Hybrid Approaches

According to Jannach et al. [2010], hybridization techniques can be grouped under the taxonomy shown in Figure 2.2, encompassing three classes: Monolithic, Ensemble, and Mixed. Ensemble systems are designed to leverage multiple off-the-shelf algorithms by combining their results into a single and more robust output. The operation of their internal recommendation algorithms can be done sequential or in parallel. Monolithic systems uses various types of data that are integrated into one recommendation algorithm. Finally, mixed systems, similarly to ensembles, use multiple recommendation algorithms as black-boxes and the items recommended are presented together side by side, instead of combining multiple results.

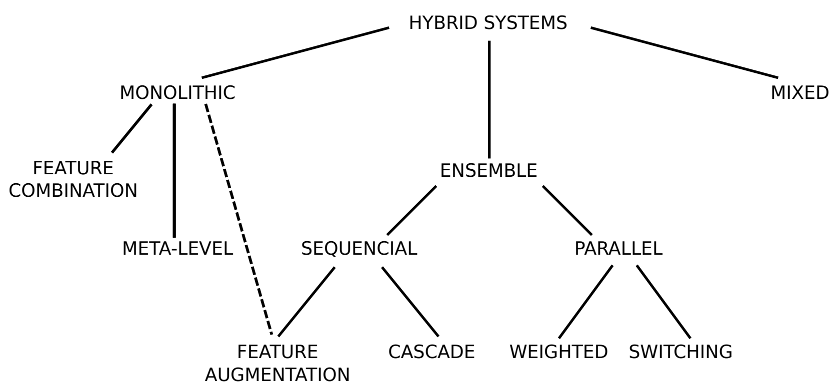


Figure 2.2: Taxonomy of hybrid systems.

In **Weighted** hybrid techniques, the outputs of the multiple recommenders are merged to a single scoring or ranking method, where each individual technique is operated independent of each other, nevertheless, they all contribute a portion of the total rating score or the ranking produced. Linear combinations of the scoring results of each recommendation component commonly fall under this category. The union/intersection of item sets, that are shared between the exclusive operation of the recommendation modules, also fall in this category.

Contrary to weighted hybridization, **Switching** does not merge multiple recommendation outputs. Instead, according to the given scenario they face, these methods are able to switch between recommendation techniques. Therefore, a decision criterion has to be established in order to chose the recommender module that will likely achieve the best performance.

In **Mixed** Hybrid recommendations, the outputs from different recommenders are presented together, but in contrast to weighted hybridization, there is no shared information between the operation of the recommendation modules. Instead, mixed hybrid approaches output a mixture of the outputs of the recommendation modules, for instance in a side-by-side recommendation lists.

Feature Combination hybridization employs features derived from multiple knowledge sources, for instance, the scores of other recommenders. Features are combined together to constitute a single input to the recommendation model. This hybrid class has not multiple tangible or well defined recommendation components, but rather a single unit recommender.

Feature Augmentation is a stacked version of feature combination, where multiple recommendation modules are chained. The output of one module is included in the input features of the next one. This strategy allows for each module to stamp its recommendation domain logic, contributing to augment the features of each item.

In **Cascade** hybridization, recommenders are organized in a hierarchy in such a way that decisions made by a higher-order recommender cannot be altered by a lower-order one. An example of this method is a recommendation module that re-ranks the items recommended by a higher-order module.

Meta-level hybrid recommendation is achieved when the model produced by one of the hybrid recommender modules serves as the input for another module. In contrast with feature augmentation, meta-level hybridization completely ingests models and not merely the features.

2.4 Touristic and Lodging Recommendation

According to surveys in recommender systems [Felfernig et al., 2007; Kabassi, 2010; Borras et al., 2014], there are multiple works in RS that aim to assist tourists and travelers in their planning process. Concretely, these different approaches recommend a set of the following items: travel destinations, touristic attractions, recreational activities, accommodation, and eatery venues. The majority of these works recommend lodgings in conjunction, or constrained, to other items, which leaves unclear how to apply them to pure lodging recommendation. Nevertheless, there are sporadic works that exclusively dedicate their effort to recommend solely accommodations.

The first work we present was proposed by Saga et al. [2008]. Their proposal leverages users' booking history, in order to create an undirected hotel-guest graph. The graph is converted into a *preference-transition-network*, which is the result of transforming the initial hotel-guest graph to an hotel-hotel graph. The intuition behind such transformations is that it allows to certain pairs of hotel nodes to be linked together by mean of the guests they hosted, a link between hotel $\mathcal{A}$ to $\mathcal{B}$ is interpreted as users that booked $\mathcal{A}$ are likely to transition to book hotel $\mathcal{B}$, as the transition is supported by users booking records, from which the graph was built. Then, recommendation is achieved in two steps, candidate selection and candidate ranking. Candidate selection is achieved after the user explicitly selects an initial hotel, and based on the preference-transition-network, the system selects neighbor nodes. The next step is candidate ranking, where candidate hotels are sorted using the scoring function $InDegree_i - OutDegree_i - C^2$, where $InDegree_i(OutDegree_i)$ is the in(out)-degree of the candidate hotel i in the transition network, and C is a penalty for the previous booking of the given hotel.

Levi et al. [2012] characterize hotel recommendation as a cold-start scenario, as users do not rate enough hotels to enable techniques such as collaborative filtering to achieve recommendation. The intuition of their proposal is to mimic how users build their opinion about a lodging while reading the reviews that guests wrote to it. They assume that readers evaluate an accommodation giving more importance to reviews wrote from people with the same background as theirs. Background is defined in terms of nationality, travel intention (single, couple, family, group, and business), and preferences on hotel traits (location, service, food, room, general, other). Therefore, hotels are modeled using the text of the reviews, where words are seen as features with different importance weights according to different groups of users under each category (nationality, travel intention, and hotel traits preferences). Then, giving to the recommender a user's nationality, travel intention, an preferences on hotel aspects, the systems is able to computed a score for each hotel, which is a compound of the feature weights of the groups that the target user belongs. They conducted an experiment using the Internet crowd-sourcing service Mechanical Turk[4] to validate their proposal, presenting evidence of how reviews can be mined to identify typical types of users for

[4]https://www.mturk.com/mturk/welcome

a given hotel.

In the work done by Zhang et al. [2015], they proposed a hybrid recommender system that uses collaborative filtering in addition to content based. Their approach consist in three stages, (1) completion of the user-item rating matrix, (2) improvement of the user-item rating matrix, and (3) personalized top-N recommendation. The first stage uses Latent Dirichlet Allocation (LDA) [Blei et al., 2003], a topic model that assumes that documents are produced from a mixture of topics. Exploiting the textual content of a document, LDA allows to backtrack how much a given document belongs to each latent topic. LDA is used to generate user/hotel representation as topic vectors, employing the reviews that a guest wrote and that a hotel received. Then, the topic vectors serve to compute user-user and a hotel-hotel similarity matrices (S_U and S_V) via Pearson correlation. The item-user matrix X and the similarity matrices S_U and S_V serve as input for their proposed CF matrix factorization technique, named Preference Factor Model (PFM). PFM uses gradient descent to complete the user-item matrix X, similarly than Singular Value Decomposition (SVD) [Billsus and Pazzani, 1998] does, adding terms considering S_U and S_V. The second step normalizes the rating scores in X, taking into consideration three main factors (a) the mean rating on each intent class, (b) the sparsity of the user-item sub-matrix in the intent category, and (c) specific user rating statistics (frequency and average). Users' intent is given by their data collection (single, couple, group, family, business, and others). The final step produces top-N recommendations with maximal marginal relevance (MMR) [Carbonell and Goldstein, 1998], which is a diversification technique used in information retrieval, that reduces redundancy while maintaining query relevance. In other words, MMR is a trade-off between relevance and duplication. Finally, their evaluation demonstrated that rating predictions can be improved by leveraging users' traveling intent.

PCA-ANFIS is a hybrid method proposed by Nilashi et al. [2015]. Their approach was employed for hotel recommendation in a multiple rating scenario, where *multiple ratings* refers to users rating multiple hotel aspects (e.g. value, rooms, location, cleanliness, check in/front desk, service, and business services). PCA-ANFIS consist in two stages, the first is (1) data preprocessing and (2) training. The first stage creates a three dimensional rating tensor, where the tensor's dimensions correspond to users, items, and the multiple ratings. The tensor is used to cluster users, using Expectation Maximization (EM) algorithm [Moon, 1996], and for each cluster Principal Component Analysis (PCA) [Wold et al., 1987] was performed for dimensionality reduction.

After applying PCA for each cluster, the second step consist in training multiple ANFIS models, in order to predict overall ratings in each cluster. Adaptive Neuro-Fuzzy Inference System (ANFIS) [Jang, 1993] is a kind of neural network that integrates fuzzy logic principles [Zadeh, 1965]. Their experiments demonstrated that PCA-ANFIS leads to the improvement in predictive accuracy of tourism multi-criteria prediction. They evaluated their proposal using a TripAdvisor's[5] test collection, demonstrating the improvement in accuracy of their multi-criteria collaborative filtering.

2.5 Summary

In this chapter we introduced the sharing economy phenomena. We highlight particularities that detach the traditional hotel scenario from the sharing economy lodging domain. We also dedicate special attention to a repurchase intention model, supported by five customers behavioral premises, investigated under the context of the lodging sharing economy, named price value, perceived risk, perceived authenticity, electronic-word-of-mouth, and price sensitivity. After characterizing the context of this work, we presented a taxonomy of recommender systems, relevant to the scope of this dissertation, which is composed of three categories: collaborative-filtering, content-based, and hybrid approaches. A section was dedicated to detail the subclasses of hybrid approaches. Finally, this chapter was concluded presenting some of the few works that are dedicated to recommend hotels. None of the current works on lodging recommendation have evaluated their proposal on the more specific sharing economy domain. In the following chapter we discuss our proposed feature model for lodging recommendation, which is inspired in the related theories of the sharing economy that were addressed in this chapter.

[5] https://www.tripadvisor.com/

Chapter 3

Contextual Socio-Economic Models For Lodging Recommendation

Recommender systems (RS) have demonstrated their effectiveness in scenarios such as recommendation of music, movies, social media content, and online purchases. In such domains, users' historical transactions are typically abundant, whereas lodgings booking transactions are far more rare and for some users very sporadic. RS typically leverage information on users, items, or interactions between both. In contrast, the target domain of this dissertation has severely sparse user profiles, which limits the applicability of traditional collaborative or content-based approaches. In the data collection used in this dissertation, the user-item matrix is very sparse (99.9997%) and most of the users have small booking transactions profiles (See Figure 4.6, 80% of the users possesses less than 5 bookings), which hinders an accurate modeling of users' distinctive preferences that is crucial for tackling the recommendation problem.

On the other hand, existing lodging recommendation approaches ignore the economic drivers that motivate users to consume lodgings in the sharing economy. Researches have found important peculiarities that substantially distinguish the hotel domain from the sharing economy lodging scenario [Möhlmann, 2015; Hamari et al., 2015; Zervas et al., 2016; Liang, 2015]. In particular, P2P lodging has been described as more than just a hotel substitute for three main reasons: (1) P2P lodgings provide a much more dynamic ecosystem as lodging supply can rapidly respond to changes in demand, (2) P2P lodgings serve 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.

In order to overcome the sharing economy challenges in the lodging domain, 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. In the following sections we first introduce the lodging recommendation problem, to further explain our approach, which is inspired by recent socio-economic studies in the domain of Airbnb, the largest lodging provider of the sharing economy.

3.1 The lodging recommendation problem

In this dissertation, the recommendation task is defined as given a location where the user wishes to sojourn, retrieve a list of lodgings sorted by their estimated relevance, composed of accommodations located at the neighborhood of the input location. Precisely, the recommendation task is a top-N recommendation task, therefore, it can also be conceived as a ranking problem. Figure 3.1 illustrates the problem, showing a user u, that inputs a target location l to the system. l is used to obtain $\mathcal{I} = \{i_1, \cdots, i_n\}$, a set of n lodgings at the surroundings of l, which is passed to a ranker $f(u, \mathcal{I}) \rightarrow \pi$, a function aimed to produce an optimal permutation π, of the set $\mathcal{I}$, prioritizing relevant items to the user u.

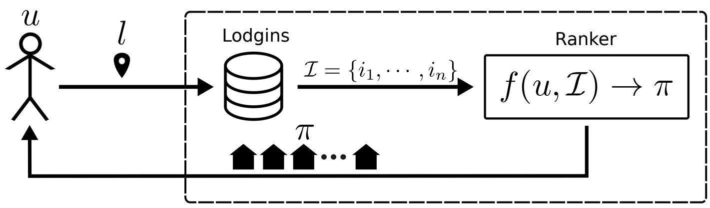


Figure 3.1: Ranking problem.

According to the outlined recommendation task, this dissertation proposes a feature-based context-aware model for the set of lodgings $\mathcal{I}$, which is explained in the following Section 3.2, that in turn, is leveraged by an effective ranker $f(u, \mathcal{I})$ capable to employ these features, as explained in Section 3.3.

3.2 Context-Aware

The context-aware model proposed in the following section is inspired by recent socio-economic studies in the domain of Airbnb [Liang, 2015]. These studies analyze five aspects of users' recurrent consumption of Airbnb lodgings, that in our model we name as preference dimensions, namely:

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

These last concepts are overlooked in RS as drivers of users consumption, especially in the lodging sharing economy domain (Section 2.4). The following sections explain the computation of each one the preference dimensions. Table 3.1 lists and summarizes the 176 features used to represent lodgings in our approach, which are aimed to capture the socio-economic context of an available accommodation in various ways. The **input** column denotes whether each feature is estimated based upon the lodging i, lodging context c, or both.

3.2.1 Perceived Value (PV)

The notion of PV can be conceived as the trade-off between the lodging's benefits versus the price of the lodging. Most of the lodgings' attributes follow under this concept (e.g. wireless Internet, number of beds, pool, tv cable, weekly price, security deposit). However, these features solely characterize rooms with their enclosed attributes and overlook the value derived from the local context. For instance, having an accommodation close to a metro station may be more valuable than the exact same lodging at the same neighborhood, but more distant from public transportation. Similarly, the nearness of other points of interest at the surroundings of a lodging may increase the value of the respective accommodation. Therefore, in order to better quantify the perceived value of a lodging, we need to consult alternative sources of information.

Web mapping services are now a reality made possible by a number of service providers such as Google Maps,[1] OpenStreetMaps,[2] Yelp,[3] Foursquare,[4] and many

[1] `https://www.google.com/maps`
[2] `https://www.openstreetmap.org/about`
[3] `https://www.yelp.com/about`
[4] `https://foursquare.com/about`

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

Table 3.1: Lodging recommendation features.

others. Usually, such services offer an API that allows to query for a geographical coordinate to get information of the surrounding geographical layout, venues, transit, and other statistics. In this particular work, we use Foursquare's API.[5]

To better capture the concept of PV, such services may be used to obtain information around the lodgings' geo-location using a fixed custom radius (as shown in Table 3.2). In order to characterize the surroundings of the accommodation in a sense that is compatible with Airbnb guests' traveling intention, we propose to use mapping services to fetch information of three venues' categories: Food, Art & Entertainment, and Travel & Transportation.

[5] The next chapter is dedicated to explain data acquisition procedures, including getting contextual data around the lodgings (Section 4.1.2)

Category	Radius	#Venues
Food	500 meters	50
Art & Entertainment	1,500 meters	50
Travel & Transportation	1,500 meters	50

Table 3.2: Venue categories for PV.

The radius used in each category are intended to capture distances one would be willing to cover in order to reach each type of venue from the accommodation, as we explain in the subsequent sections. All the features that belong to the PV category are detailed in Appendix A.1.

3.2.1.1 Food Venues

The food category includes a range of venues such as restaurants, coffee shops, cafeterias, and other culinary places. The importance of food venues has been studied as a component of touristic involvement, which has been shown to influence the selection of travel destinations [Gross and Brown, 2008; Sparks et al., 2003]. Derived from these studies, we conclude that the proximity of food venues also increases the PV of a lodging as they greatly impact the travelers' overall experience.

To determine the radius used to retrieve food venues that increase the PV of our lodgings, we used a simple criterion based on the time and distance. We assumed that the presence of food places at five minutes walking distance from the lodging add value to the accommodation. Then we established an empirical five minutes walking distance, to be the maximum distance that a pedestrian would consider to transit, in order to reach a food venue. The average preferred walking speed [Browning et al., 2006; TranSafety, 1997] estimates that 500 meters would be covered between five to six minutes by most people, which is the motivation underneath the choice of a 500 meter radius for the food category.

3.2.1.2 Travel & Transportation and Art & Entertainment Venues

Commonly, tourists choose a convenient lodging according to the proximity of points of interest they would visit during their trip [Papatheodorou, 2001]. Similarly to the criteria employed for choosing the radius used for food venues, we determined that airports, train stations, historic sites, and other venues in this category, increase the perceived value of the lodging if they can be reached by a 15 minutes car ride, or by using the public transportation.

Commuting is understood as recurring and periodic travels (routine trips) that people do between two locations. It is most commonly used to define the trips between residence and working locations. Therefore, to estimate the average distance that can be traveled in 15 minutes, abstracting geographic, environmental, and specific urban factors, we rely on studies that investigate commuting [Sandow and Westin, 2010; Lee and McDonald, 2003] and governments commuting statistics (Australia[6] and USA[7]). After reviewing and testing various approaches to translate 15 minutes, to actual geographical distances, we employed each candidate radius to get various venues samples from the mapping services, and we concluded that 1,500 meters is a feasible and convenient radius to be traversed in 15 minutes by car or public transportation.

3.2.1.3 Distance-Based Features

Among the data that may be obtained from the mapping services, two main information are particularly useful, the distance from each venue to the lodging (Φ) and the number of check-ins (Ψ) that a given venue has. Venues that have greater popularity can be considered as iconic venues of their respective neighborhood, which may be an incentive for making tourists more willing to travel a longer distance, in order to reach the venue. If we consider the proximity of a venue as a factor that increases the value of an accommodation, we may state that just as psychological maps shape geographical boundaries in people's minds [Quercia et al., 2013], the perceived distance from the accommodation to iconic venues would have a different meaning than the actual physical distance. Therefore, we propose a normalized version of the geographical distance, using the check-ins the venue possesses as a normalization factor:

$$\widehat{\Phi} = \Phi/\Psi \, . \tag{3.1}$$

Finally, splitting the data by category, we aggregate all venues' information (Ψ, Φ, and $\widehat{\Phi}$) separately by their respective group to compute various statistics, as show in Table 3.3.

3.2.2 Perceived Risk (PR)

For online sharing economy lodgings, perceived risk is the degree of perceive all possible negative outcomes, derived from booking the lodging [Liang, 2015]. Viewed from the risk classification perspective proposed by Jacoby and Kaplan [1972], the notion of

[6] https://bitre.gov.au/publications/2015/files/is_073.pdf
[7] http://www.census.gov/hhes/commuting

Feature	Description	Feature	Description
Φ_{min}	Dist. to the closest venue	Φ_{max}	Dist. to the farthest venue
Φ_{mean}	Mean dist. to the venues	Φ_{med}	Median dist. to the venues
Ψ_{min}	Lowest # check-ins	Ψ_{max}	Greatest # check-ins
Ψ_{mean}	Mean # check-ins	Ψ_{med}	Median # check-ins
$\widehat{\Phi}_{min}$	Norm. dist. to the closest venue	$\widehat{\Phi}_{max}$	Norm. dist. to the farthest venue
$\widehat{\Phi}_{mean}$	Mean norm. dist,	$\widehat{\Phi}_{med}$	Median norm. dist.
$\|V\|$	# Venues (maximum 50)		

Table 3.3: Distance-based features.

perceived risk can be dissected in three different aspects, physical risk, performance risk, and financial risk. Physical risk is a considered as guests rarely know the host. Performance risk is the uncertainty that guests cannot experience the accommodation previously to their arrival. Finally, financial risk refers to the risks that guest may incur in cancellation fees.

The two first dimensions of risk, physical and performance, are tackled by Airbnb's quality service measures (ratings) and a popularity metric (review count). When an Airbnb guest books an accommodation and the booking is concluded, the guest is asked to give explicit ratings to evaluate particular attributes of the lodging: cleanliness, accuracy, value, check-in, location, and communication, also an overall rating (star rating) is shown on the lodgings' profiles. In addition, a review count is also provided, and these information (ratings and review count) are public at the corresponding lodging profile page. These collaborative reputation strategies alleviate the perception of risk and therefore are drivers of the risk perception.

