Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Mining Bilateral Reviews for Online Transaction Prediction: A Relational Topic Modeling Approach

Jiawei Chen, Yinghui (Catherine) Yang, Hongyan Liu

http://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Vol. 32, No. 2, June 2021, pp. 541–560
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Mining Bilateral Reviews for Online Transaction Prediction: A Relational Topic Modeling Approach

Jiawei Chen,[a] Yinghui (Catherine) Yang,[b] Hongyan Liu[c]

[a] School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China; [b] Graduate School of Management, University of California, Davis, California 95616; [c] School of Economics and Management, Tsinghua University, Beijing 100084, China

Contact: chenjiawei@mail.shufe.edu.cn, https://orcid.org/0000-0001-6116-7232 (JC); yiyang@ucdavis.edu, https://orcid.org/0000-0001-6984-0102 (Y(C)Y); liuhy@sem.tsinghua.edu.cn, https://orcid.org/0000-0002-4902-1078 (HL)

Received: July 12, 2018
Revised: April 14, 2019; February 18, 2020; July 2, 2020
Accepted: September 20, 2020
Published Online in Articles in Advance: April 2, 2021

https://doi.org/10.1287/isre.2020.0981

Abstract. On most e-commerce platforms, reviews are often written by buyers to evaluate sellers or products offered by sellers. In recent years, more and more platforms allowing both buyers and sellers to write reviews for each other have emerged. These bilateral reviews are important information sources in the decision-making process of both buyers and sellers but have not been properly investigated in the literature before. We develop a comprehensive relational topic modeling approach to analyze bilateral reviews to predict transaction results. The prediction results will enable the platform to increase the chance that the buyer and seller reach a transaction by presenting buyers with offerings that are more likely to lead to a transaction. Within the framework of the relational topic model, we embed a topic structure with both shared and corpus-specific topics to better handle text corpora generated from different sources. Our model facilitates the extraction of the appropriate topic structure from different document collections that helps enhance the transaction prediction performance. Comprehensive experiments conducted on real-world data sets collected from sharing economy platforms demonstrate that our new model significantly outperforms other alternatives. The robust results obtained from multiple sets of comparisons demonstrate the value of bilateral reviews if they are processed properly. Our approach can be applied to many platforms where bilateral reviews are available.

History: Olivia Liu Sheng, Senior Editor; Xiaobai Li, Associate Editor.
Funding: This work was supported by the Fundamental Research Funds for the Central Universities of Shanghai University of Finance and Economics, and the National Natural Science Foundation of China [Grants 71771131 and 71490724].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2020.0981.

Keywords: online reviews • bilateral reviews • topic modeling • text mining • online transaction prediction

1. Introduction

After a typical transaction on a traditional e-commerce platform, the buyer can often post comments about the experience with the product or the seller. On these platforms, there are typically only reviews written by buyers to evaluate sellers and not vice versa. These reviews are known as unilateral reviews. In recent years, more and more platforms allowing both buyers and sellers to provide reviews for each other have emerged. For example, on Airbnb, a leading marketplace for accommodation sharing, 80% of hosts leave reviews for their guests and 72% of guests leave reviews for hosts (Quora 2012). Some more examples of the platforms that allow bilateral reviews are BoatSetter (boat sharing), Fiverr (services such as graphic design, translation, and programming), TaskRabbit (personal assistance, office service, and other handyman services), Eatwith (meal sharing), uShip (shipping services of large items), and Peerspace (booking spaces for events).

Platforms with bilateral reviews often have a two-stage decision-making process. For example, on Airbnb, in the first stage, the guest will pick an accommodation listing and submit a booking request, and then in the second stage, the host will review the buyer's information and decide whether to accept the booking request and complete the transaction. In this two-stage decision-making process, bilateral reviews play a very essential role. In the first stage, historical guest reviews for the listing and its host can help the guest evaluate different listings to make a choice, and in the second stage, the host will assess the guest through historical host reviews for the guest before deciding to accept the booking request or not. In this paper, our objective is to design a model based on bilateral reviews and other available information including item descriptions and user profiles to assess the chance that the buyer and seller reach a transaction. Based on this model, the platform can return a list of items with the highest chance of generating a transaction when a buyer searches for potential offerings. This will present the buyer with offerings she is more interested in booking and lower the chance of the seller rejecting the buyer's booking request by providing the seller with a buyer she is more likely to accept. The improvement in the overall efficiency will increase buyer and seller satisfaction with the transaction process; thus, the platform can generate more transactions while retaining and attracting more users.

In the aforementioned two-stage process, bilateral reviews contain essential information for both buyers and sellers to make decisions and for the platform to design a better mechanism to match buyers and sellers; therefore, it is very important to properly analyze bilateral reviews. Traditional recommendation systems only consider buyer reviews when generating the recommended products. In the problem we study, we not only need to present buyers with items they are interested in, but also need to link sellers with buyers they are less likely to reject. In order to do that, both buyer reviews and seller reviews need to be leveraged. It is also nontrivial to analyze bilateral reviews because it is not optimal to apply existing methods for unilateral reviews mechanically to analyze bilateral reviews. If we merge all the buyer reviews and seller reviews together without differentiating them and treat them as unilateral reviews, it is difficult to model the two-stage decision process. Buyer reviews and seller reviews are used in different stages of the process to evaluate different entities in the system and thus need to be treated separately instead of mixed together. It is also inferior to process buyer reviews and seller reviews entirely independently in different stages as we discuss more in Section 2. There are also challenges we face when processing bilateral reviews. Bilateral reviews contain two types of reviews. How to properly handle the similarities and differences between buyer reviews and seller reviews is a methodological challenge. Another challenge comes from the two-stage decision process. We need to clearly identify the subset of the reviews needed for each stage and model the final transaction outcome dependent on the outcome of both stages. Given these challenges, from a methodology point-of-view, analyzing bilateral reviews is very different from analyzing unilateral reviews. Almost all research conducted on online reviews in the literature is based on unilateral reviews. Little attention has been paid to the value of bilateral reviews. These research gaps motivated our study, and no previous research has studied the problem of leveraging the content of bilateral reviews in a predictive setting, especially for the two-stage decision-making problem on platforms with bilateral reviews.

To address these research gaps, we propose an integrated modeling framework to study bilateral reviews in a two-stage decision-making process including a requesting stage and an accepting stage. In the requesting stage, the buyer examines the information of a potential offering to see if it matches her preferences. The information available to the buyer contains the basic description of the offering and all the buyer reviews written for it and its corresponding seller. If the buyer has made a booking request in the requesting stage, then in the accepting stage, the seller will decide whether to accept the booking request. Before making the decision, the seller can refer to the buyer's basic profile and past seller reviews written by other sellers to see if the buyer meets the requirements.

To incorporate bilateral reviews and other available textual information such as item descriptions to predict results of potential transactions, we develop a new relational topic model (Chang et al. 2010) to analyze buyers, sellers, items, and transactions between them. In our relational topic modeling framework, individual buyer and seller reviews and item descriptions are assembled into different types of documents to describe items, buyers, and sellers. Meanwhile, topics leaned from these documents are used to generate features to predict the possibility of a transaction reached between a buyer and a seller. The model can learn topic distributions and predict transaction results simultaneously. Within the framework of our relational topic model, we adopt a topic structure with both shared and corpus-specific topics to better handle text corpora generated from different sources. This topic structure facilitates the extraction of shared topics from different types of documents and corpus-specific topics that only apply to each individual type, respectively, which helps enhance the transaction prediction performance. Topic models are a very popular and powerful model for extracting meaningful features from textual information (Huang et al. 2017, Gong et al. 2018). Unlike other ways of summarizing text (e.g., word embeddings learned from deep learning models and the traditional Term Frequency-Inverse Document Frequency representation), the features derived from topic models can be easily interpreted as opinions expressed in the reviews, which can be used to derive insights about the factors driving transaction outcome. Because of this reason, numerous studies have adopted topic modeling techniques to handle unilateral reviews (McAuley and Leskovec 2013, Büschken and Allenby 2016, Puranam et al. 2017, Ansari et al. 2018). To capture the connection between the features derived from texts and the transaction outcome, we choose the relational topic model that allows us to learn the parameters from the topic modeling component and the transaction prediction component simultaneously instead of separately. This has been proven to generate better-performing models (Chang et al. 2010). On top of generating interpretable results, our model also performs the best in terms of predictive power compared with other baselines including deep learning models.

To evaluate our proposed model, we conduct comprehensive experiments on real-world datasets collected from Airbnb and a boat sharing platform called Boatsetter. Our new model significantly outperforms baseline models in transaction prediction accuracy, and the performance improvements are reliable and robust. The comprehensive experiments conducted in this paper further demonstrate the benefit of tailoring a model to handle bilateral reviews. Our proposed model not only achieves superior predictive performance but also produces meaningful insights. We extracted coherent and comprehensive topics including both shared and corpus-specific topics, which aligns well with the motivation of introducing shared and corpus-specific topics. This provides a more comprehensive understanding of the topic structure embedded in the different types of text inputs and helps enhance the transaction prediction performance through the method we proposed. Our method handling multiple types of text corpora within the relational topic modeling framework can also identify factors from both the buyer and seller side that contribute to transaction success and failure. For example, if a host who prefers friendly guests receives a booking request from a guest who is considered friendly by other hosts, the host is very likely to accept this booking request.

Our research makes several important contributions. First, our study is the first to analyze the content of online bilateral reviews in a predictive setting, especially for the two-stage decision-making problem on platforms offering bilateral reviews. We identified a few papers (Fradkin et al. 2015, Bridges and Vásquez 2018, Quattrone et al. 2018) that use bilateral reviews collected from Airbnb as the source of their data, but their studies are very exploratory. On platforms where decisions from both buyers and sellers can affect whether a transaction will go through, reviews provided from both sides should be considered to analyze their preferences in order to predict transactions. With the consideration of bilateral reviews, platforms can significantly improve the chance that the buyer and seller reach a transaction by matching consumers and providers more effectively.

Second, we propose a new relational topic model that can better handle bilateral reviews for platforms offering bilateral reviews. We model the two-stage decision-making process using topic modeling and analyze textual information generated from different sources by incorporating shared and corpus-specific topics. Our model has several significant extensions on top of the original relational topic model. The experimental results on real-world datasets show robust performance improvements of our proposed model over many baseline models when predicting transaction results. Our modeling framework can be adapted to platforms where multiple text corpora are available to assist different users to make more informed decisions, especially in a two-stage decision-making process. Such platforms often have some of the following characteristics: transaction success highly depends on trust between buyers and sellers; the product shared or service offered involves valuable properties; the buyers and sellers need to interact with each other for a long period of time. Our method can easily benefit many sharing economy platforms and can further be adapted to platforms where different types of text inputs are available to improve the matching mechanism between different parties on the platform. For example, a job search site has jobseekers' resumes and companies' job descriptions as two different types of text sources. The decision process also has two stages with the jobseeker first sending the resume to the company and the company then deciding whether to invite the jobseeker for further evaluations. We provide more discussions on the applicability of our model in Section 6.

Our research also contributes to important information systems topics such as recommender systems (Ghoshal et al. 2015, Li et al. 2016), predictive analysis (Pant and Sheng 2015), and user review analysis (Puranam et al. 2017, Ansari et al. 2018). Because the offerings presented to buyers can be considered as the results of a recommender system, the predictive model we develop in this paper can potentially offer a better item recommendation solution for platforms with bilateral reviews such as many sharing economy platforms. We incorporate bilateral reviews into the predictive analysis process to leverage the rich information contained in bilateral reviews, thus make a significant contribution to the research on analyzing user reviews. In addition, we contribute to the stream of research which uses topic models to process textual information (Bao and Datta 2014) by creating a novel topic model tailored specifically to handle bilateral reviews in a two-stage decision process.

2. Literature Review

Topic models are a family of models that aim at discovering latent structures in large document collections. The most commonly used topic model is the latent Dirichlet allocation (LDA) model developed by Blei et al. (2003). As a generative statistical model, LDA represents the generation of documents by introducing latent topics, then each document is considered as a mixture of some topics, and each topic is regarded as a mixture of words. As an extension of the classic LDA model, the relational topic model (RTM) was proposed by Chang et al. (2010) to jointly model document contents and link structure among documents. The links between documents can be defined based on the data, for example, citations between papers, hyperlinks between web sites, and transactions between buyers and items in our case. In RTM, document contents are summarized by latent topics. In addition to topic modeling, there is also a predictive modeling component in this model, which can predict the link status between a pair of documents based on the similarity between the topic representations of the two documents. Because links between documents are conditionally dependent on the topics in RTM, parameters for the topic modeling component and the link prediction component are learned simultaneously, which was proven to be more advantageous than learning them separately. Chen et al. (2015) later provided some theoretical improvements over the original RTM. Niu et al. (2014) build on the original RTM to analyze images and their relations. The model we propose in this paper makes several extensions on top of the original RTM model, and we will discuss the details of the extensions in Section 4.2.

Numerous studies have adopted topic modeling techniques to handle unilateral reviews (McAuley and Leskovec 2013, Büschken and Allenby 2016, Puranam et al. 2017, Ansari et al. 2018). For example, Büschken and Allenby (2016) use the sentence structure contained in review texts to improve inference and prediction of consumer ratings. Although there is plenty of research analyzing unilateral reviews using topic models, bilateral reviews have never been analyzed in the topic model framework. In fact, there is very limited research analyzing bilateral reviews because they only started to attract attention in more recent years with the development of the sharing economy.

Over the past few years, more and more platforms with bilateral reviews have emerged. Several researchers have attempted to use bilateral review data, especially data from Airbnb, for various research problems. However, most of these studies are still analyzing unilateral reviews, that is, either guest reviews (Sanchez-Vazquez et al. 2017, Lawani et al. 2019, Luo and Tang 2019) or host reviews (Cui et al. 2020), but not both at the same time. Some studies just used review volume (Quattrone et al. 2016) and numeric review ratings (Fradkin 2017, Fu et al. 2017, Wang and Nicolau 2017, Chattopadhyay and Mitra 2019) without considering any review texts. A few recent papers have turned their attention to bilateral reviews (Fradkin et al. 2015, Bridges and Vásquez 2018, Quattrone et al. 2018), but they treat reviews from two sides independently as if treating unilateral reviews. Bridges and Vásquez (2018) obtained 400 publicly available reviews posted on Airbnb. Half of the reviews are written by guests and the other half are reviews written by hosts. They conduct exploratory analysis and discover that both reviews written by guests and reviews written by hosts are highly positive, negative evaluation is extremely rare, and lukewarm review is perhaps a strategy to communicate nonpositive evaluation. Fradkin et al. (2015) report some summary statistics on Airbnb reviews for both hosts and guests, and the focus of the study is an incentivized review experiment to induce additional reviews for nonrated listings. They show that nonreviewers tend to have worse experiences than reviewers. Quattrone et al. (2018) perform a linguistic analysis on bilateral reviews produced over several years on Airbnb. They label both guest reviews and host reviews with four main themes and analyze them year by year. They find that, in guest reviews, utilitarian values are discussed much more frequently than social values, and this difference substantially increases over years, and the same trend applies to host reviews. No previous research has leveraged the content of bilateral reviews in a predictive setting, especially for the two-stage decision-making problem on platforms offering bilateral reviews.

In the introduction, we provided some high-level discussions about the importance of analyzing bilateral reviews and the differences between analyzing unilateral and bilateral reviews. Here, we expand our discussion on the methodological challenges to further illustrate that analyzing bilateral reviews is nontrivial and cannot be done by simply adopting existing methods for analyzing unilateral reviews.

First, bilateral reviews have a more complex topic structure than unilateral reviews. Buyer reviews mostly evaluate items (e.g., the room is clean), and seller reviews mostly discuss the quality of the buyer (e.g., the buyer is reliable). Therefore, they will have their own corpus-specific topics. At the same time, buyer reviews can also talk about the seller of the item (e.g., the seller is responsive). Because both types of reviews can evaluate users, they may share topics such as whether someone is easy to communicate with or friendly. Existing studies handling unilateral reviews with topic modeling (McAuley and Leskovec 2013, Büschken and Allenby 2016) normally use a single topic space to generate documents, which cannot capture the inherent topic structure among bilateral reviews. Some previous studies have used the concept of probabilistic mixture models and joint topic spaces to handle more than one type of documents. Zhai et al. (2004) proposed a cross-collection generative probabilistic mixture model based on probabilistic latent semantic index (Hofmann 1999) to discover the common themes across all collections and the ones unique to each collection. Paul and Girju (2009) extend the model in Zhai et al. (2004) to an LDA-based model for cross-culture topic analysis. Hong et al. (2011) model text streams from two news sources, Twitter and Yahoo! News, and extend standard topic models by allowing each text stream to have both local topics and shared topics. Gao et al. (2012) propose a topic model to summarize trending subjects by jointly discovering the representative and complementary information from news and tweets. Under the model proposed in Paul and Girju (2009) for comparing multiple text collections, each topic is associated with two classes of topics: a collection-independent one and a collection-specific one. Previous research has demonstrated the benefit of designing a topic structure to handle multiple text collections, but this idea has not been used in the relational topic model framework or the two-stage decision process, and it has not been used to process bilateral reviews to address the online transaction prediction problem on platforms with bilateral reviews. We integrate shared and corpus-specific topics for different types of documents in a relational topic modeling framework to address the online transaction prediction problem on platforms with bilateral reviews.

Another methodological challenge analyzing bilateral review faces is that almost all platforms with bilateral reviews have a two-stage decision-making process, and for each stage, different text sources are used to facilitate users' decisions. As we illustrate more about the process of generating features used to predict transactions in Section 4.4, designing a feature space that incorporates the topic structure of bilateral reviews can potentially provide better independent variables for online transaction prediction. We fill the gap as the first study that analyzes the content of bilateral reviews, and we propose a novel relational topic model to handle the distinctive topic space presented in bilateral reviews and the two-stage decision process that is the context where bilateral reviews are often used.

The problem we study in this paper is also related to online matching, which is a classic problem on reciprocal online platforms, such as online dating and online recruiting. Most existing studies in online matching focus on reciprocal recommendation where users are provided with recommendations of other individuals with preferences of both parties satisfied, and recommendations are generated based mostly on analyzing users' attributes and interactions in a social network using methods such as graph mining (Li and Li 2012, Kutty et al. 2014). Different from existing research that pays more attention to social interactions, online transaction prediction in the two-stage decision-making process defined in our study depends heavily on textual information in historical transactions.

3. Problem Description

In this section, we formally define the problem of using bilateral reviews and basic descriptions of the goods and services to predict the probability that a buyer and a seller reach a transaction in a two-stage decision-making process.

3.1. Preliminaries

On platforms with bilateral reviews, there are two types of user identities: guest (buyer) and host (seller). The host can share access to her goods or services to the guest. The same person can be both guest and host but will be treated as separate entities. Let $G = \{g_i\}$, $i = 1, 2, \cdots, I$, be a set of guests, $L = \{l_j\}$, $j = 1, 2, \cdots, J$, be a set of items (goods or services), and $H = \{h_k\}$, $k = 1, 2, \cdots, K$, be a set of hosts.

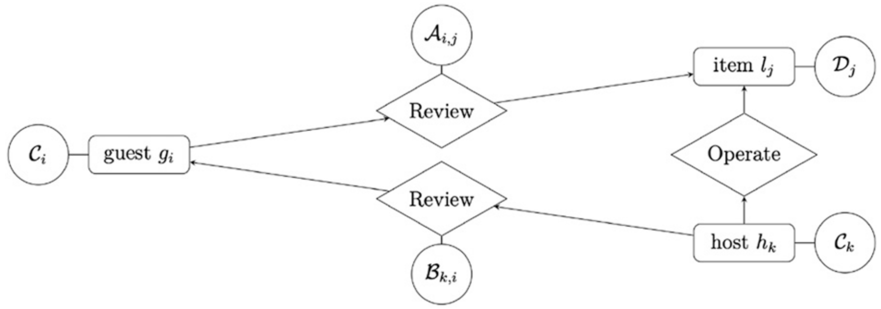
During the purchase decision-making process, the guest can search and choose a favorite item based on the information available about the host and the item. Such information includes the profile of the host, the basic description of the item, and the reviews for the item. After the guest makes a booking request, the host can decide whether to accept the request based on the information about the guest including the guest's profile and the reviews left for the guest. After the successful completion of a transaction during which guest g_i booked item l_j offered by host h_k, both the guest and the host may write reviews on each other. These reviews are called bilateral reviews. $\mathcal{A}_{i,j}$ represents the review that guest g_i has written about item l_j, and $\mathcal{B}_{k,i}$ represents the review that host h_k has written on guest g_i. Basic description for item l_j is represented by $\mathcal{D}_j$. The profile for guest g_i and host h_k contain typical features not derived from reviews (denoted by $\mathcal{C}_i$ and $\mathcal{C}_k$, respectively), such as whether the guest is domestic or international and the number of years active. Figure 1 describes the building blocks of the information we will use to address the prediction problem defined in Section 3.2.