On the other hand, the financial risk is dictated by the lodging's cancellation policy, which determines the monetary fee that the guest would incur in case of cancellation. Airbnb establishes four main cancellation policies[8] named flexible, moderate, strict, and no-refund, that respectively go from less to more aggressive measures, in order to handle the refund in case of canceling the reservation prior to arrival (one, five, seven days for the first three respectively), and grant a full refund for flexible/moderate policies, and a partial refund for the strict policy. Also, in less common circumstances, super-strict and longterm cancellation policies handle special cases, but are not considered in this dissertation due to the very small number of lodgings that fall under these cancellation modalities.

As in the previous section, we consider the importance of conceiving the lodging's features expressed in such a way that they leverage the context where lodgings are immersed. To this end, we normalized the review count r using other lodgings' review

[8]https://www.airbnb.com/home/cancellation_policies

Feature	Description
r	lodging's review count
Ω_{std}	Context review counts' standard deviation
Ω_{mean}	Context review counts' mean
$\widehat{r}$	Normalized lodging's review count

Table 3.4: PR features.

counts Ω, that are located at the surrounding area.[9] In statistics, the z-score (standard score normalization) is the distance, in terms of standard deviations from a given value (observation) to the mean (or sample mean). Therefore, we propose to use a set of review counts Ω from other lodgings at the surrounding of each accommodation. From Ω we are able to obtain a sample standard deviation and a sample mean that we employ to calculate the z-score version of the review count as shown in:

$$\widehat{r} = \frac{r - \Omega_{mean}}{\Omega_{std}} . \tag{3.2}$$

Such z-score can be interpreted as how much positively/negatively popular is the lodging when compared to its local peers.

Finally, the statistics computed in this section (shown in Table 3.4) are included in our model. To conclude this section, the full list of features that fall under this category are detailed in Appendix A.2.

3.2.3 Price Sensitivity (PS)

Price sensitivity is the extent to which the price of a product affects consumers' purchasing behaviors. Price is probably one of the most important factors, broadly and intuitively accepted to be a decisive motivator in consumers' behavior and intentions. Donthu and Garcia [1999]; Shankar et al. [1999]; Degeratu et al. [2000] offer a good compound of relevant works related to online price sensitivity. Price sensitivity may vary from one customer to another, according to the level of importance that each customer places on price relative to other purchasing criteria. It has been shown that customer's price sensitivity increases the more he or she is aware of the price dispersion of a given product [Degeratu et al., 2000]. Therefore, we argue that Airbnb users may present a high degree of price sensitivity, as a direct consequence of the information displayed on their web interface[10], which evidences the price dispersion, making it highly transparent to Airbnb users. The last statement concretely refers to the price

[9] Future sections detail the procedures to obtain 18 review counts of lodgings around any lodging in our dataset (Section 4.1.2), in order to compute the normalized review count.

[10] See *price range* `https://www.airbnb.com/s/New-York`

histogram at the Airbnb search tool (See Figure 3.2), which presents the dispersion of prices in an intuitive way. Originally, the histogram is intended to be used as a price filter. The filter is operated using two draggable slider buttons to set a custom price range that delimits minimum and maximum prices. Such histogram includes the mean price of the lodgings, aiming to guide the exploration of different prices.

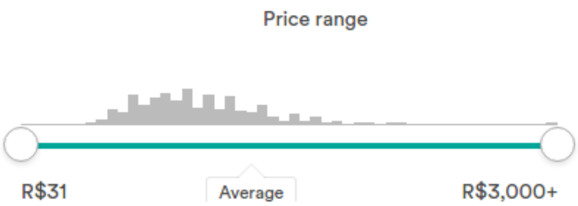


Figure 3.2: Airbnb price histogram.

By themselves, lodging prices cannot be perceived as *cheap* or *expensive* without their counterpart lodging prices. To obtain this vision of how a given price $l_\$$ stands among other prices, we propose two normalization that are similar to standard score normalization. Given a sample of nearby lodgings' prices $\mathcal{C}$,[11] where $\mathcal{C}_{mean}$ is the sample mean, and $\mathcal{C}_{std}$ is the standard deviation of $\mathcal{C}$. The first normalization is computed as shown in:

$$\widehat{\mathcal{C}}_\$ = \frac{l_\$ - \mathcal{C}_{mean}}{\mathcal{C}_{std}} . \tag{3.3}$$

In addition, assuming that the price mean $\mathcal{A}_{mean}$ displayed at the Airbnb price histogram is the true population mean. We propose to change the centering factor in equation 3.3 by $\mathcal{A}_{mean}$. As defined in:

$$\widehat{\mathcal{A}}_\$ = \frac{l_\$ - \mathcal{A}_{mean}}{\mathcal{C}_{std}} . \tag{3.4}$$

The price histogram draws a curve shaped by the lodging prices that are contained within the current search map. We assume that such histogram is built using all the lodgings contained in such area. In order to quantify the human interpretation of the histogram curve, we propose to compute the histogram's kurtosis and skewness. Skewness is a statistic [Doane and Seward, 2011] intended to score the degree of curve symmetry and kurtosis [Darlington, 1970] measures whether the data is heavy-tailed or light-tailed relative to a normal distribution. In addition, from the sample prices $\mathcal{C}$, useful statistics may be obtained, such as: lowest price, highest price, price standard deviation, price mean, kurtosis, and skewness as shown in Table 3.5.

[11] Section 4.1.2 details the procedures to obtain 18 lodgings' prices around any lodging in our dataset, in order to compute the normalized prices.

Feature	Description
$\mathcal{C}_{min}$	Lowest price
$\mathcal{C}_{max}$	Highest price
$\mathcal{C}_{std}$	Prices' standard deviation
$\mathcal{C}_{mean}$	Prices' mean
$\mathcal{C}_{kurt}$	Price kurtosis
$\mathcal{C}_{skew}$	Price skewness
$\mathcal{A}_{mean}$	Airbnb prices' mean
$\mathcal{A}_{kurt}$	Price kurtosis
$\mathcal{A}_{skew}$	Price skewness

Table 3.5: PS features.

Finally, in Appendix A.3 we show the list of features that shape the concept of PS.

3.2.4 Perceived Authenticity (PA)

In Airbnb, the concept of perceived authenticity was previously defined to be the extent to which the foreigner guest feels like natively living at the lodging's location where he or she stays. Guests solely rely on other guests' reviews to create their own authenticity perception. On Airbnb, when a booking transaction is concluded, guests are asked to write a public review at the respective lodging and host profile, and similarly, the host is requested to write a review on the corresponding guest's profile page. Hosts and guests are only able to see each other's comment when both have their reviews published on their profile, promoting the reinforcement of the reviewers' frankness.

We noticed that some reviews are very descriptive about discrete attributes of the lodging (e.g. cleanness, amenities, neighborhood characteristics). However, they are poorly informative about the guest's experience as a whole, which could be more useful to characterize the concept of authenticity we defined in past chapters. To mimic the way reviews are used by users to sense the authenticity experience of a lodging, we propose to use information retrieval techniques to create a similarity score between the reviews (documents) and a hand built authenticity lexicon (query). To this end, we use a language modeling (LM) approach [Zhai, 2008] with Dirichlet smoothing [Zhai and Lafferty, 2004].

Probabilistic LM are techniques widely used in information retrieval, their main idea is to estimate the probability $P(w|D_i)$ of a word w given document D_i. As a result, documents are ranked based on the likelihood of the query given each document language model. Based on such probability, documents that match various query terms obtain higher scores. Dirichlet smoothing is used in order to prevent scoring zero probability to unmatched terms, which allows to compute similarity scores that are interpreted as the probability of a word (query) to be important to the corpus of a text (document). In this dissertation, PA scores are computed using ElasticSearch[12] with standard text processing (Hunspell stemmer [Halácsy and Trón, 2006], lowercasing, and stopword removal) (Details in Appendix B.1).

Lexeme	Review
living	... I felt that I am **living** there ...
homey	... The apartment felt very safe and **homey** ...
true	... It's a **true** San Francisco experience ...
genuine	... heart warming, unique, **genuine** and inspirational time ...
experience	... stay here if you want a real SF **experience** ...

Table 3.6: Lexemes examples.

A lexicon is an inventory of semantic units called lexemes, commonly composed of a collection of words. The construction of our PA lexicon was done in two stages: exploration and word-selection. The exploration stage aims to first select reviews containing the word *experience*, which we presume may comprise useful information about how guests experienced the authenticity of the accommodation, and not merely discrete attributes of the room. Then, the word-selection stage consists of reading a set of reviews in order to find words that usually appear when users expressed a positive notion of PA. Table 3.6 shows a few examples of such reviews and the lexemes selected to be part of our lexicon. The lexicon can be seen in Table B.1, which is composed of a set of 33 words.

In Appendix B.2, are shown the three reviews with the highest authenticity scores, to illustrate the effectiveness of this approach, which satisfactory exemplifies how the lexicon is able to grasp and quantify a sense of authenticity expressed in textual reviews.

Feature	Description
Λ_{min}	lodging reviews' lowest authenticity score
Λ_{max}	lodging reviews' greatest authenticity score
Λ_{mean}	Mean of lodging reviews' authenticity scores
Λ_{med}	Median of lodging reviews' authenticity scores
Λ_{mean}	Skewness of lodging reviews' authenticity scores
Λ_{med}	Kurtosis of lodging reviews' authenticity scores

Table 3.7: PA features.

To characterize a lodging's degree of authenticity, we aggregate their reviews' scores, similarly to what we did in the previous sections, as shown in Table 3.7. Finally, the full list of features in this category is detailed at Appendix A.4.

[12]https://www.elastic.co/

3.2.5 Electronic-Word-of-Mouth (EWoM)

In previous chapters we explained the dynamics of how reviews are mutually cast between Airbnb entities involved in a booking transaction. When accommodations are offered in a peer-to-peer manner, usually guests write a review to hosts and their lodgings, and hosts reciprocally write a review to guests. On Airbnb, reviews are the most common manifestation of EWoM. Guests are aware that their personal opinion would help to frame the judgment of other users towards an approximation of the quality of the lodging. An opinion which expresses a positive attitude and satisfaction about the lodging would encourage the adoption of a positive feeling in other guests' minds [Tsao et al., 2015]. We assume that this guest-to-guest information flow plays an important role in the decision process of booking a lodging.

A classical task in sentiment analysis is to derive the polarity of a text, which is understood as a sense of the positiveness or negativeness of the sentiment expressed on a textual form. Sentiment analysis include techniques commonly employed for the treatment of opinions, sentiments, and evaluations from written language, leveraging computational resources [Pang and Lee, 2008; Liu, 2012]. In this dissertation, we are interested in measuring the sentiment polarity of our reviews, rather than investigate the best technique for such task. Therefore, we decided to employ a robust technique for polarity sentiment analysis, as to the best of our knowledge, there is no study that investigates the performance of sentiment analysis methods using Airbnb guest's reviews [O'Mahony and Smyth, 2009; Thelwall et al., 2012; Ribeiro et al., 2015].

Robustness is understood as the capability of generalizing the accuracy of some model, which is a property that is desirable when tackling unexplored domains [Ribeiro et al., 2015]. Vader [Hutto and Gilbert, 2014] is a robust state-of-the-art sentiment analysis technique, which extracts four polarity sentiment scores: positiveness, negativeness, neutrality, and sentiment compound. The compound score is computed by aggregating and normalizing the other 3 scores. It has been demonstrated that Vader has a consistent accuracy performance in different test collections [Hutto and Gilbert, 2014; Ribeiro et al., 2015], which makes it a good candidate for sentiment polarity detection to be employed to score Airbnb reviews.

The reviews that guests wrote on a lodging's profile page were processed using Vader. For the resulting set of scores coming from the reviews, we compute various sentiment features that aggregate the overall guests' feeling about a lodging, as we shown in Table 3.8. Appendix A.5 explicitly shows the complete list of sentiment statistics scores, that we constructed in this section, which we include in our feature-based model.

Feature	Description
Υ_{min}	lodging's Rev. Lowest sentiment score
Υ_{max}	lodging's Rev. Greatest sentiment score
Υ_{mean}	lodging's Rev. Mean sentiment score
Υ_{med}	lodging's Rev. Median sentiment scores
Υ_{mean}	lodging's Rev. Skewness of sentiment score
Υ_{med}	lodging's Rev. Kurtosis of sentiment scores

Table 3.8: EWoM features.

3.3 Contextual Learning

The term Learning-to-Rank (L2R) refers to the application of machine learning in order to build ranking models for information retrieval tasks [Liu et al., 2009]. Despite L2R merely tackles information retrieval problems, L2R has already been applied in RS [Karatzoglou et al., 2013]. In order to tackle the lodging recommendation problem we resort to learning to rank [Liu et al., 2009], which is used to leverage the contextual models for lodging recommendation we presented in previous sections.

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 3.1, 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 [Wu et al., 2008], a gradient boosted regression tree learner, which represents the current state-of-the-art in L2R [Chapelle and Chang, 2011]. LambdaMART uses gradient boosting trees [Friedman, 2001] to directly optimize ranking evaluation metrics as cost functions. Boosting leverages an ensemble of weak learners to create a unified strong model. Boosting techniques initially train a weak learner, then, iteratively subsequent models are created and added in order to correct the prediction errors, until some accuracy is achieved. By training an ensemble of regression trees, LambdaMART outputs scores used to sort documents in order to compose a ranking. In addition, the choice of LambdaMART is convenient given its capabilities to automatically selecting features that best improve the construction of its trees, which is an advantage over some other learners. RankLib [Dang, 2013] is a set of tools for L2R that has an implementation of LambdaMART, which is used in this dissertation.

3.4 Summary

In this chapter, we presented our proposed feature-based model, inspired by related theories of the sharing economy. Five sets of features, denominated as preference dimensions, were created to quantify users' repurchase intention, namely: price value, perceived risk, perceived authenticity, electronic-word-of-mouth, and price sensitivity. These features were built leveraging techniques of natural language processing, information retrieval, and feature engineering to describe how accommodations are immersed in their surroundings. Also, the recommendation problem is tackled as a ranking problem, employing LambdaMART, a state-of-the-art L2R technique in information retrieval, capable of leveraging the features we created.

In the following chapter, we detail the procedures we followed to build a test collection suitable to evaluate lodging recommender systems for the sharing economy.

Chapter 4

Data Acquisition

In this chapter we explain the procedures undertaken to build the test collection we use in our experiments. Also, we provide a characterization of the collection, comparing two cities of interest in this study. The data was collected during the period comprised between March 2016 to September 2016 with a crawler distributed in 10 machines. Airbnb[1] is a web application that permits users to list, search, and rent lodgings, enabling guests to benefit from locals' advice, having a genuine cultural exchange and a unique traveling experience. The dataset presented in this chapter is mainly comprised of public information available from Airbnb's lodgings, their reviews, and their users' profiles. This collection aims to simulate a user traveling and seeking for lodging in one of the two target cities we chose: New York, United States (NYC) and London, England (LON). This dataset constitutes a contribution to the recommender systems domain and was built to permit the evaluation of recommender systems for sharing economy platforms, with a particular focus on lodging recommendation.

4.1 Data Collection

4.1.1 City Macro-areas, Micro-areas, and lodgings

In order to explore places within the target cities where lodgings are located, we introduce the concept of macro-areas and micro-areas, which are suburban regions located in the cities. In conjunction, macro-areas and micro-areas aim to reproduce a representative overview of the true lodging supply in the city. Each micro-area belongs to a single macro-area, and macro-areas are composed of multiples micro-areas. These two concepts are similar to the concept of *neighborhoods* in an abstract way.

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

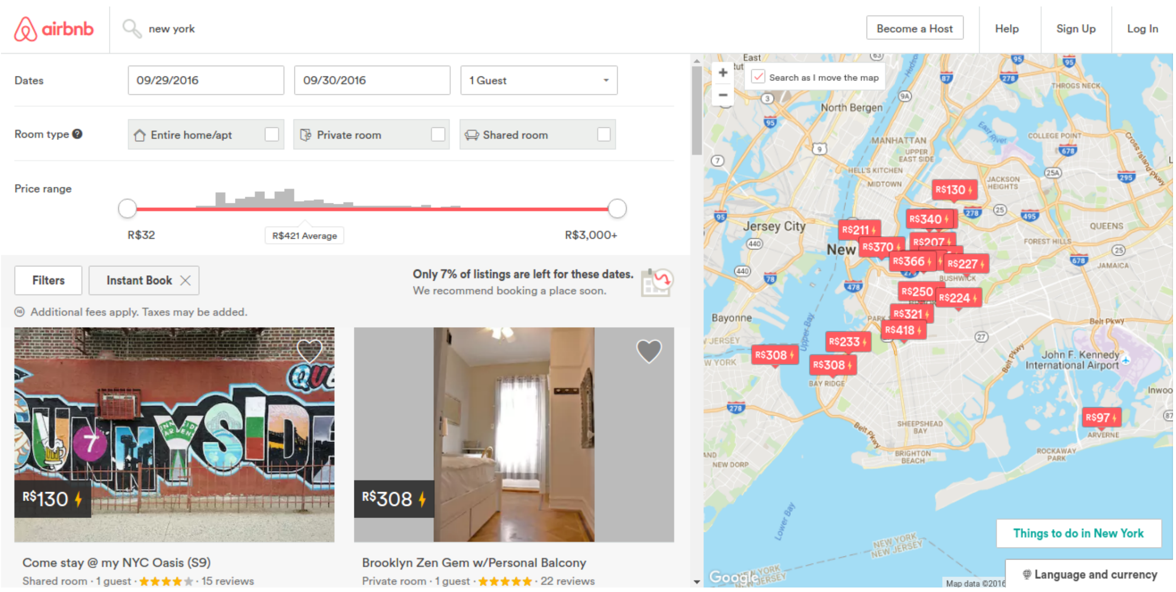
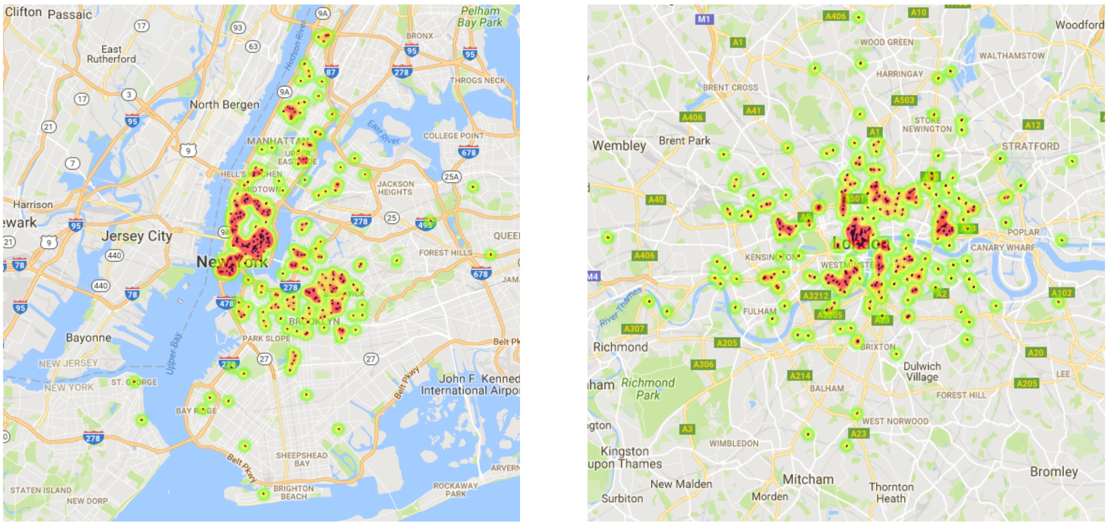


Figure 4.1: Airbnb's search tool.

Airbnb's search tool enables guest users to find lodgings that other hosts users rent on Airbnb (see Figure 4.1). The search tool is composed of a ranking list on the left-hand side and an interactive map on the right-hand side. To guide the selection of lodgings in our collection, we used Airbnb's search tool to retrieve accommodations from NYC and LON. By querying the search tool for either NYC or LON, the map automatically is positioned in a way that it covers the entire city. As an incognito user (a user with no explicit identity or booking information) we fetched a set of 300 lodgings[2] from the Airbnb's website, where the sole information that was given to the search tool was the name of the city. From this initial set of 300 lodgings in each city, NYC and LON, we only use their geo-location. Their emplacement can be viewed in Figure 4.2. Each one of these geo-location represents a macro-area (300×2 in total).