3.2. Problem Definition

Following the notations and descriptions given in Section 3.1, our overall task is to learn the probability that guest g_i will successfully transact with host h_k on item l_j: $P(Success = 1 \mid g_i, l_j, h_k)$.

In order to estimate this probability, we consider a two-stage decision-making process: a requesting stage and an accepting stage. In the requesting stage, the guest examines the information of the potential item to see if it matches her preferences. The item information available to the guest includes its basic description $\mathcal{D}_j$ and all the reviews written by guests. In our predictive context, the guest's preferences can be derived from descriptions of all the items the guest has purchased and corresponding reviews the guest has written previously. Thus, we denote the set of historical items ever purchased by guest g_i as L_i. The collection of basic descriptions for L_i can be represented by $\mathcal{D}_{L_i}$. For item l_j, we use $\mathcal{A}_{(\cdot,j)}$ to represent the

Figure 1. Summary of Entities and Related Available Information



collection of reviews written for the item l_j by any guest. For guest g_i, we use $\mathcal{A}_{(i,\cdot)}$ to represent the collection of reviews that guest g_i has ever posted for any item. In addition, guest g_i will also take host h_k's profile $\mathcal{C}_k$ into consideration when making choices. In this context, the information available for us to predict the probability that guest g_i requests to book for item l_j in the first stage includes $\mathcal{D}_j$, $\mathcal{D}_{L_i}$, $\mathcal{A}_{(\cdot,j)}$, $\mathcal{A}_{(i,\cdot)}$, and $\mathcal{C}_k$, we can write this probability as follows:

$$P\Big(Y^1 = 1 \,|\, \mathcal{D}_j, \mathcal{D}_{L_i}, \mathcal{A}_{(\cdot,j)}, \mathcal{A}_{(i,\cdot)}, \mathcal{C}_k\Big).$$

If the guest has made a booking request in the first stage, then in the subsequent stage, the host will decide whether to accept the booking request. Before making the decision, the host checks the guest's basic profile $\mathcal{C}_i$ and past reviews for the guest written by other hosts to see if the guest meets the host's preferences. In our predictive context, the host's preferences can be learned from the collection of reviews this host has written for any guest. For a given host h_k, the collection of reviews that the host has ever written is denoted by $\mathcal{B}_{(k,\cdot)}$. For a guest g_i, we use $\mathcal{B}_{(\cdot,i)}$ to represent the collection of reviews that guest g_i has ever received. In this context, the information available for us to predict host h_k's decision in the second stage includes $\mathcal{C}_i$, $\mathcal{B}_{(k,\cdot)}$, and $\mathcal{B}_{(\cdot,i)}$. Under the condition that guest g_i made a booking request for item l_j in the first stage, the probability that host h_k accepts guest g_i's booking request for item l_j in the second stage can be denoted by

$$P\big(Y^2 = 1 \,|\, Y^1 = 1, \mathcal{B}_{(k,\cdot)}, \mathcal{B}_{(\cdot,i)}, \mathcal{C}_i\big).$$

Based on the probabilities of the two stages, we can obtain the probability that guest g_i and host h_k successfully reach the transaction on item l_j as follows:

$$P\big(Success = 1 \,|\, g_i, l_j, h_k\big) = P\Big(Y^1 = 1 \,|\, \mathcal{D}_j, \mathcal{D}_{L_i}, \mathcal{A}_{(\cdot,j)}, \mathcal{A}_{(i,\cdot)}, \mathcal{C}_k\Big) \times P\big(Y^2 = 1 \,|\, Y^1 = 1, \mathcal{B}_{(k,\cdot)}, \mathcal{B}_{(\cdot,i)}, \mathcal{C}_i\big).$$

4. Model Establishment

In this section, a new relational topic model is developed to predict transaction results by analyzing item descriptions and bilateral reviews with shared and corpus-specific topics taken into consideration. Variational methods are used for approximate inference and parameter estimation, and a comprehensive analysis of complexity and scalability is provided.

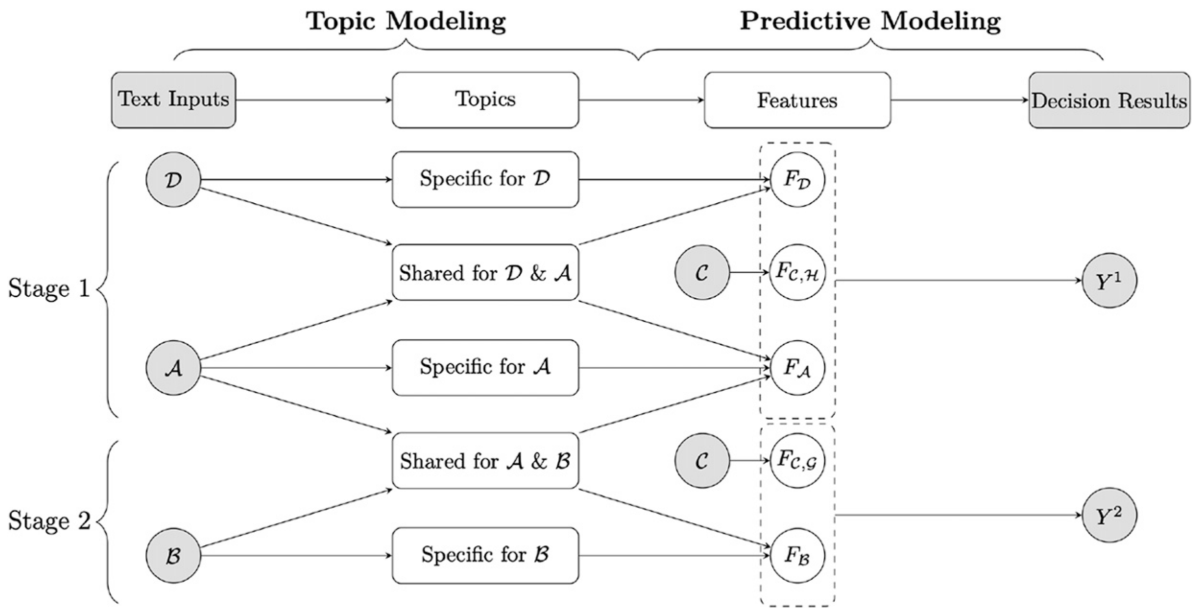
4.1. Overview of Model Structure

To model the two-stage decision-making process, we develop a generative model to process different types of documents constructed from item descriptions and bilateral reviews and use them to predict transaction results. Figure 2 graphically illustrates how we use item descriptions ($\mathcal{D}$), bilateral reviews (guest reviews $\mathcal{A}$ and host reviews $\mathcal{B}$), and guest/host profiles ($\mathcal{C}$) to predict potential transaction results (Y). Note that $\mathcal{D}$, $\mathcal{A}$, and $\mathcal{B}$ here refer to the documents we construct from these three types of text inputs, and we will explain how the documents are constructed in Sections 4.3 and 4.4.

The entire framework has two stages, and each stage has two components: topic modeling and predictive modeling. In the topic modeling part, we construct features based on topics learned from available text inputs. Then in the predictive modeling part, we consider features generated from the topic modeling part together with features presented in basic user profiles and learn the relationship between these features and transaction results for that stage. In each stage, the textual information available for decision making is different. In the first stage (stage 1) where the guest decides whether to request a booking, the guest will consider item descriptions ($\mathcal{D}$) and past guest reviews ($\mathcal{A}$). In the second stage (stage 2) where the host decides whether to accept the guest's booking request, the host will consider past host reviews ($\mathcal{B}$).

In our modeling framework, there are altogether three types of textual inputs: item descriptions, guest reviews, and host reviews. There exist some intrinsic

Figure 2. Graphical Overview of Our Relational Topic Modeling Framework



connections among these three types of textual information, and they should not be processed entirely separately. Item descriptions are generated by hosts to introduce items and hosts, and guest reviews are created by guests to evaluate items and hosts. Because these two types of textual information are both describing items and hosts, they should naturally share some common topics. On the other hand, they are written by different parties for different purposes; therefore, they should also have their own specific topics, which are called corpus-specific topics. Taking Airbnb as an example, topics related to the convenience of the location can be discussed in item (listing) descriptions and in guest reviews. However, not every topic covered in item descriptions is discussed in guest reviews and vice versa. For example, in item descriptions, hosts can talk about policies and procedures regarding the stay (e.g., check-in policy), which are very unlikely to be discussed in guest reviews. Likewise, guests can provide evaluation of the listings/hosts (e.g., had a wonderful stay) in their reviews, which will not appear in item descriptions at all. Given this intrinsic connection between these two types of text inputs, it might not be the best to assume that both item descriptions and guest reviews are generated by the same set of topics or by two separate sets of topics when implementing the topic modeling part. Similarly, guest reviews and host reviews can also share common topics while having their own specific topics. Guest reviews evaluate both items and hosts, and host reviews mostly evaluate guests. Because both types of reviews can evaluate users, they may share topics such as whether someone is easy to communicate with. On the other hand, these two types of reviews clearly have their specific topics as guest reviews can talk about items (e.g., nice room) and host reviews can evaluate guests on aspects that do not apply to hosts (e.g., following house rules). Because the content of item descriptions and host reviews have very little in common, we do not consider shared topics between them, and our experiments on real-world data sets also show that they barely have any shared topics. Under a topic modeling framework, a generative process designed to specifically incorporate the unique content structure presented in bilateral reviews and item descriptions is more likely to fit the data better, which could lead to better predictive performance.

Our topic structure with both shared and corpus-specific topics is outlined in Figure 2. In the first stage, topics about item description ($\mathcal{D}$) and guest reviews ($\mathcal{A}$) are used, and in the second stage, only topics about host reviews ($\mathcal{B}$) are used. We will explain the details in Section 4.3 and 4.4.

4.2. The Basic RTM

In the original RTM proposed by Chang et al. (2010), documents are summarized by topics, and the link status between a pair of documents can be predicted based on the similarity between the topic representations of the two documents. If the documents are abstracts of scientific papers, the links are citations. First, all the abstracts of the scientific papers are used to generate the topic space containing a set of topics. For each paper, there is a vector describing its topic representation over this topic space. For a given pair of papers, Chang et al. (2010) use the Hadamard (element-wise) product between the two topic vectors to generate a vector used as features to predict whether there is a link/citation between them. The number of features is the same as the number of topics. In the RTM model, the topic modeling component is used to generate the topic space, and the predictive modeling component is used to predict whether there is a citation. However, these two components are not independent, because citations between documents are conditionally dependent on the topics. Because document contents and links among documents are jointly modeled, parameters for the topic modeling component and the link prediction component are learned simultaneously. Chang et al. (2010) also demonstrated that this approach can achieve better results than learning parameters for these two components separately.

The model we propose in this paper shares some similarities with the original RTM model in that we do not consider topic modeling and predictive modeling as two separate and successive steps; instead, we learn the parameters for both components at the same time. In our setting, instead of predicting whether there is a citation between two papers, we predict whether there is a transaction between a guest and an item. Our model has several significant extensions on top of the original RTM model. First, the decision process we study has two stages. For each stage, there is a topic modeling component and a predictive modeling component. Our model jointly estimates the parameters for two topic modeling components and two predictive modeling components. Second, we have a topic structure with both shared and corpus-specific topics discovered from three different types of text inputs. In the original RTM model, there is only one type of text input and one topic space. Another major difference with the original RTM model is that the topic vector used to generate features in our model is not just corresponding to one single document; instead, it is constructed from three different types of textual information (i.e., guest reviews, host reviews, and item descriptions). In addition, we study whether there is a link (transaction) between two different types of entities (a guest and an item), and the link in the original RTM is between two instances of the same entity (e.g., scientific papers). Because our model extends the original RTM to the context of bilateral reviews, we call it BRTM (bilateral relational topic model).

To further illustrate the difference between our BRTM and the original RTM, we first present the modeling details of the decision problem in the second stage where the host decides whether to accept the guest's booking request. The model setup for the second stage is similar to that of the original RTM model and much easier than that for the first stage where the guest decides whether to request a booking. In the second stage, when host h_k decides to accept the booking request from guest g_i, a link is established between two constructed documents: $\mathcal{B}_{(k,\cdot)}$, which is the collection of reviews that the host has ever written, and $\mathcal{B}_{(\cdot,i)}$, which is the collection of reviews that guest g_i has ever received. Because the situation in the first stage is more complex involving two different types of text inputs (i.e., item descriptions and guest reviews) and a more complicated topic space, illustrating the host decision modeling first (in Section 4.3) will help us better understand the guest decision modeling process (in Section 4.4).

4.3. BRTM for Host Decision Modeling

In host decision modeling, we learn the hosts' preferences and guests' performances to estimate the probability of a host accepting a guest's booking request. A host h_k's preferences for guests can be represented by a topic vector extracted from $\mathcal{B}_{(k,\cdot)}$, which is the collection of reviews ever written by this host. Likewise, a guest g_i's performance can be represented by a topic vector extracted from $\mathcal{B}_{(\cdot,i)}$, which is the collection of reviews ever received by this guest. As both $\mathcal{B}_{(k,\cdot)}$ and $\mathcal{B}_{(\cdot,i)}$ contain only reviews written by hosts, they can be mapped to the same topic space through topic models and used to predict whether the host accepts the guest's request. Two types of host review documents, $\mathcal{B}_{(k,\cdot)}$ and $\mathcal{B}_{(\cdot,i)}$, are used in the second stage of our model. The unit of analysis (document) in our topic model is not individual reviews; instead, it is the collection of reviews used to represent host preferences ($\mathcal{B}_{(k,\cdot)}$) or guest performance ($\mathcal{B}_{(\cdot,i)}$). We will discuss how documents are constructed in the first stage in Section 4.4.

In BRTM, each document is first generated from topics, a latent layer between documents and words as in the basic LDA model (Blei et al. 2003). To be more specific, each document is endowed with a Dirichlet-distributed vector of topic proportions. Each word of the document is assumed to be generated by first drawing a topic assignment from those topic proportions and then drawing the word from the corresponding topic distribution. The binary dependent variable is whether a host accepts a guest's booking request. For the host decision stage, the independent variables include features constructed from topic assignments of the two documents ($\mathcal{B}_{(k,\cdot)}$ and $\mathcal{B}_{(\cdot,i)}$) and conventional features if any (e.g., guest profiles). Without loss of generality, we adopt the logit function as the link probability function, which corresponds to using logistic regression to predict the probability of a host accepting a guest's booking request. Because of the dependency between topics extracted from texts and features used for prediction, we can simultaneously learn both topic distributions and logistic regression coefficients.

Figure 3 illustrates the details of the host decision modeling without considering shared topics. For host review document $\mathcal{B}_{(k,\cdot)}$ (representing host h_k's preferences) and $\mathcal{B}_{(\cdot,i)}$ (representing guest g_i's performance), the topic proportion of the document is denoted by $\theta_k^{\mathcal{B}}$ and $\theta_i^{\mathcal{B}}$, respectively. For each word position n in document $\mathcal{B}_{(k,\cdot)}$, a topic assignment $z_{k,n}^{\mathcal{B}}$ is first generated based on $\theta_k^{\mathcal{B}}$. Next, the word on word position n, $w_{k,n}^{\mathcal{B}}$, is generated by a corresponding topic-specific word distribution $\phi_{z_{k,n}^{\mathcal{B}}}^{\mathcal{B}}$ for topic $z_{k,n}^{\mathcal{B}}$. The list of topics, each represented by its word distribution, is denoted by $\phi^{\mathcal{B}}$. Similar logic applies to document $\mathcal{B}_{(\cdot,i)}$. In addition, $\alpha_{\mathcal{B}}$ and $\eta_{\mathcal{B}}$ are hyper-parameters used in the Dirichlet distributions to generate $\theta^{\mathcal{B}}$ and $\phi^{\mathcal{B}}$. With this generative process, every document can be represented by a topic vector. The dimension of the topic vector is the same as the number of topics, and the topic vector indicates the percentage of word positions in a document belonging to each topic. For documents $\mathcal{B}_{(k,\cdot)}$ and $\mathcal{B}_{(\cdot,i)}$, the corresponding topic vectors are $\bar{\mathbf{z}}_k^{\mathcal{B}}$ and $\bar{\mathbf{z}}_i^{\mathcal{B}}$. After using topic vectors to represent host preferences and guest performances, we can integrate these vectors in fixed dimensions into the classification model for host decisions. We adopt the same treatment as in the original RTM to use the element-wise product between topic vectors as features $f_{\mathcal{B}}$:

$$f_{\mathcal{B}} = \bar{\mathbf{z}}_k^{\mathcal{B}} \circ \bar{\mathbf{z}}_i^{\mathcal{B}}. \tag{1}$$

In addition to topic-related features $f_{\mathcal{B}}$, we also include the conventional features $f_{c,\mathcal{G}}$ describing the guest's basic profile. Then, we use the logistic function to connect features and transaction results:

$P_\sigma(Y_{k,i}^2 = 1) = \sigma(\beta_{\mathcal{B}} f_{\mathcal{B}} + \beta_{c,\mathcal{G}} f_{c,\mathcal{G}} + \beta_2)$; $P_\sigma(Y_{k,i}^2 = 0) = 1 - \sigma(\beta_{\mathcal{B}} f_{\mathcal{B}} + \beta_{c,\mathcal{G}} f_{c,\mathcal{G}} + \beta_2)$, where $\sigma(x) = \frac{1}{1+\exp(-x)}$, $\beta_{\mathcal{B}}$, and $\beta_{c,\mathcal{G}}$ are logistic regression coefficients, and β_2 is the intercept term.

If we do not consider shared topics between guest reviews ($\mathcal{A}$) and host reviews ($\mathcal{B}$) as indicated in the topic modeling part of stage 2 in Figure 2, $\phi^{\mathcal{B}}$ will contain topics specific to $\mathcal{B}$, and Figure 3 will be the exact adaptation of the original RTM model to the host decision-making process if we do not include the profile-based features ($f_{c,\mathcal{G}}$). In the more general case, shared topics between $\mathcal{A}$ and $\mathcal{B}$ and specific topics for $\mathcal{B}$ are both taken into account, and so $\phi^{\mathcal{B}}$ will include these two types of topics: $\phi^{\mathcal{A}\&\mathcal{B}}$ (shared topics between $\mathcal{A}$ and $\mathcal{B}$) and $\phi^{\mathcal{B}^*}$ (specific topics for $\mathcal{B}$). The number of shared topics is $T_{\mathcal{A}\&\mathcal{B}}$, and $T_{\mathcal{B}^*}$ is the number of specific topics for $\mathcal{B}$. The hyper-parameters used in the Dirichlet distributions to generate $\phi^{\mathcal{A}\&\mathcal{B}}$ and $\phi^{\mathcal{B}^*}$ are $\eta_{\mathcal{A}\&\mathcal{B}}$ and $\eta_{\mathcal{B}^*}$. Figure 4 includes these two types of topics and is the host decision modeling component incorporated into the complete BRTM framework which also contains the guest decision modeling we illustrate in Section 4.4.

4.4. BRTM for Guest Decision Modeling

In guest decision modeling, we estimate the probability that a guest makes a booking request for a given item. The textual information available in this stage include item descriptions and guest reviews. A guest g_i's preferences for items can be learned from two constructed documents: the collection of the descriptions of all the items the guest has purchased before ($\mathcal{D}_{L_i}$) and the collection of all the reviews the guest has written about previous items ($\mathcal{A}_{(i,\cdot)}$). An item l_j can be described by two other constructed documents: its own description ($\mathcal{D}_j$), and the collection of all the historical guest reviews this item has received ($\mathcal{A}_{(\cdot,j)}$). These documents are used to generate independent variables for the prediction problem in this stage to estimate whether a guest made a booking request for an item. Table 1 illustrates the two topic vectors used to generate the independent variables.