We could eventually have chosen arbitrary points in the map that are uniformly distributed across the city. Instead, the intuition behind our approach is that by querying Airbnb's search tool as an incognito user, we forced the Airbnb recommender engine to build a non-personalized recommendation of lodgings. This induces Airbnb to adopt a diversification strategy to recommend lodgings that best suit a broader set of macro-areas. As a consequence, this methodology has two main implications: (1) the search tool deliberately displays lodgings located at places commonly searched or booked by other users, and (2) it reflects Airbnb's lodging supply density in the city, arguably producing more realistic recommendation scenarios, rather than employ

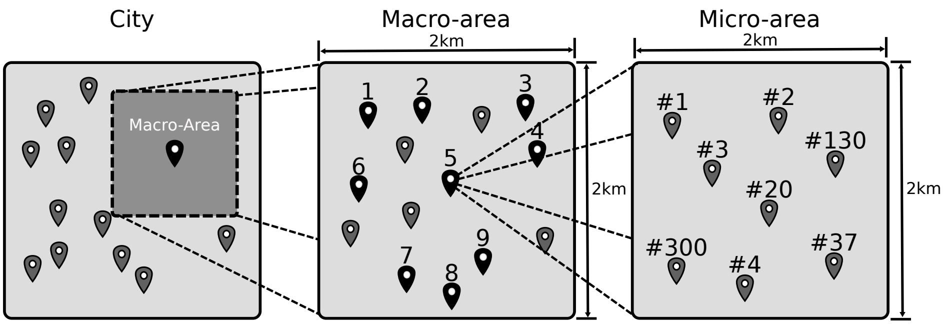
[2] Airbnb allows to retrieve at maximum 300 lodgings from their search tool, if more lodgings are needed one have to zoom-in to get different lodgings by restricting the area covered by the map.



(a) 300 macro-areas in New York. (b) 300 macro-areas in London.

Figure 4.2: Distribution of macro-areas across the city.

a simpler criteria to guide the selection of macro-areas, for instance grid sampling. This implies that we have many macro-areas gathered closely together where users commonly tend to seek for lodging (for instance downtown or touristic places), while also covering a diverse set of other suburban locations detached from the mainstream, as observed in Figure 4.2.



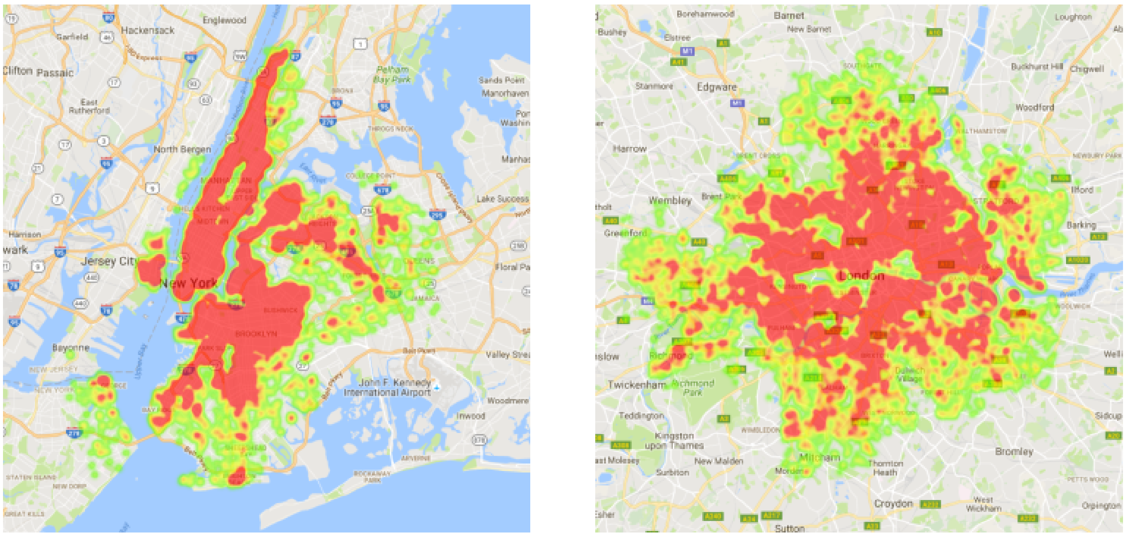
(a) 300 Macro-areas within the city.
(b) Selection of 9 random geo-locations out of 300.
(c) 300 ranked lodgings.

Figure 4.3: lodgings' sampling procedure.

One at a time, each of the 300 macro-areas are used to zoom-in Airbnb's search map to cover a two kilometers square region centered at the geo-location of a macro-area, as illustrated in Figure 4.3 (a). After centering the map at the more specific suburban region, 300 new lodgings are retrieved where we randomly selected nine of them. Their geo-location represent the locations of nine micro-areas, as seen in Figure 4.3 (b). Once again, we individually center the map at each of the nine locations (micro-areas), in order to retrieve a set of 300 new lodgings. At this time, the lodgings retrieved represent the actual micro-area, as illustrated in Figure 4.3 (c).

This process was followed with the purpose to get a more legitimized representation of macro areas, as they would be represented by nine randomly chosen regions as opposed to a single arbitrary micro-area. Notice that by using the Airbnb's search tool to retrieve the lodgings, we obtain a non-personalized ranking.[3].

Finally, Figure 4.4 shows the lodgings' density across the target cities we obtained in our collection. As observed in the figure, we obtain a high coverage of the geographic perimeter of both cities.



(a) New York (17,325 lodgings). (b) London (22,134 lodgings).

Figure 4.4: lodgings' coverage at the target. cities.

For all of the lodgings in the dataset we crawled their profiles, which contain detailed information about their attributes and the reviews left from the guests that visited the respective lodging. In Table 4.1 we show a summary of the number of lodgings per target city and the number of visitors obtained from the reviews found on the lodgings' profile pages (Table 4.1, #lodgings and #Reviews).

	#lodgings	#Reviews	#Guests
NYC	17,325	250,508	219,915
LON	22,134	266,743	223,106
Total	39,459	517,251	436,109*

Table 4.1: Test collection summary (* Unique items).

[3]This observation will be important later during our experimental setup (Section 5.5.2)

4.1.2 Context-based Features

The feature-based model we propose, CLLR, uses multiple pieces of information in order to compute normalized lodging's attributes and build a vision of the lodgings' surrounding. In this section, we detail the acquisition of the information we employed to create such features, that is obtained from two main sources: the Airbnb Search Tool and the Foursquare API.

In Section 3.2.1, we explained a group of features that aggregate venues' information (check-ins and distances to the venue), which is feasible to obtain from a mapping service. Foursquare is a Web and mobile App that allows users to discover and explore the venues of a city. It is a search and recommendation engine for users to search, evaluate, like, comment and check-in on venues. Foursquare provides an API that grants access to venues' visiting statistics and descriptive information. Many of their API allows to query for a geographical coordinate to get surrounding venues within a delimited customizable radius. Using the lodgings' geo-location, we use the API to fetch information from three main categories and their radius: Food (500m), Art & Entertainment (1,500m), and Travel & Transportation (1,500m), as previously defined during the creation of our model.

In Section 3.2.2 and Section 3.2.3, we described multiple ways to employ surrounding information to normalize lodging's attributes (review counts and prices) and also compute contextual statistics (e.g. Airbnb mean price, mean price of a sample of lodgings, price histogram kurtosis). Such surrounding information is obtained from the Airbnb search tool, by retrieving samples of lodgings within 2 kilometers and their information. By centering Airbnb search map at a lodging's geo-location, we retrieve the first search page containing 18 Airbnb lodgings of the first result page with their basic information, from which we collect their prices and review counts. Such prices and review counts serve to normalize the corresponding lodging attributes. Also, at the same result page, we collect the price histogram, which is also used to compute various features.

4.1.3 Guests' History

Each Airbnb user has a public profile that is composed of personal information, such as name, nationality, profile picture, the number of reviews obtained from other users, and other general information. When two users interact by a booking transaction one assumes the role of a guest and the other the role of a host. When a transaction is concluded, Airbnb allows both users to mutually write a review that is shown on each other's profiles and the lodging profile of the booking.

In the crawling methodology described in previous sections, we collected the reviews found at the lodgings we sampled from the target cities (see the total number of reviews in Table 4.1, #Reviews). The profile pages of such guests (reviewers) contain information to recover their booking transactions on Airbnb. Therefore, we built a crawler to get guests' profile pages and the 10 most recent reviews[4] left by hosts (Figure 4.5 (a)). Each of the host reviews found at the guests' profiles correspond to a booking transaction. However, the corresponding lodging is not displayed on the guest's profile. Consequently, the booking transactions have to be reconstructed in order to find the associated lodgings. We refer as *guest history* to the set of accommodations that a user booked on Airbnb as a guest.

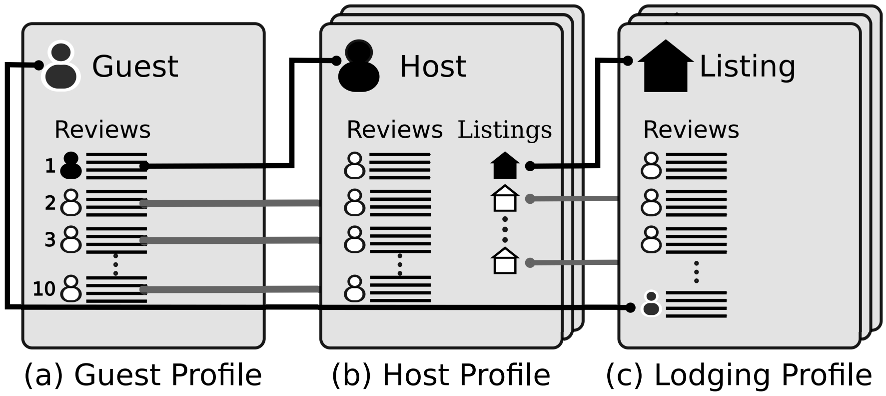


Figure 4.5: Recovering methodology used to build guests booking history.

The process we used to reconstruct guests' history starts by taking the reviews that hosts wrote to guests to retrieve their corresponding hosts' profiles (Figure 4.5, (a) and (b)). From the hosts' information that we crawled we obtain the lodgings they rent on Airbnb, from which we are allowed to obtain the corresponding lodgings' profile pages containing the reviews that guests wrote to them (Figure 4.5, (c)). Therefore, we are able to find the reviews belonging to the guests that we are interested in, hence recovering their booking transactions, which leads to discovering the accommodations they booked.

Notice that in this reconstruction process if a lodging no longer exists and it is not shown on the host's profile, it is unfeasible to recover such lodging with this methodology. Also, if the guest did not leave a review on the corresponding lodging he or she booked, we would not be able to reconstruct that specific item neither. Both limitations would lead to lose items contained among the ten most recent booking transactions that we originally crawled from guests' profiles (See Section 4.2.1 for a completeness characterization).

[4]By default the first 10 reviews are displayed at the first page of their profile

Although in this reconstruction methodology we limited the number of 10 host reviews per guest profile to be considered, it is possible to obtain more than the ten most recent booking transactions of a guest for multiple reasons.

For instance, suppose that two of our guests, u_j and u_k, share a lodging i between them in their guest booking history, however, lodging i is not between the ten most recent transactions of guest u_k. In such case, we only attempt to recover guest u_j's transaction corresponding to booking the lodging i. Then, retrieving all the reviews belonging to i we would obtain evidence to recover guest's u_j's booking history. Additionally, after having retrieved all reviews that lodging i possesses, we would also have evidence to recover the booking of lodging i by guest u_k, resulting in recovering an extra item for guest u_k than we would have expected.

Another case when we could recover items not expected between the ten most recent reviews of a guest u, is when u visited n times a given lodging i. Let's denoted by $\mathcal{R}_i$ the set of reviews belonging to the lodging i. When we obtain all reviews $\mathcal{R}_i$, as u visited n times i, we would have recovered multiple history items at once that evidence that i was booked by guest u a total of n times $r^{(1)}, ..., r^{(n)} \in \mathcal{R}$. And if the booking dates of $r^{(1)}, ..., r^{(n)}$ were not covered by the ten most recent booking transactions of u, we could eventually have recovered more than 10 items in u's guest history.

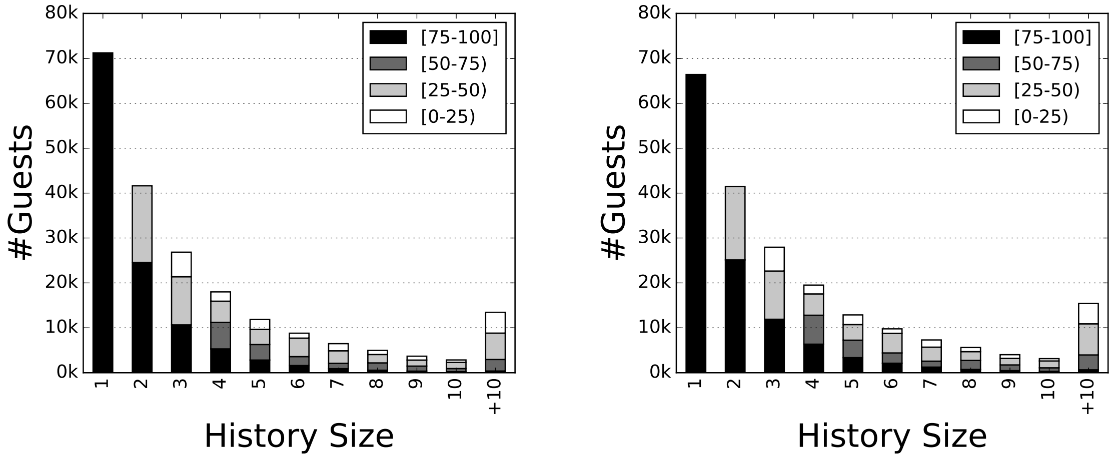
By definition, any guest in our collection has at least one lodging reconstructed in his or her booking transactions in one of the target cities.

4.2 Data Characterization

Over the following sections we present a brief characterization of the data collected in this chapter, aimed to offer an overview of the completeness, coverage, and main features contained in our dataset. We usually compare the data corresponding of the lodgings in NYC (≈17k lodgings) and LON (≈22k lodgings) in contrast to the lodgings in other parts of the world (≈480k lodgings). This characterization would also serve as companion documentation to guide users with the usage of the dataset.

4.2.1 Completeness of Guests' Booking Transactions

As a trust mechanism, Airbnb users' profiles have a review count, which is incremented by one unit each time a user receives a guest/host review from another user. If a user has both guest and host roles, we cannot determine the number of reviews that corresponds to each role, and therefore this review count cannot be taken as an exact approximation of a guest's history size. However, it may be employed as an upper bound of the total number of booking transactions that users made as a guest. Notice that for pure guest users (users that ever rented a lodging) this review count exactly matches their guest's history size.



(a) NYC guest profiles (219,915 guests). (b) LON guest profiles (223,106 guests).

Figure 4.6: Reviews in guests' profiles.

For all guests, we know their review count as it is public on their profiles. Using this upper bound approximation of the guest's history size, we are able to estimate the percentage of items recovered from all historic transactions. Assuming that the review count is equal to the guest's history sizes, Figure 4.6 shows the distribution of guests (Y axis) per review count (History Size, X axis) where the height of the bar indicates the number of guests that have the corresponding history size. The different shades of gray in each bar indicate the percentage of items recovered from these guests, where dark indicates 100% recovered and white less than 25%. As we observe, the majority of users have history sizes of less than 4 items and for these users this plot gives full detail of the number of booking transaction we failed to recover (4 colors, each one for 1 to 4 items missing). Notice that increasing the history size leads to have more items missing (not recovered), however, in most cases we reach to get at least 50% of the total guest's booking transactions across different history sizes.

Figure 4.7 emphasizes the temporarily distribution of the publication dates of the reviews of our entire dataset. The area plots depict a growing booking tendency, that bursts at the early stages of 2014. The two curves show a comparison of the number of bookings at the target cities versus the rest of the world. As we see, both fluctuations are correlated and are notoriously accentuated in the more recent years, with an even more preeminent pick at the target cities in the last years.

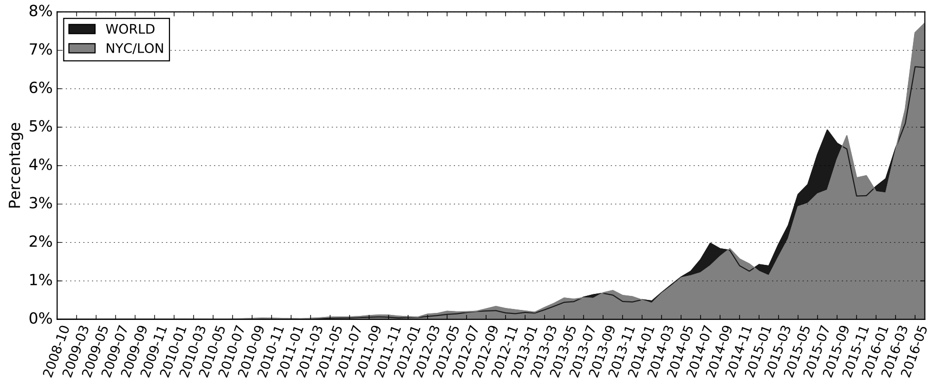


Figure 4.7: Review's creation dates.

4.2.2 Data Coverage

In this section we estimate the coverage of our data collection in terms of the estimated size of the entire Airbnb website. Based on official information [5] we are able to roughly estimate the percentage of the Airbnb data we collected. In addition with the last information, Airbnb's users, lodgings, and reviews, possesses unique IDs, which are composed of an integer number. We noticed that these IDs are incrementally generated. For instance, if we order the reviews by their creation date we obtain the same order when sorted by their IDs (both in ascending order), which validates our assumption. The same behavior was reproduced to lodgings' and users' profiles, which we know their ID and creation date. Consequently, the greatest ID that can be found on Airbnb is an upper bound of to total instances created on their site. Therefore, combining these information we were able to estimate the total percentage of the data we collected, as showed in Table 4.2.

	#lodgings	#Users	#Reviews
Collection	525,780	9,260,093	15,701,718
Total Airbnb	2,000,000*	60,000,000*	49,225,908*
Collection Percentage	26.3%	15.4%	31.9%

Table 4.2: Estimated collection completeness of the entire Airbnb (* Estimated Numbers)

Nevertheless, the total number of reviews ,left by guests to all lodgings in Airbnb, is difficult to estimate. The best estimation we were able to make was to multiply the number of lodgings that Airbnb possesses, with the mean of the number of reviews, obtained from our collection, which gives an approximation of 50 million of reviews.

[5] https://www.airbnb.com/about/about-us

4.2.3 lodging Characterization

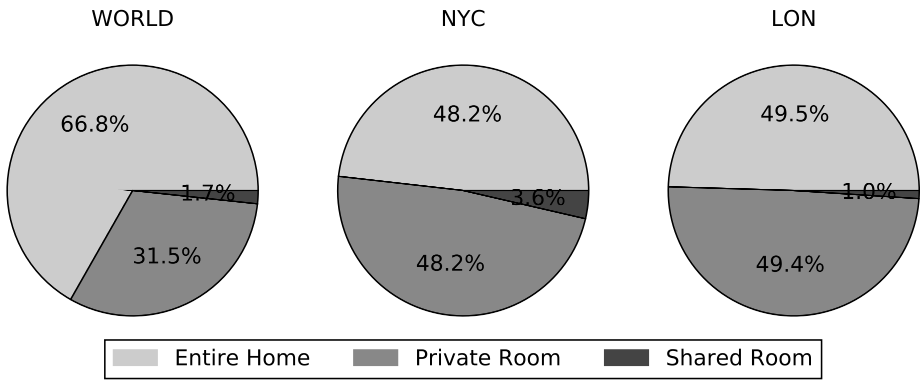


Figure 4.8: Room types.

Figure 4.8 presents the percentage of lodgings that fall under the Airbnb's room type categories, named entire home, private room, and shared room. The accommodation supply in NYC's and LON's is fairly balanced between entire homes and private rooms, despite entire homes being the predominant room type category in other places (WORLD). We also noticed that NYC has an atypically percentage of shared rooms, more than the double compared to the rest of the world.

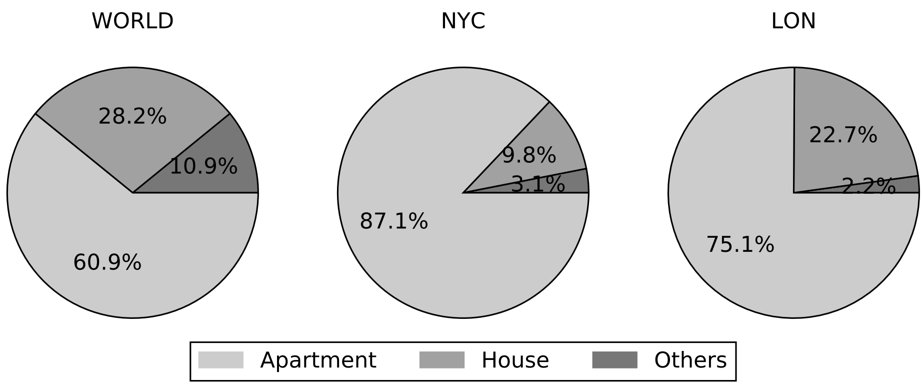


Figure 4.9: Property types.