Topic Vector 1 represents guest g_i's preferences, and Topic Vector 2 describes item l_j. Topic Vector 1 is constructed by concatenating topic vector for $\mathcal{D}_{L_i}$ and that for $\mathcal{A}_{(i,\cdot)}$. Topic Vector 2 is generated in a similar fashion. Documents $\mathcal{D}_{L_i}$ and $\mathcal{D}_j$ are all based on item descriptions; therefore, topics for item descriptions are used to construct their topic vectors.

Figure 3. Graphical Representation for Host Decision Modeling Without Shared Topics

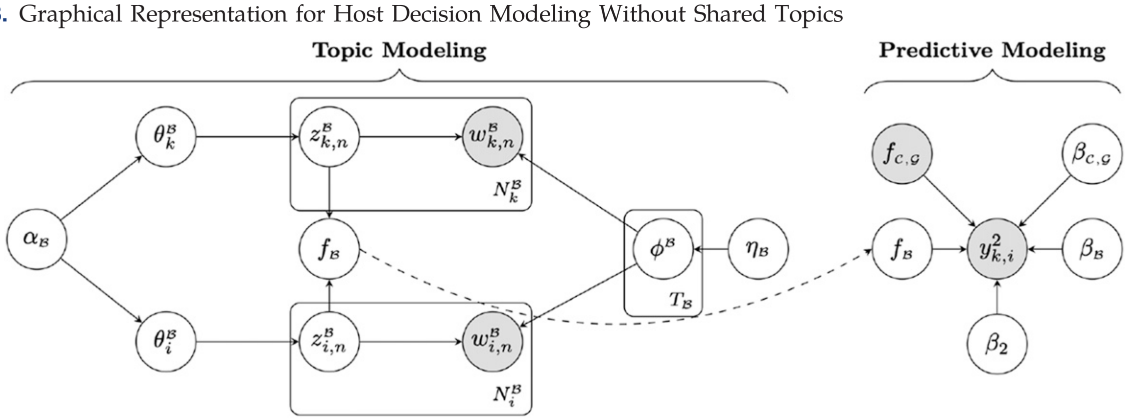
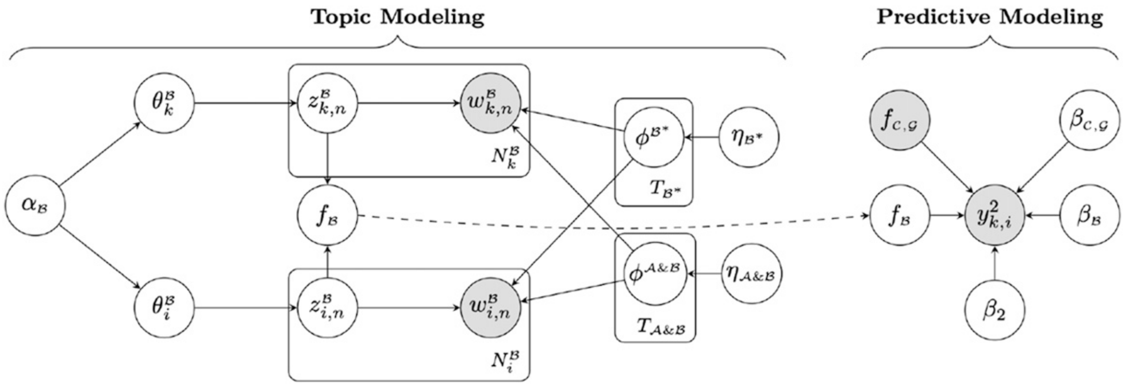


Figure 4. Graphical Representation for Host Decision Modeling



Similarly, documents $\mathcal{A}_{(i,\cdot)}$ and $\mathcal{A}_{(\cdot,j)}$ are constructed from guest reviews; thus, their topic vectors are generated based on the topics for guest reviews. We then apply element-wise product on Topic Vectors 1 and 2 to generate features based on textual information. Together with other profile-based features, they are used as the final set of independent variables to predict whether the guest makes a booking request for the item.

The graphical representation of our BRTM model for guest decision modeling is shown in Figure 5. There are four topic subspaces in the guest decision stage of BRTM, namely, one shared topic subspace $\phi^{\mathcal{D}\&\mathcal{A}}$ between item descriptions and guest reviews, another shared topic subspace $\phi^{\mathcal{A}\&\mathcal{B}}$ between guest reviews and host reviews, and two corpus-specific topic subspaces $\phi^{\mathcal{D}^*}$ and $\phi^{\mathcal{A}^*}$ for item descriptions and guest reviews, respectively. For a certain item description based document $\mathcal{D}_{L_i}$ (similarly for $\mathcal{D}_j$), its document-specific topic proportion is denoted by $\theta_i^{\mathcal{D}}$. Each word position in the document is first assigned to a latent topic $z_{i,n}^{\mathcal{D}}$: either a corpus-specific topic or a shared topic. According to the assigned topic, the observed word is generated by a corresponding topic-specific word distribution. If the topic $z_{i,n}^{\mathcal{D}}$ is a corpus-specific topic for item description documents, the observed word $w_{i,n}^{\mathcal{D}}$ is generated by $\phi^{\mathcal{D}^*}_{z_{i,n}^{\mathcal{D}}}$. Otherwise, the topic $z_{i,n}^{\mathcal{D}}$ is a shared one and the observed word is generated by $\phi^{\mathcal{D}\&\mathcal{A}}_{z_{i,n}^{\mathcal{D}}}$. For guest review based documents ($\mathcal{A}_{(i,\cdot)}$ or $\mathcal{A}_{(\cdot,j)}$), they are generated by three topic subspaces $\phi^{\mathcal{D}\&\mathcal{A}}$, $\phi^{\mathcal{A}^*}$ and $\phi^{\mathcal{A}\&\mathcal{B}}$, and the generation process is similar to that of $\mathcal{D}_{L_i}$. Note that $\phi^{\mathcal{A}\&\mathcal{B}}$ is also used in the host decision process to generate host review based documents. In this topic space, $\alpha_{\mathcal{D}}$, $\alpha_{\mathcal{A}}$, $\eta_{\mathcal{D}}$, $\eta_{\mathcal{D}\&\mathcal{A}}$, $\eta_{\mathcal{A}}$, and $\eta_{\mathcal{A}\&\mathcal{B}}$ are hyper-parameters of Dirichlet distributions to generate topic proportions for documents and word distributions for topics.

For Topic Vectors 1 and 2 in Table 1, topics for item descriptions (the second column) include $\phi^{\mathcal{D}^*}$ and $\phi^{\mathcal{D}\&\mathcal{A}}$, and topics for guest reviews (the third column) consist of $\phi^{\mathcal{D}\&\mathcal{A}}$, $\phi^{\mathcal{A}^*}$, and $\phi^{\mathcal{A}\&\mathcal{B}}$. Based on the learned topics, item description and guest review documents can be represented by topic vectors. For a given document $\mathcal{D}_{L_i}$ (or $\mathcal{D}_j$), its topic vector is denoted by $\bar{\mathbf{z}}_i^{\mathcal{D}}$ (or $\bar{\mathbf{z}}_j^{\mathcal{D}}$). Likewise, for a given document $\mathcal{A}_{(i,\cdot)}$ (or $\mathcal{A}_{(\cdot,j)}$), its topic vector is $\bar{\mathbf{z}}_i^{\mathcal{A}}$ (or $\bar{\mathbf{z}}_j^{\mathcal{A}}$). Similar to Equation (1), we can obtain features using the element-wise product between topic vectors: $f_{\mathcal{D}} = \bar{\mathbf{z}}_i^{\mathcal{D}} \circ \bar{\mathbf{z}}_j^{\mathcal{D}}, f_{\mathcal{A}} = \bar{\mathbf{z}}_i^{\mathcal{A}} \circ \bar{\mathbf{z}}_j^{\mathcal{A}}$, and then Table 1 can then be expressed more specifically as Table 2 to illustrate the corresponding features.

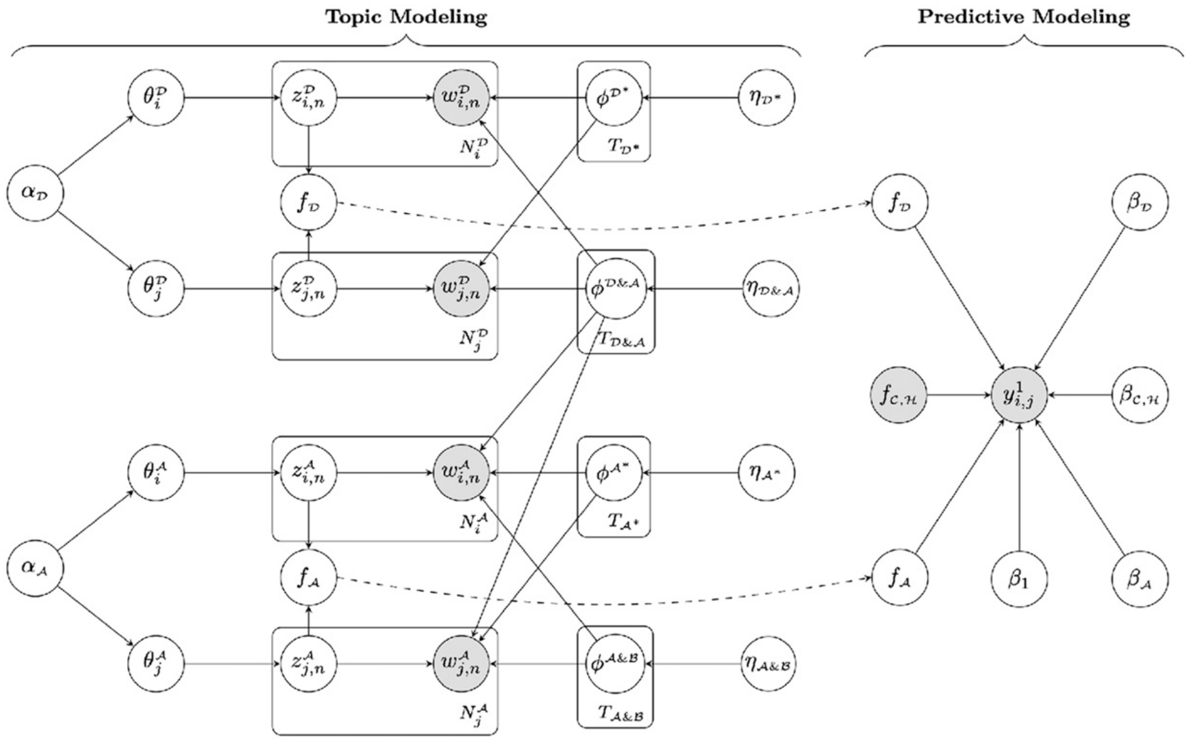
In addition to these features based on topics, we also include the host's basic profile $\mathcal{C}_k$ as additional features $f_{c,\mathcal{H}}$. Then, we use the logistic link function to connect features and the probability of a booking request in the first stage: $P_\sigma(Y_{i,j}^1 = 1) = \sigma(\beta_{\mathcal{D}} f_{\mathcal{D}} + \beta_{\mathcal{A}} f_{\mathcal{A}} + \beta_{c,\mathcal{H}} f_{c,\mathcal{H}} + \beta_1); P_\sigma(Y_{i,j}^1 = 0) = 1 - \sigma(\beta_{\mathcal{D}} f_{\mathcal{D}} + \beta_{\mathcal{A}} f_{\mathcal{A}} + \beta_{c,\mathcal{H}} f_{c,\mathcal{H}} + \beta_1)$, where $\beta_{\mathcal{D}}$, $\beta_{\mathcal{A}}$ and $\beta_{c,\mathcal{H}}$ are logistic regression coefficients, and β_1 is the intercept term. If guest g_i makes a booking request for item l_j, then $Y_{i,j}^1 = 1$, otherwise $Y_{i,j}^1 = 0$.

The differences between the original RTM model and our BRTM model can be further explained using Figures 3–5. From the modeling perspective, the original RTM is Figure 3, and our BRTM is Figure 5 (stage 1) plus Figure 4 (stage 2). In addition, BRTM also incorporates how to assemble different text inputs into different types of documents, and how different documents are used to form topic vectors representing the guest and the item.

Table 1. Topic Vectors Based on Item Descriptions and Guest Reviews

	Topics for item descriptions	Topics for guest reviews
Topic Vector 1 for guest g_i	Topic Vector for $\mathcal{D}_{L_i}$	Topic Vector for $\mathcal{A}_{(i,\cdot)}$
Topic Vector 2 for item l_j	Topic Vector for $\mathcal{D}_j$	Topic Vector for $\mathcal{A}_{(\cdot,j)}$

Figure 5. Graphical Representation of BRTM for Guest Decision Modeling



4.5. The Entire Generative Process of the Two-Stage Decision Making

To integrate item descriptions, bilateral reviews, and transaction results in both stages, we summarize the entire generative process as follows. All required variables are displayed in Table 3.

The documents used in the previous generative process are formed by a collection of reviews (or item descriptions), because they are the right type of documents needed to represent guests and items in order to construct topic vectors (Part I ~ Part IV) and then features used for transaction prediction (Part V).

Generative Process: The Two-Stage Decision Making

Part I: Preparation.
For each topic $t_{\mathcal{D}^*}$:
  Draw word distribution $\phi^{\mathcal{D}^*}_{t_{\mathcal{D}^*}} \mid \eta_{\mathcal{D}^*} \sim Dir(\eta_{\mathcal{D}^*})$.
For each topic $t_{\mathcal{D}\&\mathcal{A}}$:
  Draw word distribution $\phi^{\mathcal{D}\&\mathcal{A}}_{t_{\mathcal{D}\&\mathcal{A}}} \mid \eta_{\mathcal{D}\&\mathcal{A}} \sim Dir(\eta_{\mathcal{D}\&\mathcal{A}})$.
For each topic $t_{\mathcal{A}^*}$:
  Draw word distribution $\phi^{\mathcal{A}^*}_{t_{\mathcal{A}^*}} \mid \eta_{\mathcal{A}^*} \sim Dir(\eta_{\mathcal{A}^*})$.
For each topic $t_{\mathcal{A}\&\mathcal{B}}$:
  Draw word distribution $\phi^{\mathcal{A}\&\mathcal{B}}_{t_{\mathcal{A}\&\mathcal{B}}} \mid \eta_{\mathcal{A}\&\mathcal{B}} \sim Dir(\eta_{\mathcal{A}\&\mathcal{B}})$.
For each topic $t_{\mathcal{B}^*}$:
  Draw word distribution $\phi^{\mathcal{B}^*}_{t_{\mathcal{B}^*}} \mid \eta_{\mathcal{B}^*} \sim Dir(\eta_{\mathcal{B}^*})$.
Part II: Item Description Documents ($\mathcal{D}$).
For each document $\mathcal{D}_{L_i}$ (similar process for $\mathcal{D}_j$):
  Draw topic proportions $\theta^{\mathcal{D}}_i \mid \alpha_{\mathcal{D}} \sim Dir(\alpha_{\mathcal{D}})$.
  For each word $w^{\mathcal{D}}_{i,n}$:
    Draw topic assignment $z^{\mathcal{D}}_{i,n} \mid \theta^{\mathcal{D}}_i \sim Multi(\theta^{\mathcal{D}}_i)$.
    If $z^{\mathcal{D}}_{i,n}$ is a corpus-specific topic:
      Draw word $w^{\mathcal{D}}_{i,n} \mid z^{\mathcal{D}}_{i,n}, \phi^{\mathcal{D}^*} \sim Multi(\phi^{\mathcal{D}^*}_{z^{\mathcal{D}}_{i,n}})$.
    Else (If $z^{\mathcal{D}}_{i,n}$ is a shared topic):
      Draw word $w^{\mathcal{D}}_{i,n} \mid z^{\mathcal{D}}_{i,n}, \phi^{\mathcal{D}\&\mathcal{A}} \sim Multi(\phi^{\mathcal{D}\&\mathcal{A}}_{z^{\mathcal{D}}_{i,n}})$.
Part III: Guest Review Documents ($\mathcal{A}$).
For each document $\mathcal{A}_{(i,\cdot)}$ (similar process for $\mathcal{A}_{(\cdot,j)}$):
  Draw topic proportions $\theta^{\mathcal{A}}_i \mid \alpha_{\mathcal{A}} \sim Dir(\alpha_{\mathcal{A}})$.
  For each word $w^{\mathcal{A}}_{i,n}$:
    Draw topic assignment $z^{\mathcal{A}}_{i,n} \mid \theta^{\mathcal{A}}_i \sim Multi(\theta^{\mathcal{A}}_i)$.
    If $z^{\mathcal{A}}_{i,n}$ is a corpus-specific topic:
      Draw word $w^{\mathcal{A}}_{i,n} \mid z^{\mathcal{A}}_{i,n}, \phi^{\mathcal{A}^*} \sim Multi(\phi^{\mathcal{A}^*}_{z^{\mathcal{A}}_{i,n}})$.
    Else If $z^{\mathcal{A}}_{i,n}$ is a shared topic for $\mathcal{D}$ and $\mathcal{A}$:
      Draw word $w^{\mathcal{A}}_{i,n} \mid z^{\mathcal{A}}_{i,n}, \phi^{\mathcal{D}\&\mathcal{A}} \sim Multi(\phi^{\mathcal{D}\&\mathcal{A}}_{z^{\mathcal{A}}_{i,n}})$.
    Else If $z^{\mathcal{A}}_{i,n}$ is a shared topic for $\mathcal{A}$ and $\mathcal{B}$:
      Draw word $w^{\mathcal{A}}_{i,n} \mid z^{\mathcal{A}}_{i,n}, \phi^{\mathcal{A}\&\mathcal{B}} \sim Multi(\phi^{\mathcal{A}\&\mathcal{B}}_{z^{\mathcal{A}}_{i,n}})$.
Part IV: Host Review Documents ($\mathcal{B}$).
For each document $\mathcal{B}_{(k,\cdot)}$ (similar process for $\mathcal{B}_{(\cdot,i)}$):
  Draw topic proportions $\theta^{\mathcal{B}}_k \mid \alpha_{\mathcal{B}} \sim Dir(\alpha_{\mathcal{B}})$.
  For each word $w^{\mathcal{B}}_{i,n}$:
    Draw topic assignment $z^{\mathcal{B}}_{k,n} \mid \theta^{\mathcal{B}}_k \sim Multi(\theta^{\mathcal{B}}_k)$.
    If $z^{\mathcal{B}}_{k,n}$ is a corpus-specific topic:
      Draw word $w^{\mathcal{B}}_{k,n} \mid z^{\mathcal{B}}_{k,n}, \phi^{\mathcal{B}^*} \sim Multi(\phi^{\mathcal{B}^*}_{z^{\mathcal{B}}_{k,n}})$.
    Else (If $z^{\mathcal{B}}_{k,n}$ is a shared topic):
      Draw word $w^{\mathcal{B}}_{k,n} \mid z^{\mathcal{B}}_{k,n}, \phi^{\mathcal{A}\&\mathcal{B}} \sim Multi(\phi^{\mathcal{A}\&\mathcal{B}}_{z^{\mathcal{B}}_{k,n}})$.
Part V: Transaction Status (Y).
For each pair of guest g_i and candidate item l_j:
  Draw binary link indicator:
  $Y^1_{i,j} \mid \bar{\mathbf{z}}^{\mathcal{D}}_i, \bar{\mathbf{z}}^{\mathcal{D}}_j, \bar{\mathbf{z}}^{\mathcal{A}}_i, \bar{\mathbf{z}}^{\mathcal{A}}_j \sim P_\sigma(\cdot \mid f_{\mathcal{D}}, f_{\mathcal{A}}, f_{\mathcal{C},\mathcal{H}}, \beta_{\mathcal{D}}, \beta_{\mathcal{A}}, \beta_{\mathcal{C},\mathcal{H}}, \beta_1)$.
For each pair of host h_k and candidate guest g_i:
  Draw binary link indicator:
  $Y^2_{k,i} \mid \bar{\mathbf{z}}^{\mathcal{B}}_k, \bar{\mathbf{z}}^{\mathcal{B}}_i \sim P_\sigma(\cdot \mid f_{\mathcal{B}}, f_{\mathcal{C},\mathcal{G}}, \beta_{\mathcal{B}}, \beta_{\mathcal{C},\mathcal{G}}, \beta_2)$.