In Airbnb the property type covers a wide categorization of housing types that can range from castle and tree house to more conventional types, for instance houses and apartments. In Figure 4.9 we present the percentages of lodgings in Airbnb of the house and apartment property types against the rest of the categories to illustrate their predominance of the apartment category. We also notice that the house type is particularly rare for lodgings in NYC.

Following the same categorization of property types, in Figure 4.10 we present the mean price in US Dollars, for each of the property types described before with

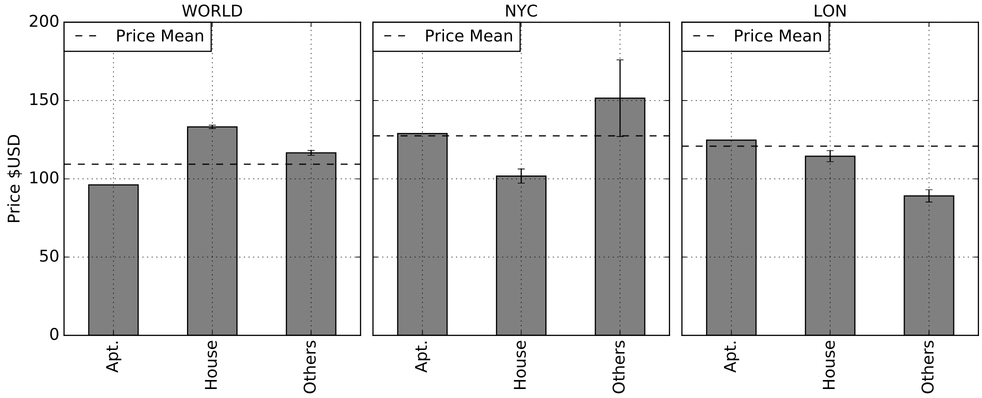
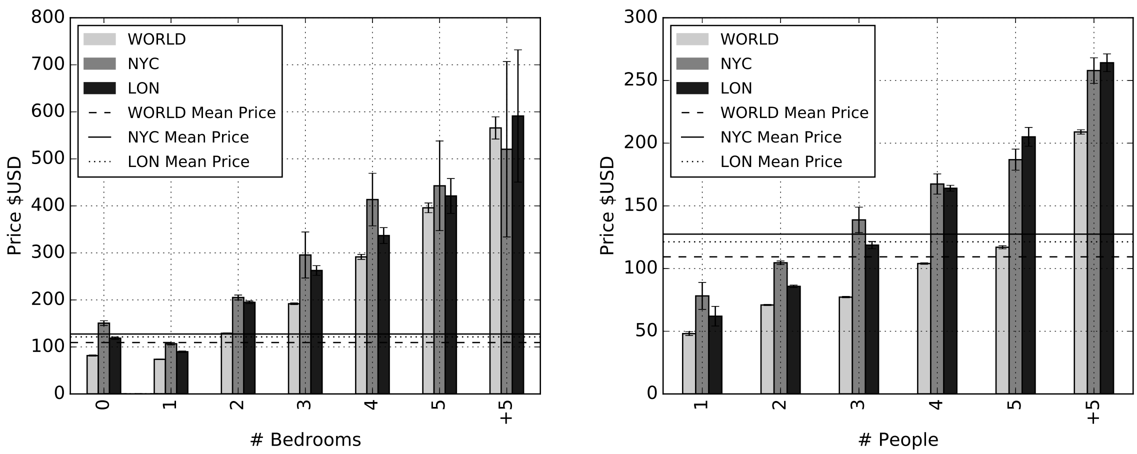


Figure 4.10: lodgings' prices by their property type.

their respective 95% confidence interval and the overall price mean. In such plot we perceive that apartment prices in NYC and LON are more expensive compared to the rest of the world, which is the opposite (houses being more expensive) at other places. An extended visualization of the prices with respect to the number of people (that the lodging can hold) and the number of beds the lodging possess can be found at Figure 4.11.



(a) lodging's price by number of bedrooms

(b) lodging's price by room's capacity

Figure 4.11: lodging's price by space attributes.

In Figure 4.12 is shown a compound of the most common amenities offered by Airbnb lodgings. In this figure we see that the great majority of Airbnb lodgings provide free wireless Internet and allow the usage of their kitchen. Also, we noticed that the usage of TV is restricted/missing in 40% of the lodgings at the target cities, contrary to the rest of the world where it is the 4^{th} most common amenity. As well, the sharing usage/availability of the washer machine in lodgings in NYC would not be

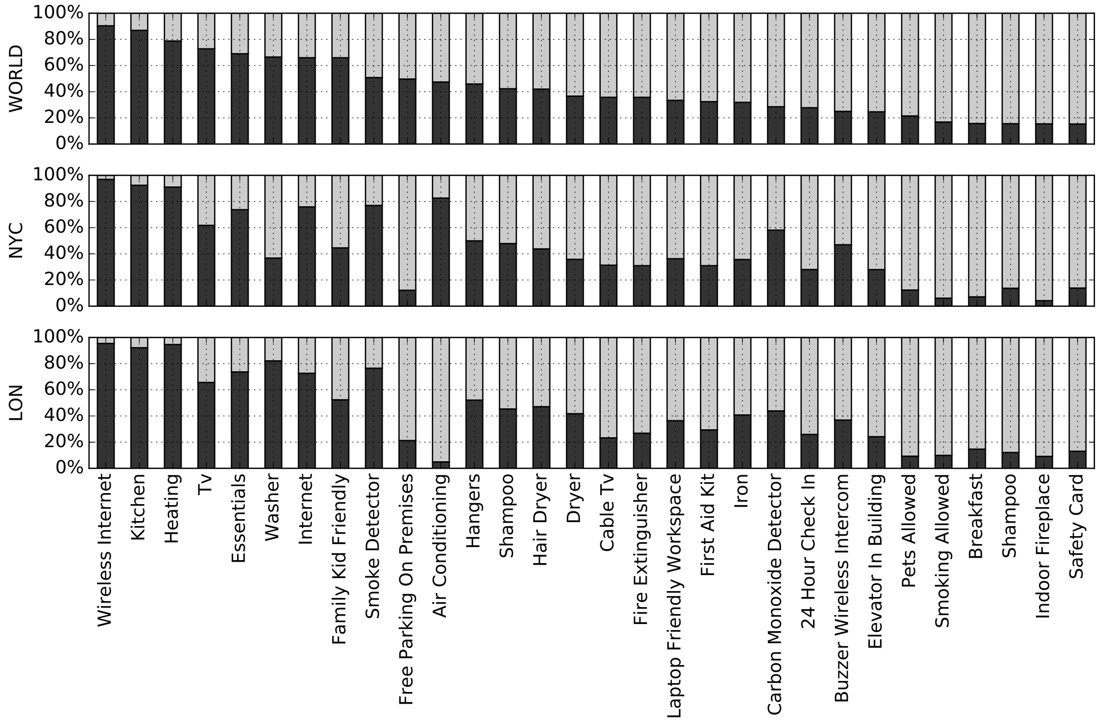


Figure 4.12: Percentage of lodgings offering amenities.

permitted or is available. Furthermore, NYC and LON offer less kid friendly lodging options than lodgings in other locations in the world. We also assume that due to the superiority of the apartment room category in NYC and LON (Figure 4.8), smoke detectors, carbon monoxide detectors, and buzzer are common amenities in these cities, which again explains the reason to have few lodgings offering parking spots compared to other cities.

As for many goods found on Internet the frequency of purchases, views, and ratings are long tailed [Anderson, 2006], which means that some items may posses high popularity than usual. In this lodging domain, the number of reviews per lodging is also long tailed, as we can see in Figure 4.13, where we see a greater number of lodgings with few reviews and substantially a great number of lodgings with outstanding review counts.

Violin plots of the number of bedrooms, beds, bathrooms, person capacity (maximum capacity), and guests included (default number of guests included) are shown in Figure 4.14. Violin plots are comparable to box plots, their main difference is that they drawn the probability density of the data, allowing to show multiple peaks, including a marker indicating the position of the median (white dots).

A visual examination of Figure 4.14 permits to derive the most common accommodation configuration for NYC and LON, which is typically a single bathroom,

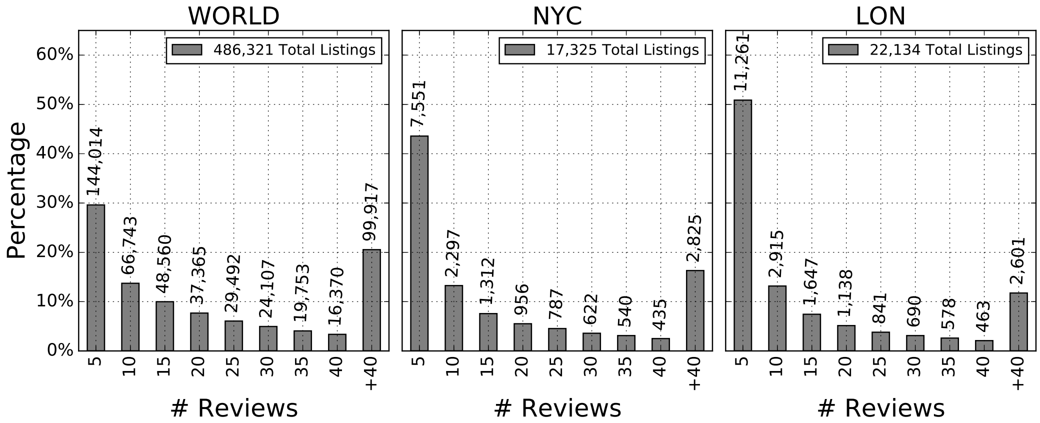


Figure 4.13: Number of reviews per lodging.

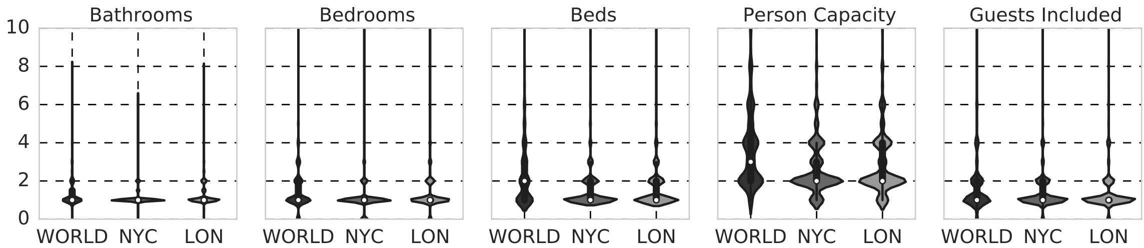


Figure 4.14: Space configuration.

providing 1 or 2 bedrooms/beds, in places that often accommodates up to maximum 4 persons, which are aimed to receive 1 to 2 persons. While in the rest of the world Airbnb's lodgings seem to have a greater capacity to hold more persons if necessary, with the disposal of more bedrooms and beds, holding more than 6 guests, which we presume is evidence suggesting that Airbnb lodgings are more spacious in other cities.

The anonymity policy of ratings in Airbnb does not permit to display the rating that each user gave to a lodging, therefore rating scores are only displayed when the room has been booked and rated by multiple users. In Figure 4.15 and 4.16 we show the rating scores that lodgings posses. We notice that ratings are biased towards high scores and that a considerable number of lodgings do not posses ratings. Also, we noticed that lodgings with no scores are particularly high at the target cities.

4.3 Summary

In this chapter, we detailed the procedures followed to build a data collection aimed to evaluate recommender systems in the lodging domain and more specifically in the sharing economy. We discussed the crawling steps we followed to sample lodgings at New York (United States) and London (England) and we explain the methodology

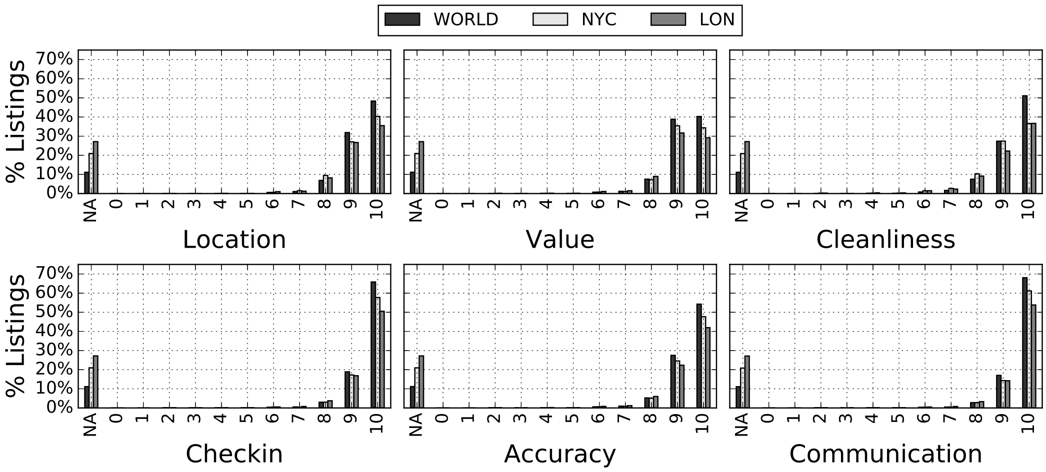


Figure 4.15: Ratings.

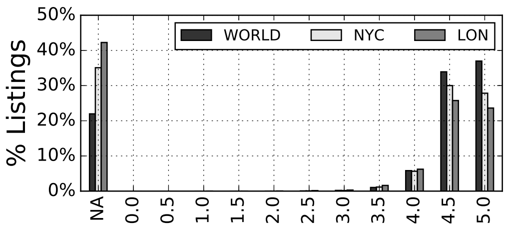


Figure 4.16: Star rating.

employed the recover guests' historical booking transactions to be able to build users' profiles. We also estimated the size of our data collection with respect to the entire Airbnb site and we conclude with a characterization of the dataset.

In the following chapter we expose a methodology framework that permits evaluating the performance of recommender systems in the lodging domain.

Chapter 5

Experimental Setup

In this chapter we present a comprehensive off-line evaluation using the data collected in Chapter 4, which is a test collection contribution from the author, which is intended to facilitate the evaluation of recommender systems for the lodging domain in the sharing economy. This evaluation framework serves to assess the performance of the model we proposed in this work, CLLR, which is a model inspired by theories of users' repurchase intention in the lodging sharing economy domain.

Typically, travelers engage an exploration process using the search tool of a particular lodging provider, to our scenario this is explicitly a search map,[1] which users operate to find a convenient accommodation in a particular suburban region that we denominate with the term macro-area. This browsing nature is considered in our experimental design, therefore, we conceive users as a geographical query, which mimics the action of centering the search map on the desired region that the traveler aims to sojourn, and the recommendation task is to suggest accommodations at the given geo-location.

5.1 Problem Definition

In this dissertation, the recommendation task is reduced to a ranking problem. Concretely, a test case is a tuple $\langle u, l, t, \mathcal{I}, i^* \rangle$, where u is a target user, l is the location where the user 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. Notice that for any tuple $\langle u, l, t, \mathcal{I}, i^* \rangle$, there is only a single relevant document $i^* \in \mathcal{I}$. Therefore, the

[1] Airbnb search map `https://www.airbnb.com/s`

ground truth of the sorting is a ranking that places i^* as the first suggestion to be recommended to the user u, as indeed, i^* was the item preferred by the user u.

5.2 Test Collection

Our test collection is comprised of multiple test cases that are instances of the problem we previously defined. Test cases are chosen randomly from all the bookings contained in our data collection. For simplicity, the term *test case* is used to refer to the tuple $\langle u, l, t, \mathcal{I}, i^* \rangle$, and is composed of:

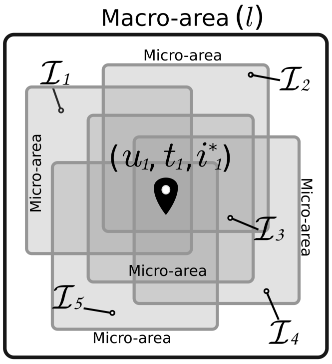
1. A target user u.
2. An input geo-location l.
3. A booking date t of the visit.
4. A set of lodgings $\mathcal{I}$ at maximum 2km from l, which includes a target lodging $i^* \in \mathcal{I}$.
5. The target lodging i^*

The lodgings in $\mathcal{I}$ posses relevance scores Y, associated with each lodging $i \in \mathcal{I}$, where $y_j = 1(i_j = i^*) \forall y_j \in Y$ and $1()$ is the indicator function, that scores a relevance of one for the target lodging i^* and zero otherwise.

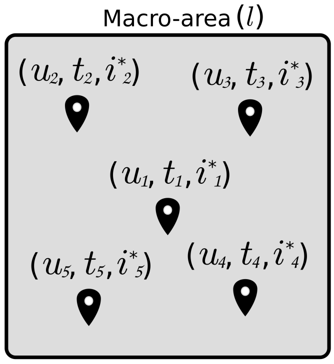
5.2.1 Geographical interpretation of test cases

The term macro-area is a concept that refers to a geographical area that can be compared to a neighborhood and it is composed of nine micro-areas, which are also suburban regions that comprise lodgings. These concepts were introduced in early chapters (Chapter 4.1.1). Macros-areas and micro-areas, with the lodgings they contain constitute part of our data collection. In order to simulate the exploration nature of a user searching for accommodation on a given location l, we use the lodgings contained in micro-areas to replicate the test case varying the candidate set $\mathcal{I}$. Analogously, the term candidate-set and micro-area refer to the same concept: lodgings gathered around a given location.

Figure 5.1 (a) illustrates a macro-area, which is intended to represent the input location l. To have a more robust evaluation, and rely on statistical tests, five simulations of the user u seeking for lodging in l are performed. In Figure 5.1 (a), we



(a) 5 micro-areas $\mathcal{I}_i$.



(b) 5 target guests.

Figure 5.1: Macro-area with overlapping micro-areas.

illustrate a user u_1, with five test-case simulations, where the only variant between them is the permutation of the micro-area $\mathcal{I}_k$ (candidate sets). Notice how the micro-areas $\{\mathcal{I}_1, \dots, \mathcal{I}_5\}$ contain the target lodging i^*. Despite being five different test cases, they all share the same user u_1, booking date t, and target lodging i^*, such as $i^* \in \mathcal{I}$. As we previously mentioned, we aim to simulate a user seeking for lodging, where the macro-area (neighborhood) is explicitly informed to the recommendation engine. These five distinct test cases portray a user using the search interface to browse and explore the neighborhood l (macro-area), where he or she is intended to sojourn (Figure 5.1 (a)). This methodology is not merely convenient in a way that it permutes the items contained in the candidate set (shaping a broader coverage of diverse scenarios), but also, it avoids positioning the target lodging exactly at the same location with respect to the micro-area itself (Notice how i^* is not always exactly at the center within micro-area $\{\mathcal{I}_1, \dots, \mathcal{I}_5\}$, Figure 5.1 (a)). All the five simulations are included in our experiment as independent test cases.

On the other hand, our test collection includes 25 user simulations for each macro-area, which is the product of simulating five different target users, as illustrated in Figure 5.1 (b). Our data collection comprises 300 use cases at each of the two target cities, completing 15,000 test cases, decomposed in 7,500 test cases per target city (25 test-cases/macro-area $\times$ 300 use cases). All these numbers are summarized in Table 5.1.

5.3 Training and Test Procedure

The partition of the test collection in multiple subsets is a common practice found in many recommender systems to evaluate top-N scenarios. The last practice aims to use the partitions in order to assemble training and test sets, which are employed

City	#Test Cases	#Macro areas	#Target Guests	Unique Guests
NYC	7,500	300	1,500	1,414
LON	7,500	300	1,500	1,434
Total	15,000	600	3,000	2,847

Table 5.1: Test cases summary.

to accurately evaluate recommender systems [Cremonesi et al., 2010; Deshpande and Karypis, 2004; Hurley and Zhang, 2011]. Cross-fold validation [Devijver and Kittler, 1982; Kohavi et al., 1995] is probably one of the most standard evaluation procedures, which consist into partitioning the data into various subsets, allowing to iteratively combine them to compose variations of the training and test sets, in such a way that it allows to repeatedly perform evaluations, with the purpose to estimate the performance of the model as the average of the results across the different evaluation rounds. Such procedure has the goodness to create robust models to over-fitting.