Table 2. Features Based on Item Descriptions and Guest Reviews

	Topics for item descriptions: $\phi^{\mathcal{D}^*}$ and $\phi^{\mathcal{D}\&\mathcal{A}}$	Topics for guest reviews: $\phi^{\mathcal{D}\&\mathcal{A}}$, $\phi^{\mathcal{A}^*}$, and $\phi^{\mathcal{A}\&\mathcal{B}}$
Topic Vector 1 for guest g_i	Topic vector for $\mathcal{D}_{L_i}$: $\bar{\mathbf{z}}_i^{\mathcal{D}}$	Topic vector for $\mathcal{A}_{(i,\cdot)}$: $\bar{\mathbf{z}}_i^{\mathcal{A}}$
Topic Vector 2 for item l_j	Topic vector for $\mathcal{D}_j$: $\bar{\mathbf{z}}_j^{\mathcal{D}}$	Topic vector for $\mathcal{A}_{(\cdot,j)}$: $\bar{\mathbf{z}}_j^{\mathcal{A}}$
Feature vector	$f_{\mathcal{D}}$	$f_{\mathcal{A}}$

Table 3. Summary of the Required Variables

Variable	Description	Variable	Description
$\mathcal{D}$: $\mathcal{D}_j$ and $\mathcal{D}_{L_i}$	Item description documents	$\eta = \{\eta_{\mathcal{D}^*}, \eta_{\mathcal{D}\&\mathcal{A}}, \eta_{\mathcal{A}^*}, \eta_{\mathcal{A}\&\mathcal{B}}, \eta_{\mathcal{B}^*}\}$	Hyper-parameters to generate Φ
$\mathcal{A}$: $\mathcal{A}_{(\cdot,j)}$ and $\mathcal{A}_{(i,\cdot)}$	Guest review documents	$\Theta = \{\theta^{\mathcal{D}}, \theta^{\mathcal{A}}, \theta^{\mathcal{B}}\}$	Topic proportions
$\mathcal{B}$: $\mathcal{B}_{(k,\cdot)}$ and $\mathcal{B}_{(\cdot,i)}$	Host review documents	$\alpha = \{\alpha_{\mathcal{D}}, \alpha_{\mathcal{A}}, \alpha_{\mathcal{B}}\}$	Hyper-parameters to generate Θ
Y: $Y^1_{i,j}$ and $Y^2_{k,i}$	Transaction results	$Z = \{z^{\mathcal{D}}, z^{\mathcal{A}}, z^{\mathcal{B}}\}$	Topic assignments
$\Phi = \{\phi^{\mathcal{D}^*}, \phi^{\mathcal{D}\&\mathcal{A}}, \phi^{\mathcal{A}^*}, \phi^{\mathcal{A}\&\mathcal{B}}, \phi^{\mathcal{B}^*}\}$	Word distributions	$F = \{f_{\mathcal{D}}, f_{\mathcal{A}}, f_{\mathcal{C},\mathcal{H}}, f_{\mathcal{B}}, f_{\mathcal{C},\mathcal{G}}\}$	Extracted features
$T = \{T_{\mathcal{D}}*, T_{\mathcal{D}\&\mathcal{A}}, T_{\mathcal{A}}*, T_{\mathcal{A}\&\mathcal{B}}, T_{\mathcal{B}}*\}$	Number of topics	$\beta = \{\beta_{\mathcal{D}}, \beta_{\mathcal{A}}, \beta_{\mathcal{C},\mathcal{H}}, \beta 1, \beta_{\mathcal{B}}, \beta_{\mathcal{C},\mathcal{G}}, \beta 2\}$	Logistic regression coefficients

In other settings where an individual review can be treated as a document, Part I ~ Part IV of our generative process can also be directly used to find shared and corpus-specific topics. Part V constructs features used for our predictive model; therefore, it only applies to our document definition.

4.6. Learning

To infer distributions and estimate parameters under our model, we focus on computing the posterior distribution of the latent variables (Z, Θ, Φ) conditioned on the observed textual documents (W) and links (Y) between them: $p(Z, \Theta, \Phi \mid W, Y, \alpha, \eta, \beta)$.

As with many hierarchical Bayesian models, the exact posterior inference for our model is intractable to compute because of the coupling between Θ and Φ in the summation over latent topics (Blei et al. 2003). Thus, we approximate the inference using variational methods (Blei et al. 2006) with comparable accuracy and improved computational efficiency. We develop a variational expectation-maximization (EM) algorithm to effectively predict the transaction success probability in the two-stage decision-making process. Technical details of learning steps are available in Online Appendix A. Our algorithm is fast to converge and easy to scale up. Comprehensive analysis of complexity and scalability is also provided in Online Appendix B. The analysis demonstrate that our model based on the topic space with both shared and specific topics has the same time complexity as the approach using all separate topic spaces for different documents and has lower space complexity.

We also show in the simulation in Online Appendix C that our proposed topic structure with shared and corpus-specific topics models the document generation process more accurately if the documents are indeed generated based on shared and corpus-specific topics. Therefore, in theory, our model should fit the data better. Our topic model is also a more general model that covers the case where there are limited or no shared topics.

5. Experiments

In this section, we conduct several experiments with real-world datasets to validate our ideas and evaluate our proposed models.

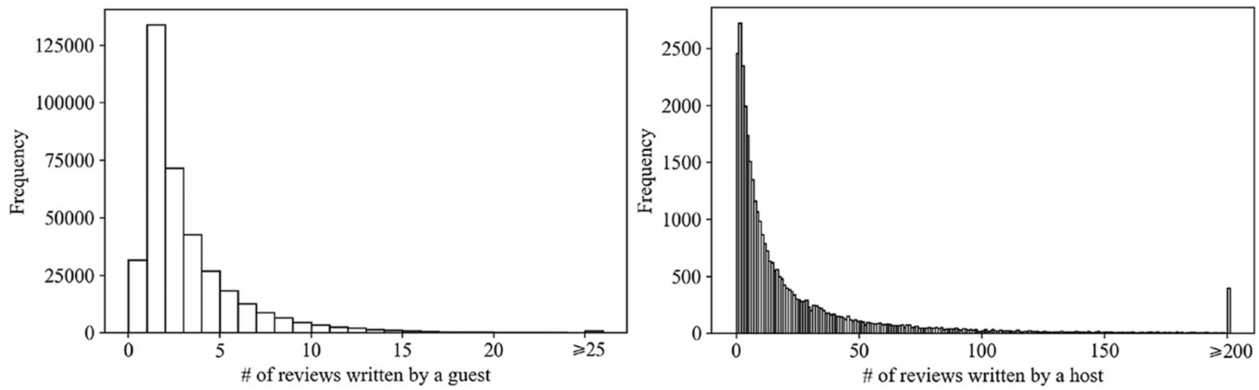
5.1. Data Description

We build our data set by collecting publicly available information from Airbnb, which is one of the most prevalent sharing economy platforms. First, we collect listings located in New York City through Inside Airbnb (http://insideairbnb.com). Then for the guests and hosts associated with the NYC listings, we crawl Airbnb.com for their relevant information up to the end of 2016, including listing descriptions, bilateral reviews, and user profiles (domestic or international, the number of years since joined, the number of verified sources, and the number of connected accounts). Descriptive statistics of our Airbnb-NYC data set are provided in Table 4. In addition to the Airbnb data set for New York City, we also conducted robustness tests using a data set from Airbnb

Table 4. Summary of the Airbnb-NYC Data Set

No. of guests	No. of hosts	No. of listings	No. of transactions	No. of reviews written by all guests	No. of reviews written by all hosts
374,525	34,416	43,395	425,336	1,488,456	1,706,788

Figure 6. Histograms of the Number of Reviews Written by a Guest and a Host



transactions in London and another data set from a boat sharing platform called Boatsetter.

The reviews written by all the guests in this data set include those written for listings not located in New York City. This is also the case for the reviews written by all the hosts contained in the data set. The histograms in Figure 6 depict the distribution of the number of reviews written by a guest and a host, respectively. The median and mean for the number of reviews a guest has written are 2 and 2.9, respectively, and for a host, the median is 9 and the mean is 24.0.

According to Airbnb's chief executive officer (CEO) and cofounder Brian Chesky, 80% of hosts leave a review for their guests and 72% of guests leave a review for hosts (Quora 2012). Although not every successful transaction will be reviewed by both the guest and the host, our proposed modeling framework is still workable, because the data we use to generate predictions are reviews pooled together for each guest and each host. A small percentage of transactions, which have no reviews, will not affect the aggregated review collection used to build the prediction model.

On Airbnb, some hosts may leave similar reviews for different guests. In our New York City data set, 19.5% of the host reviews are similar in content to some other reviews written by the same host. When generating the review collections used in our models, we only keep one review for each group of highly similar reviews by the same host and remove the redundant ones. Please see Online Appendix D for the details of the process we follow to identify and filter redundant reviews and a detailed description of host reviews, including an analysis to demonstrate the topic variance among the reviews.

In addition, 18.36% of the transactions in our data are instant bookings that result in automatic booking confirmation after the guest's request. We provide results with and then without instant bookings, and in both cases, our proposed model obtains significant performance improvements.