On the other hand, other methodologies partition the test collection into training and test sets ruled by the temporary order of the data, where the information contained in the train set is composed of events that occur earlier than the events in the test set. Evaluation frameworks had implemented sliding time windows [Cheng et al., 2016; Ahmed et al., 2013; Matsubara et al., 2012] aiming to grant the concession of valid data to be employed in the training, according to certain time constraint, which keeps the evaluation congruent with respect to the natural sequence that events occurred, keeping the test data ahead in time with reference to the training data.

Then, combining the methodologies previously discussed, we sort the test cases $\mathcal{T}$ in our test collection by their booking date, in ascending order, and we split the test collection into 12 folds $\mathcal{T} \equiv \bigcup_{i=1}^{12} t^{(i)}$ of equal size (same number of test cases) $\left|t^{(i)}\right| = \left|t^{(j)}\right| \; \forall \; i, j \in [1, 12]$, such that $t^{(i)} \lhd t^{(i+1)}$, meaning that the booking dates of the test cases in $t^{(i)}$ contain older dates than the booking dates of the test cases contained in $t^{(i+1)}$, as illustrated in Figure 5.2, which shows the temporal coverage of each of the folds.

Next, we create height time windows, that are composed of five adjacent folds such as $w^{(i)} \equiv \bigcup_{i}^{i+4} t^{(i)}$. Within each time window $w^{(i)}$, the first four folds $t^{i+j} \forall j \in [0, 3]$ are used to train the model using 4-folds cross-validation, and the last one $f^{(i+4)}$ is used to test the performance of the model. Figure 5.3, illustrates the sliding window, with the respective 4-folds cross-validation and the test fold.

In this evaluation any feature derived from lodgings' review count (PA 3.2.4,

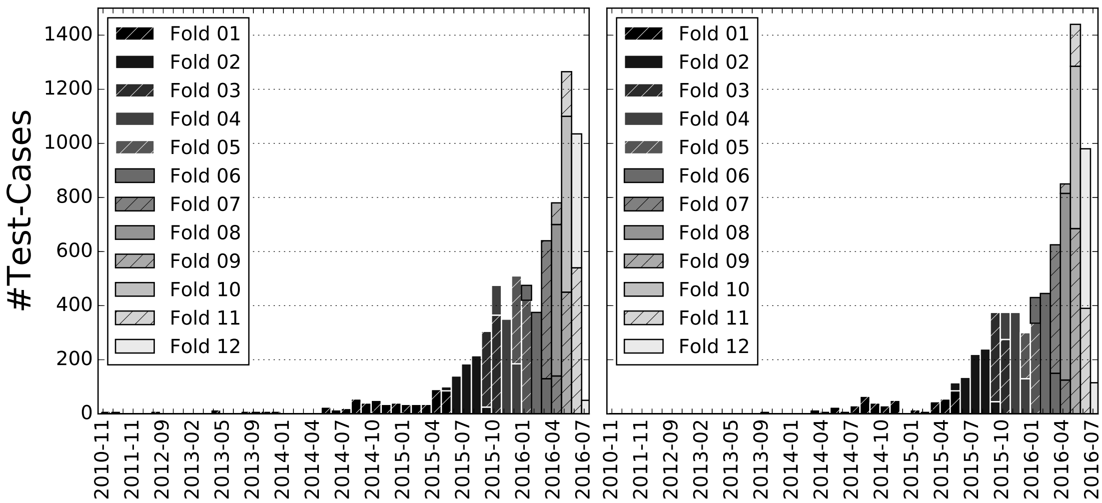


Figure 5.2: Folds' temporal distribution (NYC left and LON right).

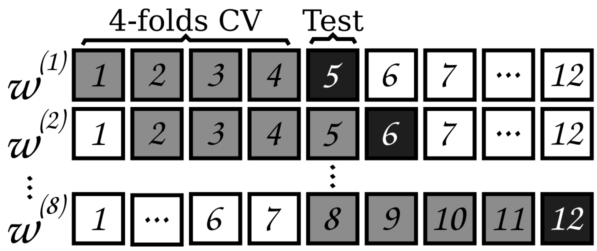


Figure 5.3: Sliding time windows.

EWoM 3.2.5, and PR 3.2.2) and the review count itself, need to be adjusted according to the time frame they are used, discounting the reviews that are ahead in time in order to keep the evaluation coherent.

Hyper-parameter search was performed via exhaustive search (grid-search) using the 4-fold cross-validation in the first time window (2-train, 1-validation, 1-test) over specified parameters (See Table C.1). However, no significant improvement were observed, therefore same default parameters were used across all time windows. In our experiments, we used standard recommended hyper-parameters for LambdaMART as follow, $1,000$ trees with 10 leaves each, minimum leaf support 1, unlimited threshold candidates for tree splitting, learning rate 0.1, and early stopping after 100 non-improving iterations.

5.4 Evaluation Metrics

5.4.1 Mean Reciprocal Rank

The reciprocal rank (RR) is a ranking metric, which is defined as the reciprocal of the position r at which the first relevant document was retrieved. RR ranges between one and zero, and it is 1/1 if the first relevant document was retrieved at position rank 1, 1/2 if the first relevant document was retrieved at position rank 2, and so on. The averaged version of RR, across a set of queries Q, is called the Mean Reciprocal Rank (MRR), as explicitly define in:

$$\mathrm{MRR} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{\mathrm{r}_i}. \tag{5.1}$$

5.4.2 Empirical Feature Efficiency

In order to analyze the impact that features have in a model, we define the empirical feature efficiency (EFE) score. Ablation studies intentionally omit the usage of features at the training stage of the model [Richardson et al., 2006; Horvitz and Apacible, 2003; Mooney and Roy, 2000] to observe changes in the performance. To asses an empirical measurement of the feature efficiency, repeated measurements of the model's performance are compared to the performance of the model that makes integral usage of the features.

Concretely, the ablation experiment consist to submit a model $\mathcal{M}^{(i)}$ through a training and evaluation procedure, where $\mathcal{M}^{(i)}$ is a model that uses the set of features $\mathcal{F}^{(i)} = \mathcal{F} - \{f_i\}$ to train, and f_i is the selected feature excluded in the model. $\mathcal{M}^{(\mathrm{all})}$ denotes the model that was trained using the set of all features $\mathcal{F}$. To assess the contribution of the feature $f^{(i)}$ to $\mathcal{M}^{(\mathrm{all})}$, we compare the metric performance $metric(M^{(i)})$ obtained from the evaluation of $M^{(i)}$ with the $metric(M^{(all)})$ obtained from evaluating $\mathcal{M}^{(\mathrm{all})}$. Then to facilitate the observation of performance decay/gain when omitting $f^{(i)}$, $metric(M^{(i)})$ is expressed in terms of the percentage of $metric(M^{(all)})$, as show in:

$$EFE = \frac{metric(M^{(all)}) - metric(M^{(i)})}{metric(M^{(all)})} \times 100\,, \tag{5.2}$$

consequently, we interpret three possible scenarios:

1. $metric(M^{(i)}) < metric(M^{(all)})$ is understood as a **positive** contribution from $f^{(i)}$

to $\mathcal{M}^{(\mathrm{all})}$, as in the absence of the feature the performance decays, and the EFE becomes a positive score that empirically quantifies a feature importance.

2. $metric(M^{(i)}) > metric(M^{(all)})$ is interpreted as a **negative** contribution from $f^{(i)}$ to $\mathcal{M}^{(\mathrm{all})}$, as the absence of the feature lead to improve the model's performance. The EFE is interpreted as the degree of noise that $f^{(i)}$ induced to the model $\mathcal{M}^{(\mathrm{all})}$.

3. $metric(M^{(i)}) = metric(M^{(all)})$ is understood as **no contribution** from $f^{(i)}$ to the model's accuracy, suggesting that $f^{(i)}$ is an irrelevant feature in the model, and the EFE is equal to zero.

5.4.3 Least Square Improvement

Least Square Improvement (LSI) [Friedman, 2001] is a feature relevance score for tree-based models. LSI have been used to quantify feature importance scores of the features in ensembles of regression trees [Lucchese et al., 2015], such as the boosted regression tree learners in LambdaMART. The intuition behind LSI is to be able to quantify feature importance by measuring the capability of a node (feature with a threshold) to discriminate relevant items of training instances.

Concretely, given an ensemble of trees T and $\mathcal{F}$ the set of features that compose the splitting nodes $n \in t$ of a tree $t \in T$, where feature f is used in n, LSI is defined as:

$$i_n^2 = \frac{n_l n_r}{n_l + n_r} (\bar{y}_i - \bar{y}_r)^2 \,, \tag{5.3}$$

where $n_l(n_r)$ is the number of the training instances that were used to create the model in the left(right) child of the splitting node n, and $\bar{y}_l(\bar{y}_r)$ is the mean value assumed by the relevance score in the left(right) child of n. Then, the gain g_i is estimated by summing up the gains across all the split nodes $n \in t$, for all trees $t \in T$ where feature $f^{(i)}$ is used, as:

$$g_i = \sum_{t \in T} \sum_{n \in t} i_n^2 1(v_n = f^{(i)}) \,, \tag{5.4}$$

where v_n is the splitting feature used in node n and 1() is the indicator function. A positive LSI indicates relevance and zero indicates no relevance.

5.5 Baselines

This evaluation setup is intended to simulate real-world scenarios for online lodging recommendation. Therefore, we selected three plausible baselines that appeal to real-world implementations of recommender systems, which are described in the sections below.

5.5.1 Non-Personalized Popularity Ranking

Bias-centric models are non-personalized recommendation techniques that focus in predicting centered rating scores [Aggarwal, 2016] (such as averages). The items' scores computed by such methods are usually values that are centered in reference to the ratings that users gave to such item, and those scores can also be used to compute top-N recommendations. Popularity-recommenders [Schafer et al., 2001; Wei et al., 2007; Burke, 2002] fall into the category of bias-centric models, where the predicting rating scores are merely explained by item's popularity. One of the reasons they work in practice is that customers often wish to know about the most popular items as a mean of indicate an item not to be missed or an important item to pay attention to [Burke, 2002]. Furthermore, popularity-based recommendations are very common as they are well received by users and they are easy to compute [Schafer et al., 2001; Leino, 2014]. Despite failing to enhance item discovery [Herlocker et al., 2004], they are intuitive baselines that perform reasonable well by reason of the probability of a user to dislike popular items is low [Steck, 2011; Celma and Cano, 2008].

The popularity recommender employed in this work is a bias-centric model based on the lodging's review count (number of reviews a lodging possesses). In such recommender, lodgings are sorted prioritizing lodgings with greater popularity. Ties are solved randomly sorting conflicting items.

5.5.2 Airbnb Ranking

The Airbnb search engine is indeed a recommender system. Most of our test collection was retrieved using the Airbnb search tool, including the candidate sets that compose our test cases. Such candidate sets correspond to the lodgings that need to be ranked in our evaluation framework, which is the recommendation problem we outlined. Candidate sets were obtained following a crawling procedure which mainly consisted in centering the map on a given location to retrieve lodgings from the Airbnb's recommender. Consequently, all the candidate sets that compose the test cases in our collection have already been ranked by the Airbnb's recommender system.

When such lodging lists were collected, Airbnb intentionally achieved recommendation for an anonymous user (user with no booking history). Therefore, we can compare the performance of our proposed model on equal circumstances that the Airbnb's recommendation engine, hence becoming an ideal baseline. The Airbnb recommender allows to compare the performance of our proposed model facing a real-world setup.

5.5.3 Bayesian Personalized Ranking Matrix Factorization

Bayesian Personalized Ranking Matrix Factorization (BPRMF) [Rendle et al., 2009] is a classic state-of-the-art matrix factorization model for item recommendation commonly employed as baseline in many experimental setups for recommender systems [Li et al., 2017; Yuan et al., 2016, 2017]. BPRMF is optimized for top-N item recommendations and is intended to work with binary relevance data. In this work, the recommendation problem is a top-N recommendation taks, where only a single item is relevant. Therefore, in accordance to the problem outlined, compare the performance of BPRMF against our proposal allows to asses the efficiency of our model.

BPRMF is classified as a collaborative filtering (CF) method. Contrary to other CF, BPRMF is not rating prediction oriented, which means that it does not attempts to predict the actual rating scores in order to produce rankings. Also, BPRMF approaches the recommendation task as a ranking problem. MyMediaLite [Gantner et al., 2011] includes an implementation of BPRMF, and is used in our experiments. Hyperparameter tuning is performed via grid search using the training folds of the first round.

5.6 Summary

In this chapter, we described an evaluation framework that is intended to simulate genuine real-world lodging scenarios in the sharing economy, while safeguarding the temporal integrity of the data using time constraints during the evaluation framework. Such experimental methodology is conducted in eight evaluation rounds, splitting the data in 12 folds that are grouped in eight sliding time windows. Each time window is comprised of five adjacent folds, where the first four folds provide training data to learn a model, performing a 4-fold cross-validation strategy, and the last remaining fold is used to test the performance of the model. In the following chapter, we present and analyze the results of the evaluation of CLLR, the model we proposed in this work, by employing the evaluation methodology presented in here.

Chapter 6

Experimental Results and Analysis

In this chapter we present and analyze the evaluation of our proposed model, CLLR, which is inspired in five preference dimensions of repurchase intention: perceived value, perceived risk, perceived authenticity, electronic-word-of-mouth, and price sensitivity. We also discuss the evaluation results obtained according to the experimental methodology we described in this work. Our experiments were designed to investigate how recommendation is improved when modelling lodgings using the preference dimensions of repurchase intention and approaching the recommendation problem as a ranking task. In the upcoming sections we present the results and discuss the findings of our investigation, dedicating a section for each of the outlined research questions:

- **RQ1**: How accurate is CLLR for lodging recommendation?
- **RQ2**: How robust is CLLR for lodging recommendation?
- **RQ3**: How do single features contribute to the performance of CLLR?
- **RQ4**: How do our results relate to existing theories of the sharing economy?

6.1 Model Effectiveness

In this section, we discuss the results concerning the effectiveness of CLLR, to address the research question **RQ1**.

In Figure 6.1 (a), we present the MRR derived from the experiments. Also, for each result the error bars correspond to the 95% confidence interval (CI). The label CLLR stands for the model we trained using all features that we proposed in Chapter 3. In addition, in Figure 6.1 (b) we summarized the resulting outcomes of comparing a pair of RS, where each cell corresponds to the evaluation of two competing models, indicated

(a) MRR and 95% CI.

(b) Win-Loss significance (P-values).

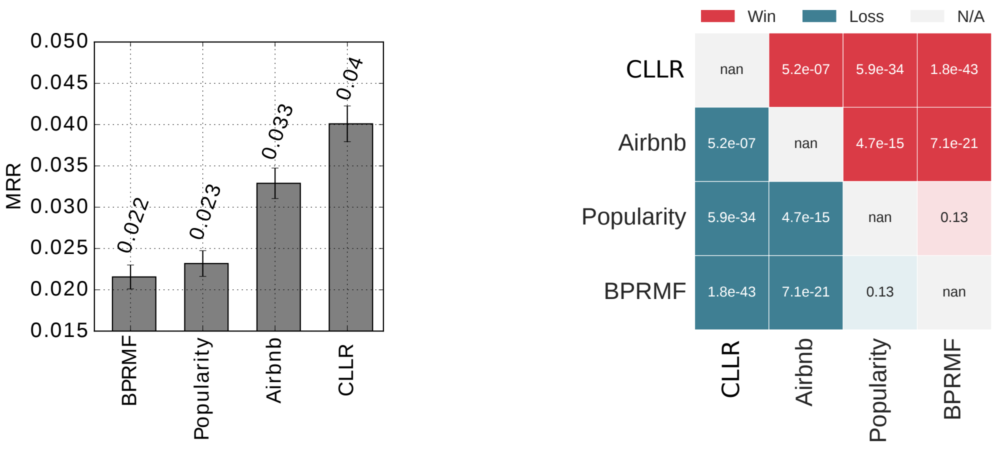


Figure 6.1: Accuracy.

by the row's/column's label. Reading the matrix row-wise, the red cell indicates that the model in the corresponding row has a larger MRR (win) than the model at the given column, and reciprocally, blue indicates that the model in the row has a lower MRR (loss). The number in the cell corresponds to the p-value of a two-tailed paired t-test comparing the result. To interpret the p-values in such matrix, the Bonferroni correction [Dunn, 1961] is considered, in order to prevent misleading interpretation of statistical significance in multiple comparisons $\alpha/m = 0.05/6 = .0083$. To facilitate visualization, the intensity of the color indicates smaller p-values (greater significance).

According to Figure 6.1 (a) and (b) the MRR obtained from BPRMF suggests that collaborative filtering suffered from the sparsity problem. The sparsity issue has already been pinpointed in the hotel domain by other authors [Zhang et al., 2015; Saga et al., 2008]. Supporting the last observation, the mean sparsity at each time window is ≈ 99.99, suggesting that most users' booking records may not be sufficient to accomplish useful factorization or the computation of user-item similarities, which are critical to produce recommendations in CF. The MRR of the Popularity recommender resembles the MRR of BPRMF, resulting in a statistically tied performance between them. Despite being a black-box, Airbnb's recommender seems arguably more sophisticated than one may expect, as it obtained a greater MRR than popularity and BPRMF. Finally, the MRR of our proposed model is superior to the given baselines, demonstrating the effectiveness of our proposal.

To improve the understanding of CLLR's effectiveness, Figure 6.2 shows details of the models' effectiveness for each of the evaluation time windows. The red stars mean that CLLR performed statistically better ($\alpha = 0.05$) than all the baselines, and the blue stars denote statistical significance against **either** the Airbnb **or** the Popularity baseline ($\alpha = 0.05$), according to a two-tailed paired t-test. We first note that CLLR performed consistently better than the rest of the recommenders across all time windows, except for the 4^{th} one, where the Airbnb recommender was the best.

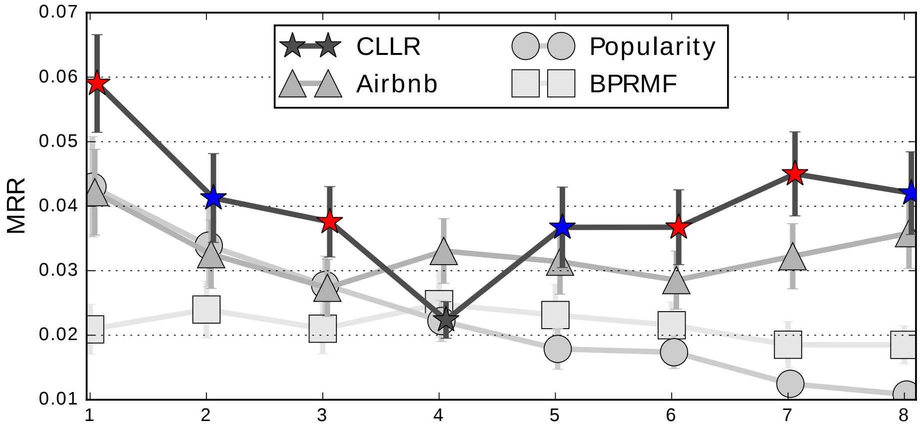


Figure 6.2: MRR by time window. Error bars denote 95% confidence interval.

There are various reasons that may explain the behavior of the performance decay in the 4^{th} time window. It is well known that predictive models perform poorly if the training/test data is not uniform. Atypical events on the training may cause the model to learn irrelevant patterns. Likewise, anomalous events on the test can make the performance of the model differ from the performance achieved during the training. We posit two possible scenarios that may cause the training to diverge from the test data. Notice that the effectiveness of the popularity baseline decays (Figure 6.2), indicating that users stopped consuming only popular items, perhaps as a consequence of the increase of the lodging supply. Also, because data folds tend to get packed as the time window slides to more recent dates (Figure 5.2), training data may be restricted to only the immediate time period of the testing, which in case of seasonable phenomena may produce atypical behavior. 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. Most of the training data are in the winter, whereas test data resides in spring. This observation suggests that seasonal events may have hindered the model's generalization capabilities on this particular time window.

To better understand the drop in performance at the fourth round in Figure 6.2, we perform a complementary ablation experiment. Figure 6.3 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 considerably improves the performance of the model. By discarding information that potentially misled CLLR into making wrong predictions we can have evidence of the cause of the decay in performance.

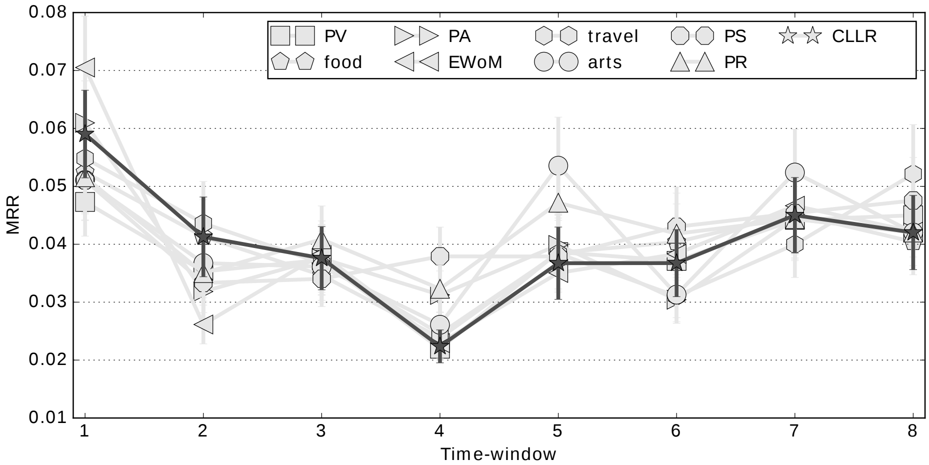


Figure 6.3: MRR by time window. Error bars denote 95% confidence interval.

In summary, these results show evidence that the information contained in the features of our proposed model succeed to tackle the task of recommending lodgings in the domain of the sharing economy.

6.2 Model Robustness

In order to address RQ2 regarding the robustness of the CLLR, we investigate the efficiency of the model while removing entire feature groups. Robustness refers to the capability of the model to favorably generalize its accuracy under different circumstances. To this end, we explore the robustness in terms of omitting various sources of information. In Figure 6.3 and Figure 6.4, the labels: PR (perceived risk), PV (perceived value), PA (perceived authenticity), EWoM (electronic-word-of-mouth), PS (price sensitivity), Food (food venues), Arts (arts & entertainment venues), and Travel (travel & transportation venues), correspond to the models trained while removing the corresponding set of features.

Figure 6.4 (a) shows the MRR of the different models with the corresponding 95% confidence interval (CI). Also, by reading Figure 6.4 (b) row-wise, we illustrate the outcome (win/loss) of the different comparatives. As we see, models' performance is affected by removing information components. Nevertheless, when compared to

(a) MRR and Confidence Interval (0.95) (b) Win-Loss Significance (P-values)

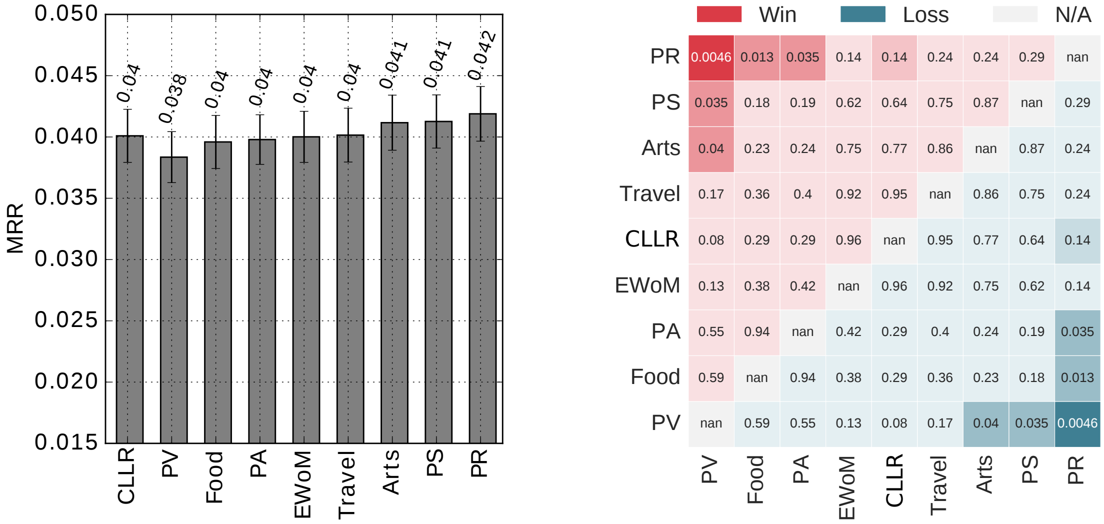


Figure 6.4: Robustness

the model employing the complete set of features (CLLR), they show a statistically equivalent performance ($\alpha = 0.05$) (Bonferroni $\alpha = 0.0014$).

Furthermore, to have a broader understanding of the robustness of our model, we measured the MRR obtained while varying the sliding time window. In Figure 6.3 we show the variation of the MRR in different evaluation rounds. As one may observe, they oscillate around the performance of the model using the complete set of features (CLLR). From these results, we conclude that our model is robust and performs uniformly while removing various sets of features.

6.3 Feature Efficiency

In this section we analyze the performance implications that single features have in CLLR, in order to address the research question RQ3. In addition, the findings also serve as evidence to validate the correctness of the assumptions we made while constructing the preference dimensions of CLLR. Many of the features we built had their foundation under the assumption that:

- **H1** A user's repurchase intention toward a lodging is influenced by it's surrounding facilities.
- **H2** A user's repurchase intention toward a lodging is influenced by other available lodgings.

Both H1 and H2 assume that the physical region where lodgings are located is important. H1 is restricted to state that the surrounding urban configuration (for instance the presence of venues) is important to discriminate relevant lodgings. H2 highlights the psychological interpretation of raw attributes, such as prices, review count, and distances to venues, are interpreted according to the context.

Empirical feature efficiency (EFE) and least square improvement (LSI) are used to quantify features' importance. To facilitate the analysis, EFE is presented in Table 6.1, grouped by the preference dimensions of CLLR. The symbols △ (▲) denote statistical significant according to a two-tailed paired t-test $\alpha \leq 0.05$ ($\alpha \leq 0.01$), between MRR measurements. Furthermore, in Figure 6.5 we present the LSI of the features that compose the ensemble of decision trees in our model (using all features). The scores are averaged across the time windows. Noticed that in Figure 6.5, sub-figures (a), (b), and (c), have different scales in their x-axis, and that scores are sorted from greatest to lowest while reading the figures top-down and from left to right.

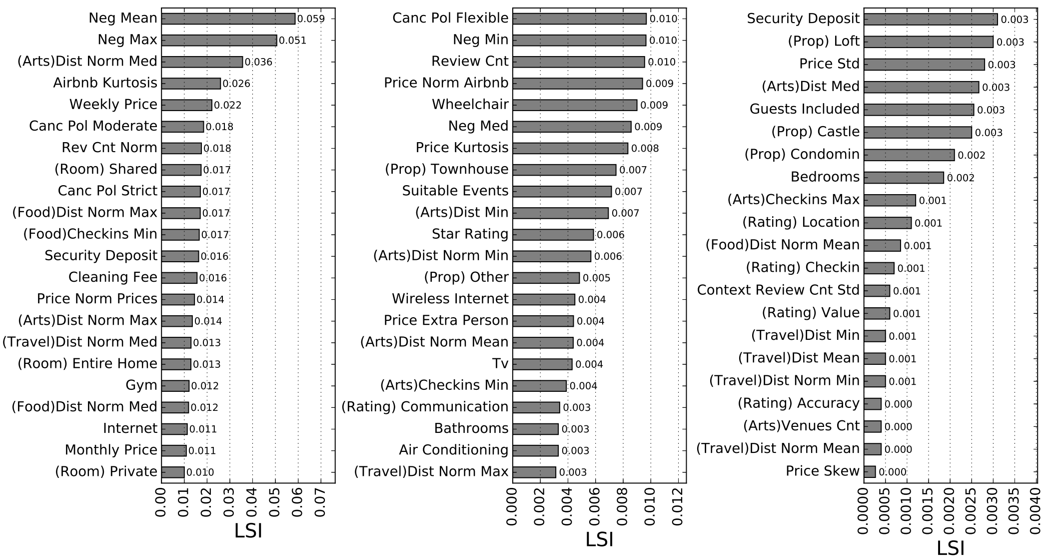


Figure 6.5: Averaged LSI ($LSI = 0$ are omitted).

Only statistically significant scores are considered to support the discussion of our findings, in conjunction with LSI scores. The following sections group the findings to facilitate their interpretation.

6.3.1 Review Count

The review count normalized by the review counts of the lodgings surrounding the accommodation had the greatest positive contribution in the model (See Table 6.1, Rev Cnt Norm). In the absence of such feature the performance greatly decays (8.07%),

	Feature	(+/-)%	**Feature**	(+/-)%	**Feature**	(+/-)%
PR	Rev Cnt Norm	8.1$^{\blacktriangle}$	Context Review Cnt Std	5.4$^{\blacktriangle}$	Star Rating	4.0$^{\blacktriangle}$
	(Rating) Location	3.0$^{\blacktriangle}$	Canc Pol No Refunds	1.4	(Rating) Value	1.4
	Canc Pol Flexible	0.0	Canc Pol Strict	0.0	Context Rooms Cnt	0.0
	Canc Pol Super Strict 30	0.0	(Rating) Cleanliness	0.0	Canc Pol Super Strict 60	0.0
	(Rating) Communication	−0.2	(Rating) Accuracy	−0.3	Context Review Cnt Mean	−0.6
	(Rating) Checkin	−0.8	Canc Pol Moderate	−1.1	Review Cnt	−9.8$^{\blacktriangle}$
FOOD	(Food)Dist Max	5.7$^{\blacktriangle}$	(Food)Dist Mean	4.3$^{\blacktriangle}$	(Food)Checkins Max	2.7
	(Food)Checkins Min	2.7$^{\triangle}$	(Food)Dist Med	2.5	(Food)Dist Norm Med	1.8
	(Food)Venues Cnt	1.2	(Food)Checkins Med	0.7	(Food)Dist Norm Min	0.1
	(Food)Checkins Mean	−0.5	(Food)Dist Norm Mean	−1.7	(Food)Dist Norm Max	−2.2$^{\blacktriangle}$
	(Food)Dist Min	−4.5$^{\triangle}$				
ARTS	(Arts)Checkins Mean	3.5$^{\blacktriangle}$	(Arts)Checkins Min	2.9$^{\blacktriangle}$	(Arts)Dist Mean	2.8
	(Arts)Venues Cnt	2.4	(Arts)Checkins Max	1.4	(Arts)Dist Norm Med	0.6
	(Arts)Dist Norm Mean	0.5	(Arts)Dist Max	0.5	(Arts)Checkins Med	0.3
	(Arts)Dist Norm Min	0.0	(Arts)Dist Norm Max	−1.6	(Arts)Dist Min	−2.3
	(Arts)Dist Med	−3.3$^{\triangle}$				
TRAVEL	(Travel)Dist Mean	7.1$^{\blacktriangle}$	(Travel)Checkins Med	2.8$^{\blacktriangle}$	(Travel)Dist Med	1.6
	(Travel)Dist Max	1.3	(Travel)Dist Norm Mean	1.0	(Travel)Dist Min	0.1
	(Travel)Dist Norm Med	0.0	(Travel)Checkins Min	0.0	(Travel)Venues Cnt	−0.8
	(Travel)Checkins Mean	−0.9	(Travel)Dist Norm Min	−1.0	(Travel)Checkins Max	−1.5
	(Travel)Dist Norm Max	−1.7				
PS	Price Kurtosis	4.5$^{\blacktriangle}$	Airbnb Skew	3.9$^{\triangle}$	Price Norm Airbnb	2.6$^{\triangle}$
	Airbnb Kurtosis	1.9	Context Price Mean	1.6	Airbnb Mean	0.2
	Price Skew	−0.3	Price Norm Prices	−0.4	Price Std	−1.1
EWoM	Comp Skew	6.5$^{\blacktriangle}$	Pos Skew	6.2$^{\blacktriangle}$	Neg Med	6.1$^{\blacktriangle}$
	Neg Mean	5.1$^{\blacktriangle}$	Neu Med	4.6$^{\triangle}$	Comp Med	4.1$^{\triangle}$
	Neu Max	3.7	Neu Kurtosis	2.5	Comp Min	2.4
	Neg Kurtosis	2.2	Neg Max	1.9	Comp Mean	1.7
	Pos Mean	1.7	Comp Kurtosis	0.9	Pos Kurtosis	0.6
	Neg Skew	0.2	Neu Mean	0.2	Pos Max	−0.5
	Pos Med	−1.7	Neu Skew	−2.1	Neu Min	−2.4
	Neg Min	−3.6	Comp Max	−4.1$^{\triangle}$	Pos Min	−5.8$^{\triangle}$
PA	Auth Min	6.3$^{\blacktriangle}$	Auth Kurtosis	4.3$^{\blacktriangle}$	Auth Mean	2.1
	Auth Med	0.2	Auth Max	−0.6	Auth Skew	−1.7
PV	Security Deposit	3.9$^{\blacktriangle}$	Monthly Price	2.2	(Room) Shared	2.1
	Hangers	2.1$^{\blacktriangle}$	Wheelchair	1.3	(Prop) Castle	1.2$^{\triangle}$
	Elevator Building	1.0	Washer	1.0	Cable Tv	1.0
	Iron	0.9	Bedrooms	0.6$^{\triangle}$	Shampoo	0.6
	Fire Extinguisher	0.6	First Aid Kit	0.5	Smoke Detector	0.4
	Person Capacity	0.4	Lock Bedroom Door	0.4	Cleaning Fee	0.3
	Internet	0.3	(Bed) Couch	0.2	Doorman	0.1
	(Prop) House	0.0	Laptop Friendly	0.0	Safety Card	0.0
	Pets In Prop.	0.0	Essentials	0.0	Indoor Fireplace	0.0
	Air Conditioning	0.0	Hot Tub	0.0	Heating	0.0
	Carb.Monox.Det.	0.0	Smoking Allowed	0.0	24 Hour Check In	0.0
	(Bed) Pull Out Sofa	0.0	Suitable Events	0.0	(Prop) Chalet	0.0
	(Room) Entire Home	0.0	(Prop) Yurt	0.0	(Prop) Villa	0.0
	(Prop) Tent	0.0	(Prop) Loft	0.0	(Prop) Lighthouse	0.0
	Tv	0.0	(Prop) Hut	0.0	(Prop) Dorm	0.0
	(Prop) Condomin	0.0	(Prop) Igloo	0.0	(Prop) Cave	0.0
	(Prop) Apartment	0.0	Wireless Internet	0.0	(Prop) Bungalow	0.0
	(Prop) Boat	0.0	Kitchen	−0.0	(Room) Private	−0.0
	Dryer	−0.1	Beds	−0.2	Price Extra Person	−0.2
	(Prop) Townhouse	−0.3	Parking	−0.3	Weekend Price	−0.4
	(Bed) Futon	−0.5	(Prop) Other	−0.7	Guests Included	−0.8
	(Bed) Real Bed	−0.8	Bathrooms	−1.1	Gym	−1.2
	(Prop) Cabin	−1.2	Breakfast	−1.7	Breakfast	−1.7
	Pets Allowed	−2.0$^{\triangle}$	Hair Dryer	−2.1$^{\blacktriangle}$	(Bed) Airbed	−2.2$^{\blacktriangle}$
	Weekly Price	−2.2	Pool	−2.9$^{\blacktriangle}$	Price	−3.1$^{\triangle}$

Table 6.1: Empirical Feature Efficiency (EFE).

making the normalized version of the review count the most informative feature, according to the EFE score. Furthermore, the LSI of the normalized review count obtained the 7^{th} greatest score (Figure 6.1, (a)), which agrees with the previous observation.

Surprisingly, by removing the feature review count (not normalized) the model had the greatest increase in performance (Table 6.1, Review Cnt -9.8%), making the review count the noisiest feature in our model, according to its EFE. Notice that the normalized review count has a LSI almost twice than the one obtained for the non-normalized (Figure 6.1, (a) and (b)). Despite both showing LSIs that contribute to the model, the EFE suggests that the feature review count seems to fail to replace the discriminative capabilities of its normalized form, conversely, the normalized feature seems to be a powerful substitute of the non-normalized feature, regardless of both being closely related.

It has been shown that redundant features degrade the performance for some models [Langley and Sage, 1994; Pazzani, 1996; Zhu et al., 2003; John et al., 1994], which may explain such behavior. Indeed, the analysis of the features' correlation discards trivial redundancy (low correlation), suggesting that the normalized version is a substantially a different feature, with greater discriminatory capabilities, according to the results. Finally, we highlight the EFE score (5.4) and non-zero LSI of the standard deviation of the review counts of 18 accommodations around the lodging (Table 6.1, Context Review Cnt Std) (Figure 6.1, (c)).

The inferiority of the feature importance for review count against its normalized form supports assumption H2, which states that the interpretation of some features is done in accordance with their context. In addition, the importance of the standard deviation of the contextual review counts is also congruent with such finding.

6.3.2 Distance-Based Features

The features in this category were built employing the information of venues and their distances around the accommodation. They are intended to describe the value derived from the physical context where lodgings are immersed. Not all the features in this category obtained significant results to state that they all contribute to the model, nevertheless, the following features obtained a statistically significant EFE:

- Food's maximum/mean distance from the venue to the accommodation,
- Arts & Entertainment venues' mean/maximum number of check-ins.
- Travel & Transportation venues' mean distance and their check-in's median.

The number of positive contributions of these features contrasts with the low occurrence of irrelevant attributes (Table 6.1, Food, Arts, and Travel have few features with 0.0 EFE). In particular, the travel & transportation mean distance from the venues to the accommodation obtained the second greatest EFE score (7.1), which is an interesting result that clearly evokes the touristic domain we are dealing with. These findings suggest that the characteristics of the physical context (food, arts, entertainment, travel, and transport venues) are important factors as many tourism theories pinpoint [Papatheodorou, 2001; Sparks et al., 2003; Gross and Brown, 2008]. Furthermore, the LSI systematically categorized the normalized distances (by the number of check-ins) from the lodging to the venues as relevant attributes in the model, in contrast to the non-normalized distance features, which mostly obtained a LSI equal to zero, or smaller than their normalized counterparts. Conversely, the EFEs of the normalized distances suggest that they poorly contributed to the models' performance.

These last results slightly differ from each other, contrary to previous findings where we were able to observe EFE and LSI converging to the same conclusions. A conservative interpretation would not safely state that such results fully support assumption H2, however these results encourage the exploration of the validity of such findings. Nevertheless, they are also evidence that support the correctness of assumption H1.

6.3.3 Price Sensitivity

The normalized lodging price by the Airbnb mean-price of nearby lodgings obtained a positive EFE score (1.6). In contrast, the feature *price* obtained the noisiest score in the PV category according to the EFE (-3.1) (Table 6.1, Price Norm Airbnb and Price). In addition, the normalized versions of prices obtained a greater LSI than price, which obtained a LSI equal to zero (Figure 6.1, (a) and (b)). Similarly to the analysis made in Section 6.3.1 (review count versus its normalized form), these two features do not show trivial correlation, and the superiority of the normalized version supports the appropriate intuition we followed to create price sensitivity features. Furthermore, we notice that context price kurtosis (symmetry score of the curve of prices around the accommodation) and the skewness of Airbnb's price histogram (long-tail score for the curve shaped by the price histogram) have a positive contribution to the model (EFE 4.5 and 3.9 respectively) and non-zero LSIs.

Such results empirically demonstrate the utility of employing contextual information to normalize prices. By using such features, the feature-model gained discriminative capabilities in terms of determine relevant lodgings, hence, supporting assumption

H2.

6.3.4 Perceived Risk

According to the EFE and LSI, the star-rating (overall rating-score) has the greatest importance score than any of the other ratings, as one would have expected (Table 6.1, Star Rating and Figure 6.5, (b)). Airbnb allows guests to evaluate the convenience of the lodging's location using the location rating-score, which obtained the second(third) best EFE(LSI) (Table 6.1, (Rating) Location)(Figure 6.5) among the rating scores. Such results suggest that the model learned that guests give importance to the location ratings of the lodgings and consequently favorably support the correctness of assumption H1, which states that including context characteristics improves the model's capability to recommend relevant items.

6.3.5 Perceived Value

The features in this category mostly describe physical aspects of the accommodation and are agnostic to the lodging's context. Also, many of them are indicator variables, taking the values $\{0, 1\}$ to indicate the absence or presence of some categorical effect [Suits, 1957], such as the property types, bed types, room types, cancellation policies, and amenities.

According to the results, most of the indicator variables have poor performance implications or null contribution (e.g. all categories of room type, Figure 6.1, PV). Also, lodging's traits that one would suppose to be important, such as number of beds, rooms, and bathrooms, surprisingly obtained scores indicating to be noisy or not discriminative. Only security deposit was highlighted by the importance scores as a relevant feature. On the other hand, our intuition is that normalizing or combining PV features would considerably enhance their usefulness. For features such as weekly price, monthly price, and cancellation policies, the LSI and EFE scores differ from each other. Further experiments are needed to investigate their true usefulness.

Contrasting the results of features that leveraged lodgings' context, the features in this category obtained inferior feature importance. These findings suggest that features that merely describe physical characteristics of the lodging are less effective that the ones that leverages the context, which is evidence that validates assumption H1 and H2, that state that considering context improves recommendation.

6.3.6 Electronic-Word-of-Mouth and Perceived Authenticity

The sets of features in EWoM and PA were obtained using the text contained in the lodging's reviews. The concept of EWoM was modeled using the reviews to obtain sentiment polarity scores (positive, negative, neutrality, and compound). Also, a PA score was computed using a language model to obtain the similarity between the reviews and a PA lexicon.

The negative sentiment features remarkably obtained the two greatest LSI scores out of all the features in our model (Figure 6.5 (a)). However, only the mean average of the negative scores showed a positive contribution according to EFE. Excluding the negative polarity dimension, the rest of the categories comprise at least one feature with EFE scores oscillating between 4.0 to 6.5 (Table 6.1, EWoM), indicating that they enhance the accuracy of the model. On the other hand, they did not obtain LSIs demonstrating their contribution to the model's performance. The PA features obtained two features with EFE contributing to the model (Table 6.1, Auth Min, Auth Kurtosis), However, the LSI does not support the importance of any feature under this category.

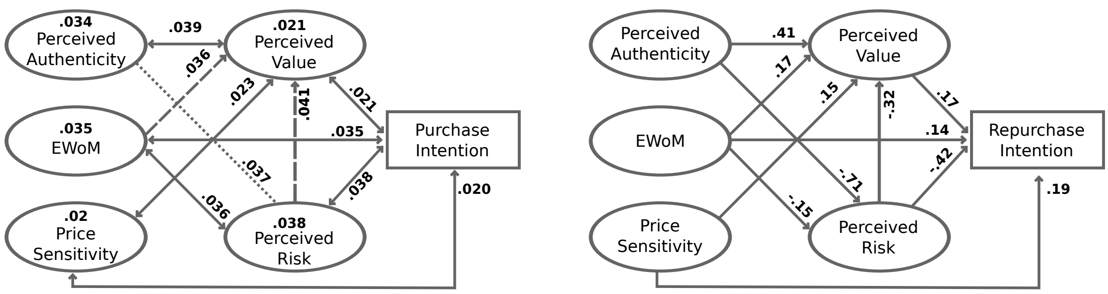
The importance for the negative sentiment polarity scores were supported by both of the metrics we used. On the other hand, the results for the rest of the features under this category has to be taken with caution, in order to avoid misleading conclusions. However, these results motivate further investigations of the usefulness of polarity sentiment analysis and IR techniques, in order to create more robust features.

6.4 Preference Dimensions and the SEMRI

The repurchase intention model proposed by Liang [2015] (SEMRI, Section 3.2), was validated via structural equation modeling, where model fit was achieved from a questionnaire that was applied to Airbnb's users, in order to evaluate multiple concepts involved in repurchase intention. The SEMRI aims to explain how these criteria mutually interact and translate to repurchase intention. Despite the data-driven methodology adopted in this dissertation be different, CLLR was built under similar premises. Therefore, in this section, we further analyze our results in contrast to the original SEMRI, in order to address RQ4.

Let $\mathcal{D} = \{\mathrm{PV}, \mathrm{PR}, \mathrm{PS}, \mathrm{EWoM}, \mathrm{PA}\}$ denote the preference dimensions in CLLR. When users book a lodging, they implicitly state their preferences towards the features of the room. Instead of applying a questionnaire, we indirectly observe the users' preferences for lodging's traits, directly from their bookings. Therefore, given a feature f, belonging to a preference dimension $d \in \mathcal{D}$, if f demonstrates to be discriminative when modeling and predicting users preferences, we assume that it truly implies that users give importance to the preference dimension d.

Purchase intention (PI) concerns about the aspects that affect users purchase decisions. On the other hand, repurchase intention (RI) explains the drivers in repeated consumption (continuous repurchases) [Chen et al., 2016]. In our experiments, we cannot observe what influences users to repeatedly consume. Instead, we are able to observe the aspects that led to consumption. For such reason, in Figure 6.6 (a) we show the **purchase intention** model we built. The diagram aims to illustrate the extent to which one(multiple) concept(s) explain users' preferences. In contrast, the SEMRI in Figure 6.6 (b) should be interpreted as the degree in which one(multiple) concept(s) correlate and translate into **repurchase intention**. Users' purchase intentions and repurchase intention are not necessarily aligned concepts. However, they are complementary concepts and can be analyzed together.



(a) MRR and the preference dimensions. (b) SEMRI.

Figure 6.6: CLLR's validation.

In Figure 6.6 (a) the ellipses represent the preference dimensions. The numbers within ellipses are the MRR obtained from the evaluation of a model using solely the corresponding features. Edges $A \overset{mrr}{\rightarrow} B$ indicate that two preference dimensions were combined $A \cup B$ to train a model. For any combination of feature dimensions the MRR obtained from the evaluation is indicated by the edge's weight. An arrow from $A \overset{mrr}{\rightarrow} B$ indicates that the MRR of combining $A \cup B$ is statistically different from the MRR of $\overset{mrr}{B}$ (two-tailed t-test ($\alpha \leq 0.05$)). Double arrows $A \overset{mrr}{\leftrightarrow} B$ indicates that significance is found from $\overset{mrr}{\leftrightarrow}$ against $\overset{mrr}{A}$ and $\overset{mrr}{B}$. Dashes on the edges facilitate the visualization of statistical significance. All edges pointing to *Purchase Intention* are simply the MRR of the preference dimension.

The first of our findings is that perceived risk, PR, is the factor that most impacts PI and RI in their respective model. The risk involved in booking a lodging is the main driver for consumption and repeated consumption. Also, perceived value, PV, is greatly influenced by PR in both models. Conversely, PV does not influence PR, which is supported by our findings and the SEMRI. Our model does not support that there is a real interaction between PA and PR, moreover, electronic-word-of-mouth (EWoM) has a greater effect in PR in our model than the one observed in SEMRI. For the rest of the edges, our model often indicates that there is a two-sided effect, which is a curious finding not considered in the studies we covered in this dissertation. These results are encouraging in a way that further studies may strength the implications of such findings. Despite the implications of these results are restricted to our lodging sharing economy domain, they are interesting directions to be undertaken in other sharing economy domains. The full list of results is shown in Appendix D.1, which fully includes all the combinations of the preference dimensions and their statistical significance.

6.5 Summary

In this chapter, we tested the efficiency of multiple baselines and compared their results against CLLR, our proposed model, which demonstrated to achieve better overall accuracy, and local accuracy at each evaluation round. In addition, we explored the robustness of our proposal and observed that removing parts of the features does not significantly degrade its performance.

We also investigated the feature importance and the correctness of the assumptions we made during the design of the feature-based model. The results indicated that features that were normalized by the corresponding attributes of other available lodgings showed to have improved discriminative power, which is in line with how users interpret the lodgings attributes. In particular, we found that risk factors, such as the lodgings' review count, are one of the most important features in our model. Similarly, other lodging attributes, such as price, demonstrated to be sensitivity to contextual information. Furthermore, we showed that negative opinions in the reviews are important features to be considered in order to discriminate relevant items.

Finally, we summarized the predictive capabilities of the individual preference dimensions in our model into a comprehensive diagram that permitted contrasting the repurchase intention model that inspired us with the repurchase intention we created. In the next chapter, we summarize the outcomes of our investigation and we conclude this work with future directions.

Chapter 7

Conclusions and Future Works

Recommender systems (RS) allow users to have a customized interaction with online platforms, while at the same time, they alleviate the information overload found in online environments. The tourism industry has largely adopted the Internet as one of its main sales' channel. Nevertheless, few RS have been proposed to assist users seeking for lodging, in order to simplify the complexity of the traveling planning process. In recent years, we have witnessed the emergence of innovative online companies, commonly designated as the sharing economy. Such phenomena have significantly changed the way people consume, particularly in the tourism industry, which hosts one of the major protagonists of the sharing economy, namely the Airbnb online lodging platform.

The rapid expansion and adoption of the sharing economy has been studied in different areas. However, it reminded unexplored for RS as we showed in the literature we covered. Few works in RS have dedicated their effort to tackle hotel recommendations, nonetheless none of them has explored the sharing economy. For such reason, we proposed the CLLR, which is a recommender system for the lodging domain, with a focus on the sharing economy. Our model leverages learning-to-rank techniques and implements the socio-economy theories of users' repurchase intention we explored in this dissertation, which served to frame five preference dimensions, namely: perceived value, perceived risk, perceived authenticity, electronic-word-of-mouth, and price sensitivity.

In order to assess the effectiveness of our approach, a test collection and an evaluation framework were specifically created. The evaluation methodology is intended to (1) simulate a user exploring a location where he or she aims to sojourn, (2) preserve the integrity of the temporal nature of the problem, (3) permit the comparison of the effectiveness of a recommender system against common theoretical and practical recommendation benchmarks.

The sharing economy has the potential to become an important portion of the digital economy. Consequently, the recommender systems' community should also extend their efforts to tackle the challenges that such domain may pose. To this end, this dissertation bridged socio-economic studies of the sharing economy with recommender systems, discussing practical implications in both areas.

The following sections summarize the conclusions drawn from our investigation and the main contributions of this work, to finally conclude with directions for future works.

7.1 Summary of Contributions

We summarize the main contributions of this dissertation as follows. **CLLR** In Chapter 3, we proposed a feature-based model for lodging recommendation, which is sustained in five preference dimensions, namely, perceived value, perceived risk, electronic-word-of-mouth, perceived authenticity, and price sensitivity. The preference dimensions are modeled using sentiment analysis, information retrieval, and feature engineering approaches. The features we create are leveraged by a learning-to-rank technique to tackle recommendation of lodgings in the sharing economy.

Test Collection and Characterization In Chapter 4, we created a data collection in order to evaluate recommender systems in sharing economy platforms, with a particular focus on lodging recommendation. The collection was built using a methodology for sampling lodgings that has the advantage to create plausible lodging scenarios for RS. Also, we proposed a strategy to reconstruct users booking profiles. Alternative sources of data were used to enrich the collection with contextual information. Furthermore, in Section 4.2.3 a characterization was performed to demonstrate the completeness and coverage of our data, as well to contrast the peculiarities of the cities we chose to simulate recommendation.

Evaluation Methodology In Chapter 5, we proposed an evaluation framework, intended to simulate users searching for an accommodation exploring the search map of a lodging provider. In addition, such evaluation considers the temporal nature of the problem and was used to evaluate our proposed CLLR against multiple baselines, including state-of-the-art algorithms for recommender systems and real-world recommenders that operate in current lodging providers of the sharing economy.

Evaluation and Model Validation In Chapter 6, we presented the results of our empirical evaluation, which were used to demonstrate the performance and robustness of our model. In Section 5.4.2 we proposed a new metric to assess empirical feature efficiency, which in conjunction to least squared improvement was used to measure features' contribution to the model. We used such metrics to characterize the functionality that single features have in the model. In addition, in Section 6.4, we characterized the interaction of the preference dimensions and we compared them to existing theories of the lodging sharing economy. Such analysis allowed to expand the understanding of practical implications for recommender systems and sharing economy.

7.2 Summary of Conclusions

In this dissertation, we proposed the CLLR in order to fill the gap in lodging recommender systems for the sharing economy environment. We demonstrated that the CLLR has a greater performance than well known state-of-the-art collaborative filtering algorithms. In addition, our model showed significantly improvements compared to real-world recommenders in the industry of lodging recommendation.

The analysis and understanding of the features in our model allowed the assessment of their effectiveness, which demonstrated the importance that contextual information have in the tourism and lodging domain. Also, our results suggested that features that are normalized based on the attributes of other available lodgings have improved discriminative power.

In addition, we performed a characterization of the preference dimensions in our model in contrast to the SEMRI model that inspired us. This characterization provided further evidence that the users' perceived risk is one of the most important factors, which highly influences purchase intention and perceived value. Moreover, our model suggested that there is no relation between perceived authenticity and perceived risk as the SEMRI model states. Electronic-word-of-mouth (EWoM) has a greater effect in PR than the one observed in previous studies. Finally, we discovered two-side effects that were not considered by other previous studies in repurchase intention of Airbnb's customers.

7.3 Directions for Future Research

Derived from the research we conducted during this dissertation, we propose the following directions for future research:

- In this work we restricted the usage of context information according to theories of the sharing economy, however, future works may be willing to explore traditional context information such as season information or the time of the year.

Similarly, social features, demographic features (e.g. well-being and wealth), and hierarchical features can also be included as part of the recommendation model.

- As explained in Section 3, our model was not intended to tackle personalization, however, our proposed model can be easily extended to handle personalization via normalization of features based on users booking records. For instance, personalization can be achieved by normalizing the prices of the candidate set using the booking prices in the booking history of a user, therefore, achieving personalized pricing features. Furthermore, other personalization techniques can also be used to compute personalization scores that can be added as new features in the model.

- During our experimental analysis some salient features exhibited results encouraging further investigation of their possible contribution to CLLR. Further exploration of polarity sentiment analysis and information retrieval techniques may enhance the discriminative capabilities under each preference dimension and significantly improve the model.

- The test collection we built in this dissertation contains lodgings' information omitted during the modeling of our features. For instance, textual description of the lodgings (e.g. summary, description, space, house rules, location description, transit and access, neighborhood overview). Also, lodgings' images and hosts' information constitute interesting sources of information. Such data can be employed to model new features for CLLR.

- Another direction is to weight reviews according to their publication date, in order to penalize older reviews and prioritize reviews published near the time of the booking.

- The sharing economy lodging scenario is intrinsically multi-language, which enlarges the set of experiments that can be conducted in order to tackle the multi-language issue, which was not considered in this work.

- Employing various sources of contextual information from other mapping services (e.g. Google Maps, Open Street Maps, Yelp) can also lead to interesting findings.

- Finally, future works may be willing to explore other machine learning models capable to leverage the features we proposed.

Appendix A

Feature Model

Features			
(Arts) Checkins Max	(Arts) Checkins Mean	(Arts) Checkins Med	(Arts) Checkins Min
(Arts) Dist Max	(Arts) Dist Mean	(Arts) Dist Med	(Arts) Dist Min
(Arts) Dist Norm Max	(Arts) Dist Norm Mean	(Arts) Dist Norm Med	(Arts) Dist Norm Min
(Arts) Venues Cnt.	(Bed) Airbed	(Bed) Couch	(Bed) Futon
(Bed) Pull Out Sofa	(Bed) Real Bed	(Food) Checkins Max	(Food) Checkins Mean
(Food) Checkins Med	(Food) Checkins Min	(Food) Dist Max	(Food) Dist Mean
(Food) Dist Med	(Food) Dist Min	(Food) Dist Norm Max	(Food) Dist Norm Mean
(Food) Dist Norm Med	(Food) Dist Norm Min	(Food) Venues Cnt.	(Prop.) Apartment
(Prop.) Boat	(Prop.) Bungalow	(Prop.) Cabin	(Prop.) Camper/Rv
(Prop.) Castle	(Prop.) Cave	(Prop.) Chalet	(Prop.) Condomin.
(Prop.) Dorm	(Prop.) House	(Prop.) Hut	(Prop.) Igloo
(Prop.) Lighthouse	(Prop.) Loft	(Prop.) Other	(Prop.) Tent
(Prop.) Townhouse	(Prop.) Villa	(Prop.) Yurt	(Room) Entire Home
(Room) Private	(Room) Shared	(Travel) Checkins Max	(Travel) Checkins Mean
(Travel) Checkins Med	(Travel) Checkins Min	(Travel) Dist Max	(Travel) Dist Mean
(Travel) Dist Med	(Travel) Dist Min	(Travel) Dist Norm Max	(Travel) Dist Norm Mean
(Travel) Dist Norm Med	(Travel) Dist Norm Min	(Travel) Venues Cnt.	24 Hour Check In
Air Conditioning	Bathrooms	Bedrooms	Beds
Breakfast	Breakfast	Buzzer/Wireless Intercom	Cable Tv
Carb.Monox.Det.	Cat(S)	Cleaning Fee	Dog(S)
Doorman	Dryer	Elevator Building	Essentials
Family/Kid Friendly	Fire Extinguisher	First Aid Kit	Guests Included
Gym	Hair Dryer	Hangers	Heating
Hot Tub	Indoor Fireplace	Internet	Iron
Kitchen	Laptop Friendly	Lock Bedroom Door	Monthly Price
Other Pet(S)	Parking	Person Capacity	Pets Allowed
Pets In Prop.	Pool	Price	Price Extra Person
Safety Card	Security Deposit	Shampoo	Smoke Detector
Smoking Allowed	Suitable Events	Tv	Washer
Washer / Dryer	Weekend Price	Weekly Price	Wheelchair
Wireless Internet			

Table A.1: PV Features

Features			
(Rating) Accuracy	(Rating) Checkin	(Rating) Cleanliness	(Rating) Communication
(Rating) Location	(Rating) Value	Canc. Pol. Flexible	Canc. Pol. Moderate
Canc. Pol. No Refunds	Canc. Pol. Strict	Canc. Pol. Super Strict 30	Canc. Pol. Super Strict 60
Context Review Cnt. Mean	Context Review Cnt. Std	Context Rooms Cnt	Rev Cnt. Norm
Review Cnt.	Star Rating		

Table A.2: PR Features

Features			
Airbnb Kurtosis	Airbnb Mean	Airbnb Skew	Context Price Mean
Price Kurtosis	Price Norm Airbnb	Price Norm Prices	Price Skew
Price Std			

Table A.3: PS Features

Features			
Auth Kurtosis	Auth Max	Auth Mean	Auth Med
Auth Min	Auth Skew		

Table A.4: PA Features

Features			
Comp Kurtosis	Comp Max	Comp Mean	Comp Med
Comp Min	Comp Skew	Neg Kurtosis	Neg Max
Neg Mean	Neg Med	Neg Min	Neg Skew
Neu Kurtosis	Neu Max	Neu Mean	Neu Med
Neu Min	Neu Skew	Pos Kurtosis	Pos Max
Pos Mean	Pos Med	Pos Min	Pos Skew

Table A.5: EWoM Features

Appendix B

PA Lexicon, Reviews Examples, and ES parameters

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			

Table B.1: Hand build authenticity lexicon.

Listing B.1: ElasticSearch Similarity Configuration

```
{"settings":{
    "index":{
        "similarity":{
            "default":{"type":"LMDirichlet"}
        },
        "number_of_shards":1,
        "number_of_replicas":0
    },
    "analysis":{
        "filter":{
            "en_US":{
                "type":"hunspell",
                "language":"en_US"
            },
            "english_stop":{
                "type": "stop",
                "stopwords": "_english_"
            },
            "english_keywords":{
                "type": "keyword_marker",
```

```
                "keywords":  ["MustSetSomeWord"]
            },
            "english_stemmer":{
                "type": "stemmer",
                "language":  "english"
            },
            "english_possessive_stemmer":{
                "type": "stemmer",
                "language":  "possessive_english"
            }
        },
        "analyzer":{
            "english":{
                "tokenizer": "standard",
                "filter":[
                "en_US",
                "english_possessive_stemmer",
                "lowercase",
                "english_stop",
                "english_keywords",
                "english_stemmer"
                ]
            }
        }
    }
},
"mappings":{
    "review":{
        "properties":{
            "comment":{
                "type":"string",
                "analyzer":"english"
            }
        }
    }
}
}'
```

Review
Our stay with $Host^{(1)}$ and his family was by far the best Airbnb experience Ive had as a guest! Upon arrival I felt slightly annoyed with my partners choice of accommodation inland. I thought the beach would be ideal in my pregnant state. I could not have been more wrong! Every part of our experience was rich with relaxation and nourishment! And we were able to make daily trips to a wonderful beach on the south side of the island. On our journeys to the beach we discovered villages that felt like they had stopped in time; multi-generation family houses upon olive farms and wine orchards. Stunning mountains with peaks into the sun washed skies. The beach we chose to visit, upon $Host^{(1)}$'s recommendation, was free of mass-tourism. We often had Long beaches with piercing blue water all to ourselves. And not far from our favorite beach we discovered what became our favorite restaurant, Jimmys. We ate fantastic fish and Greek delights such as homemade stuffed grape leaves, locally harvested salads and crisp Cretan white wines. At home we were endlessly spoiled my $Host^{(1)}$ and his gorgeous and gracious wife. Each morning we were delivered a new organic delight; fresh baked breads, local honey, cheese and homemade Sheeps milk yogurt. An assortment of melons and heavenly figs, pears, oranges the list is endless. We were warmly invited to meet on the terrace for shared dinners, which was often abundant with local dishes prepared by $Host^{(1)}$ wife. And of course a local wine and lots of story telling. They invited as a there guests to a neighbors wedding. 500 guests! We were once again spoiled with there wariness and open hearted nature. We spent a lovely evening talking to a backdrop of traditional Greek music more incredible food and conversation. Our last night ended with a meal under the stars. $Host^{(1)}$ brought out his telescope and gave us a tour of the moon. And we ate snails collected by their two sons and carefully and lovingly prepared. The family is full of warmth and generosity. I felt welcome on the farm but I also felt we had a lot of privacy in our flat. This is an authentic experience with genuinely warm and interesting people. If you want the best of Crete, stay a fews day on the farm with $Host^{(1)}$ and his family. You will go home feeling like you had a real life experience, and not something packaged and sold out of a catalogue. Thank you to the entire family for hosting us. We made some lasting memories.
$Host^{(2)}$ and $Host^{(3)}$ (Feb 10th-11th) These guys are seriously rock stars (no pun intended). Communication to get checked in went great because both were at work when I arrived. These guys are here to make everyone that comes to their hot spot of a listing have a memorable story to share with their friends and family back home. $Host^{(2)}$ and $Host^{(3)}$ have multiple bunk beds and multiple rooms for multiple people to experience a multitude of memories. These guys have a mini hostel in the Mission District! $Host^{(3)}$ got in first from work and wanted to get to know me, why I was traveling, what I was out in San Francisco for and right away offered me a beer. After 10-15 minutes of chatting and getting to know him, his roommate $Host^{(2)}$ came in and was all ready to do some rock climbing. Almost simultaneously, they both asked if I wanted to join and I said Absolutely! We went and I was able to meet a few of their friends that were into climbing as well. After about an hour of rock climbing at Mission Cliffs, which was right down the road from them (\$20-\$25 for a one day/night pass by the way), we all went to an awesome dinner spot at a place called Southern Pacific. The food was absolutely fantastic. Even better then the food, the hospitality was sincerely genuine and unexpected. These guys truly make you feel like you've been best friends with them for years. 11 out of 10 stars. Easily the best AirBNB experience I've had so far.In addition, because these guys have multiple spots for people to sleep, I was fortune enough to meet three awesome individuals also traveling from different areas of the world! $Guest^{(1)}$, $Guest^{(2)}$ and $Guest^{(3)}$. $Guest^{(1)}$ is originally from Israel, but now lives in the US, and has done commercial real estate for 25+ years back in Israel and all over the United States. We talked for quite a long time and it was so cool connecting with him, talking about business and listening to his perspectives on different things in life. $Guest^{(2)}$ and $Guest^{(3)}$ are buddies traveling together from Edinburgh, Scotland and I was fortunate enough to be awake reading when they got home from Taco Tuesday on their night out around 1:00am. We had the most in depth, open conversation about life, what we want to do in life and some of our past experiences traveling. Vivid memories and conversations I'll be able to take with me, and new friends and connections to keep in touch with as the years go on! $Host^{(2)}$ and $Host^{(3)}$ have a very unique place where it's easy to connect and meet new people in an open and young environment. It's worth every penny and more, book while you can! Check out their wall to see all the faces that have gone through Treat Street!
I stayed 10 days at $Host^{(4)}$'s flat. I felt very welcome and part of a community of very nice people from the first day.During my stay I had several very interesting long chats with other guests at $Host^{(4)}$'s flat, like a couple from Paris and a couple from Utrecht, in the beautiful kitchen or on the overwhelming terrace. It was like who already had discovered a few places passed his/her experience to the new guests while having a bear, a tea or a coffee or we just talked about life in general. $Host^{(4)}$ arrange a contact to a student from Germany still studying in Napoli, who had stayed at her place before, for me, who gave me several very interesting tips. $Host^{(4)}$ was always very hospital. She cooked pasta for us and I had breakfast with her. $Host^{(4)}$ likes to have at least one guest with her staying in the flat. She enjoys talking to guests, about what you are planning to see during the day, politics, your home town, your family, Caravaggio paintings and many things. $Host^{(4)}$ is an actor who lives the challenging live of an artist. You can really feel it she is an artist type of person. I stayed at the single room with the private bath room. I slept very well. The room was perfect for me. It all felt extremely comfortable. I didn't miss a Hotel room. The light in the room is fantastic and the view through the windows is unique with the church roofs and towers around you and Capri saying good morning to you from the sea. I always felt like it is completely up to me, if I prefer to have some privacy in my room or if I would like to talk to $Host^{(4)}$ or other guests instead. The internet bandwidth is in the range of 1 mbps. The signal strength is sometimes a bit low in the single room. I think next time I would bring a small repeater to solve this problem. The location of $Host^{(4)}$'s flat turned out to be perfect for my discovery of the town and the greater Napoli area. The central train station, the Alibus stop (shuttle bus to the airport), the harbor, the metro, stops of other Narrow-gauge railway systems, which will get you to Sorrento or Pozzuoli for example and even places where you can go swimming in the sea are all in working distance. $Host^{(4)}$'s flat offers the chance to discover a very vivid and truly authentic Mediterranean place. I have been to many other places in Italy before. Still I felt Napoli and especially the area around $Host^{(4)}$'s flat was a new experience to me. I felt like I had not really understood this kind of Southern Italian mentality and lifestyle before. Now, looking behind, I think it is one of the top must see places in Europe. The street food is so interesting, tasty and very cheap. Don't miss the Reggia di Caserta, plan a whole day.

Table B.2: Three reviews with the greatest PA scores

Appendix C

Grid Search

Hyper-parameter	Test values
Learning rate	$\{0.0001, 0.001, 0.01, 0.1, 0.25, 0.5\}$
# leaf	$\{2, 5, 10, 25, 50\}$
# tree	$\{2, 5, 10, 25, 50\}$
mls	$\{5, 10, 20, 30, 40, 50\}$

Table C.1: Grid Search.

Appendix D

Preference Dimensions and MRR

$\mathcal{F}_i$	**MRR**	$\mathcal{F}_j$	**MRR** (+/-)%
PR	0.038	EWoM	0.036 (-7.3)
		PA	0.037 (-4.59)
		PS	0.037 (-3.1)
		PV	0.041 (6.48)
PS	0.02	EWoM	0.032 (62.87) ▲
		PA	0.037 (87.57) ▲
		PR	0.037 (89.69) ▲
		PV	0.023 (18.7) ▲
EWoM	0.035	PA	0.039 (11.26) ▲
		PR	0.036 (0.61)
		PS	0.032 (-9.77) ▲
		PV	0.036 (1.64)
PA	0.034	EWoM	0.039 (17.21) ▲
		PR	0.037 (9.12)
		PS	0.037 (9.59) ▲
		PV	0.039 (17.05) ▲
PV	0.021	EWoM	0.036 (74.31) ▲
		PA	0.039 (90.35) ▲
		PR	0.041 (98.06) ▲
		PS	0.023 (12.78) ▲

Table D.1: Preference dimensions, MRR Gain/Loss

Bibliography

Aggarwal, C. C. (2016). *Recommender Systems*. Springer.

Ahmed, M., Spagna, S., Huici, F., and Niccolini, S. (2013). A peek into the future: Predicting the evolution of popularity in user generated content. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 607--616. ACM.

Ajzen, I. and Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological bulletin*, 84(5):888.

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette Books.

Balabanović, M. and Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66--72.

Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, 67(8):1595--1600.

Billsus, D. and Pazzani, M. J. (1998). Learning collaborative information filters. In *Icml*, volume 98, pages 46--54.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993--1022.

Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46:109--132.

Borras, J., Moreno, A., and Valls, A. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, 41(16):7370--7389.

Browning, R. C., Baker, E. A., Herron, J. A., and Kram, R. (2006). Effects of obesity and sex on the energetic cost and preferred speed of walking. *Journal of Applied Physiology*, 100(2):390--398.

Buhalis, D. and O'Connor, P. (2005). Information communication technology revolutionizing tourism. *Tourism recreation research*, 30(3):7--16.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331--370.

Burke, R. (2007). Hybrid web recommender systems. In *The adaptive web*, pages 377--408. Springer.

Cantallops, A. S. and Salvi, F. (2014). New consumer behavior: A review of research on ewom and hotels. *International Journal of Hospitality Management*, 36:41--51.

Carbonell, J. and Goldstein, J. (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*, pages 335--336. ACM.

Celma, Ò. and Cano, P. (2008). From hits to niches?: or how popular artists can bias music recommendation and discovery. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, page 5. ACM.

Chang, E.-C. and Tseng, Y.-F. (2013). Research note: E-store image, perceived value and perceived risk. *Journal of Business Research*, 66(7):864--870.

Chapelle, O. and Chang, Y. (2011). Yahoo! learning to rank challenge overview. In *Proceedings of the Yahoo! Learning to Rank Challenge*, pages 1--24.

Chen, J. V., Yen, D. C., Kuo, W.-R., and Capistrano, E. P. S. (2016). The antecedents of purchase and re-purchase intentions of online auction consumers. *Computers in Human Behavior*, 54:186--196.

Cheng, J., Adamic, L. A., Kleinberg, J. M., and Leskovec, J. (2016). Do cascades recur? In *Proceedings of the 25th International Conference on World Wide Web*, pages 671--681. International World Wide Web Conferences Steering Committee.

Cremonesi, P., Koren, Y., and Turrin, R. (2010). Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 39--46. ACM.

Dabholkar, P. A. and Sheng, X. (2012). Consumer participation in using online recommendation agents: effects on satisfaction, trust, and purchase intentions. *The Service Industries Journal*, 32(9):1433--1449.

Dang, V. (2013). Ranklib.

Darlington, R. B. (1970). Is kurtosis really "peakedness?". *The American Statistician*, 24(2):19--22.

Degeratu, A. M., Rangaswamy, A., and Wu, J. (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):55--78.

Deshpande, M. and Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, 22(1):143--177.

Devijver, P. A. and Kittler, J. (1982). *Pattern recognition: A statistical approach*, volume 761. Prentice-Hall London.

Doane, D. P. and Seward, L. E. (2011). Measuring skewness: a forgotten statistic. *Journal of Statistics Education*, 19(2):1--18.

Donthu, N. and Garcia, A. (1999). The internet shopper. *Journal of advertising research*, 39(3):52--52.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52--64.

Felfernig, A., Gordea, S., Jannach, D., Teppan, E., and Zanker, M. (2007). A short survey of recommendation technologies in travel and tourism. *OEGAI Journal*, 25(7):17--22.

Fodness, D. and Murray, B. (1997). Tourist information search. *Annals of tourism research*, 24(3):503--523.

Fraiberger, S. P. and Sundararajan, A. (2015). Peer-to-peer rental markets in the sharing economy. *NYU Stern School of Business Research Paper*.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189--1232.

Gantner, Z., Rendle, S., Freudenthaler, C., and Schmidt-Thieme, L. (2011). MyMediaLite: A free recommender system library. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys 2011)*.

Grayson, K. and Martinec, R. (2004). Consumer perceptions of iconicity and indexicality and their influence on assessments of authentic market offerings. *Journal of consumer research*, 31(2):296--312.

Gretzel, U. and Yoo, K. H. (2008). Use and impact of online travel reviews. *Information and communication technologies in tourism 2008*, pages 35--46.

Gross, M. J. and Brown, G. (2008). An empirical structural model of tourists and places: Progressing involvement and place attachment into tourism. *Tourism management*, 29(6):1141--1151.

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current issues in Tourism*, 18(12):1192--1217.

Halácsy, P. and Trón, V. (2006). Benefits of resource-based stemming in hungarian information retrieval. In *Workshop of the Cross-Language Evaluation Forum for European Languages*, pages 99--106. Springer.

Hamari, J., Sjöklint, M., and Ukkonen, A. (2015). The sharing economy: Why people participate in collaborative consumption. *Journal of the Association for Information Science and Technology*.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5--53.

Horvitz, E. and Apacible, J. (2003). Learning and reasoning about interruption.

Hurley, N. and Zhang, M. (2011). Novelty and diversity in top-n recommendation–analysis and evaluation. *ACM Transactions on Internet Technology (TOIT)*, 10(4):14.

Hutto, C. J. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Jacoby, J. and Kaplan, L. B. (1972). The components of perceived risk. In *SV-Proceedings of the third annual conference of the association for consumer research*.

Jang, J.-S. (1993). Anfis: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3):665--685.

Jannach, D. and Adomavicius, G. (2016). Recommendations with a purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 7--10. ACM.

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge University Press.

John, G. H., Kohavi, R., Pfleger, K., et al. (1994). Irrelevant features and the subset selection problem. In *Machine learning: proceedings of the eleventh international conference*, pages 121--129.

Kabassi, K. (2010). Personalizing recommendations for tourists. *Telematics and Informatics*, 27(1):51--66.

Karatzoglou, A., Baltrunas, L., and Shi, Y. (2013). Learning to rank for recommender systems. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 493--494. ACM.

Kashyap, R. and Bojanic, D. C. (2000). A structural analysis of value, quality, and price perceptions of business and leisure travelers. *Journal of travel research*, 39(1):45--51.

Kim, D. J., Ferrin, D. L., and Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems*, 44(2):544--564.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., and Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):441--504.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137--1145. Stanford, CA.

Langley, P. and Sage, S. (1994). Induction of selective bayesian classifiers. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, pages 399--406. Morgan Kaufmann Publishers Inc.

Law, R., Buhalis, D., and Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, 26(5):727--750.

Lee, B. S. and McDonald, J. F. (2003). Determinants of commuting time and distance for seoul residents: The impact of family status on the commuting of women. *Urban Studies*, 40(7):1283--1302.

Leino, J. (2014). User factors in recommender systems: Case studies in e-commerce, news recommending, and e-learning.

Levi, A., Mokryn, O., Diot, C., and Taft, N. (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*, pages 115--122. ACM.

Li, G., Zhang, Z., Wang, L., Chen, Q., and Pan, J. (2017). One-class collaborative filtering based on rating prediction and ranking prediction. *Knowledge-Based Systems*.

Liang, L. J. (2015). *Understanding repurchase intention of Airbnb consumers: perceived authenticity, EWoM and price sensitivity*. PhD thesis, University of Guelph.

Litvin, S. W., Goldsmith, R. E., and Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism management*, 29(3):458--468.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1--167.

Liu, T.-Y. et al. (2009). Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, 3(3):225--331.

Lucchese, C., Nardini, F. M., Orlando, S., Perego, R., and Tonellotto, N. (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*, pages 895--898. ACM.

Masiero, L. and Nicolau, J. L. (2012a). Price sensitivity to tourism activities: looking for determinant factors. *Tourism Economics*, 18(4):675--689.

Masiero, L. and Nicolau, J. L. (2012b). Tourism market segmentation based on price sensitivity finding similar price preferences on tourism activities. *Journal of Travel Research*, 51(4):426--435.

Matsubara, Y., Sakurai, Y., Prakash, B. A., Li, L., and Faloutsos, C. (2012). Rise and fall patterns of information diffusion: model and implications. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 6--14. ACM.

Möhlmann, M. (2015). Collaborative consumption: determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of Consumer Behaviour*, 14(3):193--207.

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6):47--60.

Mooney, R. J. and Roy, L. (2000). Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries*, pages 195--204. ACM.

Nicolau, J. L. and Masiero, L. (2013). Relationship between price sensitivity and expenditures in the choice of tourism activities at the destination. *Tourism Economics*, 19(1):101--114.

Nilashi, M., bin Ibrahim, O., Ithnin, N., and Sarmin, N. H. (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):542--562.

O'Connor, P. (2008). User-generated content and travel: A case study on tripadvisor.com. *Information and communication technologies in tourism 2008*, pages 47--58.

O'Mahony, M. P. and Smyth, B. (2009). Learning to recommend helpful hotel reviews. In *Proceedings of the third ACM conference on Recommender systems*, pages 305--308. ACM.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1--135.

Papatheodorou, A. (2001). Why people travel to different places. *Annals of tourism research*, 28(1):164--179.

Pazzani, M. J. (1996). Searching for dependencies in bayesian classifiers. In *Learning from Data*, pages 239--248. Springer.

Quercia, D., Pesce, J. P., Almeida, V., and Crowcroft, J. (2013). Psychological maps 2.0: a web engagement enterprise starting in london. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1065--1076. ACM.

Rendle, S., Freudenthaler, C., Gantner, Z., and Schmidt-Thieme, L. (2009). Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452--461. AUAI Press.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175--186. ACM.

Ribeiro, F. N., Araújo, M., Gonçalves, P., Benevenuto, F., and Gonçalves, M. A. (2015). A benchmark comparison of state-of-the-practice sentiment analysis methods. *arXiv preprint arXiv:1512.01818*.

Richardson, M., Prakash, A., and Brill, E. (2006). Beyond pagerank: machine learning for static ranking. In *Proceedings of the 15th international conference on World Wide Web*, pages 707--715. ACM.

Saga, R., Hayashi, Y., and Tsuji, H. (2008). Hotel recommender system based on user's preference transition. In *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, pages 2437--2442. IEEE.

Sandow, E. and Westin, K. (2010). The persevering commuter–duration of long-distance commuting. *Transportation Research Part A: Policy and Practice*, 44(6):433--445.

Schafer, J. B., Konstan, J., and Riedl, J. (1999). Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158--166. ACM.

Schafer, J. B., Konstan, J. A., and Riedl, J. (2001). E-commerce recommendation applications. In *Applications of Data Mining to Electronic Commerce*, pages 115--153. Springer.

Shankar, V., Rangaswamy, A., and Pusateri, M. (1999). The online medium and customer price sensitivity. *University Park: E-business Research Center*.

Shardanand, U. and Maes, P. (1995). Social information filtering: algorithms for automating "word of mouth". In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210--217. ACM Press/Addison-Wesley Publishing Co.

Sivapalan, S., Sadeghian, A., Rahnama, H., and Madni, A. M. (2014). Recommender systems in e-commerce. In *2014 World Automation Congress (WAC)*, pages 179--184. IEEE.

Sparks, B., Bowen, J., and Klag, S. (2003). Restaurants and the tourist market. *International Journal of Contemporary Hospitality Management*, 15(1):6--13.

Steck, H. (2011). Item popularity and recommendation accuracy. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 125--132. ACM.

Suits, D. B. (1957). Use of dummy variables in regression equations. *Journal of the American Statistical Association*, pages 548--551.

Taylor, J. W. (1974). The role of risk in consumer behavior. *The Journal of Marketing*, pages 54--60.

Thelwall, M., Buckley, K., and Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163--173.

TranSafety, I. (1997). Study compares older and younger pedestrian walking speeds. *Road Management & Engineering Journal*.

Tsao, W.-C., Hsieh, M.-T., Shih, L.-W., and Lin, T. M. (2015). Compliance with ewom: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management*, 46:99--111.

Wei, K., Huang, J., and Fu, S. (2007). A survey of e-commerce recommender systems. In *Service systems and service management, 2007 international conference on*, pages 1--5. IEEE.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37--52.

Wu, Q., Burges, C. J. C., Svore, K. M., and Gao, J. (2008). Ranking, boosting, and model adaptation. Technical report, Microsoft Research.

Yuan, F., Guo, G., Jose, J. M., Chen, L., Yu, H., and Zhang, W. (2017). Boostfm: Boosted factorization machines for top-n feature-based recommendation. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pages 45--54. ACM.

Yuan, F., Jose, J. M., Guo, G., Chen, L., Yu, H., and Alkhawaldeh, R. S. (2016). Joint geo-spatial preference and pairwise ranking for point-of-interest recommendation.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3):338--353.

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *The Journal of marketing*, pages 2--22.

Zekanovic-Korona, L. and Grzunov, J. (2014). Evaluation of shared digital economy adoption: Case of airbnb. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*, pages 1574--1579. IEEE.

Zervas, G., Proserpio, D., and Byers, J. (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).

Zhai, C. (2008). Statistical language models for information retrieval. *Synthesis Lectures on Human Language Technologies*, 1(1):1--141.

Zhai, C. and Lafferty, J. (2004). A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems (TOIS)*, 22(2):179--214.

Zhang, K., Wang, K., Wang, X., Jin, C., and Zhou, A. (2015). Hotel recommendation based on user preference analysis. In *Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference on*, pages 134--138. IEEE.

Zhu, J., Rosset, S., Hastie, T., and Tibshirani, R. (2003). 1-norm support vector machines. In *NIPS*, volume 15, pages 49--56.