5.2. Data Issues

5.2.1. Two-Stage Data. Our model depicts two stages of decision making, and thus ideally, two sets of outcomes should be used when implementing the model. If the model is adopted by Airbnb, the platform will have access to the outcome of the first stage where the guest makes a booking request for a listing. However, the data we use in the experiments only contains successful transactions without showing the intermediate results in the requesting stage. Naturally, leveraging the additional outcomes in the first stage should most likely increase the performance of the prediction. In the experiments, we show that even without the results in the first stage, we can still achieve a significant performance increase over other alternatives. In order for our model to be applied directly to the data without the results in the first

Table 5. All Features for the Modified Model

	Topics for item descriptions: $\phi^{\mathcal{D}^*}$ **and** $\phi^{\mathcal{D}\&\mathcal{A}}$	Topics for guest reviews: $\phi^{\mathcal{D}\&\mathcal{A}}$, $\phi^{\mathcal{A}^*}$, **and** $\phi^{\mathcal{A}\&\mathcal{B}}$	Topics for host reviews: $\phi^{\mathcal{A}\&\mathcal{B}}$ **and** $\phi^{\mathcal{B}^*}$	Profile features
Topic Vector 1 for guest g_i	Topic Vector for $\mathcal{D}_{L_i}$: $\bar{\mathbf{z}}_i^{\mathcal{D}}$	Topic Vector for $\mathcal{A}_{(i,\cdot)}$: $\bar{\mathbf{z}}_i^{\mathcal{A}}$	Topic Vector for $\mathcal{B}_{(\cdot,i)}$: $\bar{\mathbf{z}}_i^{\mathcal{B}}$	
Topic Vector 2 for item l_j and its host h_k	Topic Vector for $\mathcal{D}_j$: $\bar{\mathbf{z}}_j^{\mathcal{D}}$	Topic Vector for $\mathcal{A}_{(\cdot,j)}$: $\bar{\mathbf{z}}_j^{\mathcal{A}}$	Topic Vector for $\mathcal{B}_{(k,\cdot)}$: $\bar{\mathbf{z}}_k^{\mathcal{B}}$	
Feature vector	$f_{\mathcal{D}}$	$f_{\mathcal{A}}$	$f_{\mathcal{B}}$	$f_{\mathcal{C},\mathcal{G}}$ and $f_{\mathcal{C},\mathcal{H}}$

Information Systems Research, 2021, vol. 32, no. 2, pp. 541–560, © 2021 INFORMS

Generative Process: Part V*

For each tuple of guest g_i, item l_j and host h_k:
Draw binary link indicator:
$$Y_{i,j,k} \mid \bar{\mathbf{z}}_i^{\mathcal{D}}, \bar{\mathbf{z}}_j^{\mathcal{D}}, \bar{\mathbf{z}}_i^{\mathcal{A}}, \bar{\mathbf{z}}_j^{\mathcal{A}}, \bar{\mathbf{z}}_k^{\mathcal{B}}, \bar{\mathbf{z}}_i^{\mathcal{B}} \sim P_\sigma(\cdot \mid f_{\mathcal{D}}, f_{\mathcal{A}}, f_{\mathcal{C},\mathcal{H}}, f_{\mathcal{B}}, f_{\mathcal{C},\mathcal{G}}, \beta_{\mathcal{D}}, \beta_{\mathcal{A}}, \beta_{\mathcal{C},\mathcal{H}}, \beta_{\mathcal{B}}, \beta_{\mathcal{C},\mathcal{G}}, \beta_0).$$
Meanwhile, the link probability is also rewritten as:
$$P_\sigma(Y_{i,j,k} = 1) = \sigma(\beta_{\mathcal{D}} f_{\mathcal{D}} + \beta_{\mathcal{A}} f_{\mathcal{A}} + \beta_{\mathcal{C},\mathcal{H}} f_{\mathcal{C},\mathcal{H}} + \beta_{\mathcal{B}} f_{\mathcal{B}} + \beta_{\mathcal{C},\mathcal{G}} f_{\mathcal{C},\mathcal{G}} + \beta_0);$$
$$P_\sigma(Y_{i,j,k} = 0) = 1 - \sigma(\beta_{\mathcal{D}} f_{\mathcal{D}} + \beta_{\mathcal{A}} f_{\mathcal{A}} + \beta_{\mathcal{C},\mathcal{H}} f_{\mathcal{C},\mathcal{H}} + \beta_{\mathcal{B}} f_{\mathcal{B}} + \beta_{\mathcal{C},\mathcal{G}} f_{\mathcal{C},\mathcal{G}} + \beta_0). \quad (2)$$

stage, we modify Part V of our entire generative process by making the following changes.

This modification is reasonable because we are essentially addressing the prediction problem of estimating the probability of a successful transaction given all the observed information. Having the intermediate results will most likely help the predictive accuracy, but it is not a necessary piece of information we have to obtain in order to build a predictive model. All features used in this modified model combine the features from each of the two stages as summarized in Table 5.

5.2.2. Negative Instances. In our data set, we can only observe the final listing a guest booked (i.e., the positive instances) and do not observe the listings which are viewed but rejected by the guest (i.e., the negative instances). The BRTM model adopting the logistic link function in Equation (2) needs both positive and negative instances for model training. Although the negative instances will be observed by Airbnb if our method is adopted, we do need to address this issue of lacking negative instances to evaluate our model in the experiments. We address this issue via two different strategies. The first strategy (BRTM-Rank) is based on a learning-to-rank model (Rendle et al. 2009), where the positive instances are considered ranked higher than the unobserved instances. In our scenario, the probability of an observed transaction is assumed higher than the probability that this guest would reach a transaction on listings active at that time but not actually booked. Instead of labeling negative instances for logistic regression, this strategy adopts a pairwise ranking loss function named Bayesian personalized ranking (Rendle et al. 2009) to maximize the difference of preference between observed and unobserved instances. The second strategy (BRTM-Sample) is to generate negative instances by sampling unobserved instances, and this strategy is also adopted in recommendation systems research where negative feedback is lacking (Hu et al. 2008, Pan et al. 2008, Pan and Scholz 2009, He et al. 2016). Negative instances are randomly selected from listings active at the time of the transaction and similar to the actually booked listing. Because the number of negative instances is far more than that of positive ones, we adopt a widely used subsampling solution (He and Garcia 2008) to balance the ratio of negative instances to positive ones in the model training stage, which is set at 1:1 for BRTM-Sample. The implementation details of these two strategies are provided in Online Appendix E (under Sections E.1 and E.2). In Section E.3 of Online Appendix E, we provide more discussions on other alternatives to handle unobserved negative instances in our context.

5.3. Baseline Models and Other Experimental Settings

We compare our proposed BRTM model with several alternative baselines listed in Table 6, including collaborative filtering, LDA, deep learning, RTM, and several other related models. Please see Online Appendix F for detailed setup of these baseline models.

For fair comparison, RTM-GH and BRTM-SEP use the same number of features (independent variables) as BRTM. Please see Online Appendix G (under Section G.1) for how the feature spaces of these three

Table 6. Summary of Alternative Baselines

Name	Description
RAND	Random selection
RANK-RATING	Rank listings according to the average ratings (Sarwar et al. 2001)
CF-G	Collaborative filtering based on reviews written by guests
LDA-G	Latent Dirichlet allocation based on reviews written by guests
STL	Supervised topic labeling (Abrahams et al. 2012)
DL-CNN	Deep learning model with convolutional neural network (Krizhevsky et al. 2012)
RTM-G	Relational topic model based on reviews written by guests
RTM-GH	Relational topic model that treats guest reviews and host reviews combined as one corpus and listing descriptions as the second corpus
BRTM-SEP	Bilateral relational topic model that handles bilateral reviews and item descriptions without considering shared topics; each text corpus has its own topic space separately

models (RTM-GH, BRTM-SEP, and BRTM) differ and the explanation for why BRTM is superior.

We split the transactions with reviews in 2016 into three subsets: the training set (the first eight months), the validation set (the next one month), and the test set (the last three months). The validation set is used for parameter selection and the test set is used for performance evaluation.

We evaluate BRTM-Rank and BRTM-Sample in the context of top-N recommendation. For each transaction in the validation or test set, we combine the actually booked listing with 19 other listings active at the time of the transaction to build the candidate set of size 20. Our model is then applied to predict the probability of successful transaction for each listing in the candidate set, and we choose the top N listings based on the predicted probability. To evaluate the top N listings, we adopt three widely used metrics in information retrieval and recommendation systems: hit rate (HR), mean reciprocal rank (MRR), and normalized discounted cumulative gain (NDCG). HR can be calculated by checking whether the actually booked listing is contained in the predicted top n listings, and MRR and NDCG further consider the rank of the booked listing in the predicted top n listings. For all these three metrics, the higher the score the better the model is.

For the models that involve LDA or RTM, the number of topics is provided as a fixed parameter before the learning process starts. To choose the number of topics, we first use perplexity score (Blei et al. 2003) on the validation set to select several candidate choices, and then use the predictive performance on the validation set to make the final selection. As the result, we choose 60 as the final number of topics for each individual type of documents. In addition, we examine the similarity between topics learned from individual type of documents to determine the number of shared and corpus-specific topics. The number of corpus-specific topics for item descriptions, guest reviews and host reviews are $T_{\mathcal{D}^*} = 38$, $T_{\mathcal{A}^*} = 29$, and $T_{\mathcal{B}^*} = 51$, respectively, and the number of shared topics between guest reviews and item descriptions is set as $T_{\mathcal{D}\&\mathcal{A}} = 22$ and that between guest reviews and host reviews is $T_{\mathcal{A}\&\mathcal{B}} = 9$. A detailed description about how to select the number of topics is provided in Online Appendix H. The hyperparameters of the generative process for the models based on LDA or RTM are fixed at $\alpha = 0.1$ and $\eta = 0.01$.

5.4. Experimental Results

5.4.1. Predictive Performance Comparisons. Following the experimental settings, we perform a comprehensive comparison of predictive performances between our proposed model and the alternatives. Both BRTM-Rank and BRTM-Sample significantly outperform all the comparable baselines based on HR, MRR, and NDCG, which fully demonstrates the robust benefit of designing specific models to properly handle bilateral reviews.

Because of the space limit, we only present the detailed results of BRTM-Sample based on HR in the main text because it also produces the logistic regression results. The rest of the results are in Online Appendix I (under Sections I.1 and I.2). As displayed in Table 7, BRTM-Sample shows a significantly better HR compared with other baselines as N varies.

Table 7. Comparison of HR with Baselines

Top-N	RAND	RANK-RATING	CF-G	LDA-G	STL	DL-CNN	RTM-G	RTM-GH	BRTM-SEP	BRTM-Sample
1	0.050 (+304.8%)	0.067 (+204.2%)	0.099 (+106.4%)	0.110 (+84.8%)	0.121 (+69.0%)	0.167 (+22.0%)	0.110 (+85.4%)	0.177 (+15.3%)	0.179 (+14.0%)	0.204
2	0.100 (+264.5%)	0.140 (+160.5%)	0.177 (+105.7%)	0.206 (+76.4%)	0.233 (+56.3%)	0.303 (+20.1%)	0.215 (+69.3%)	0.312 (+16.3%)	0.317 (+14.6%)	0.363
3	0.149 (+230.6%)	0.220 (+124.0%)	0.247 (+100.1%)	0.294 (+67.9%)	0.342 (+44.3%)	0.417 (+18.4%)	0.312 (+58.2%)	0.431 (+14.4%)	0.434 (+13.6%)	0.493
4	0.200 (+202.5%)	0.304 (+99.5%)	0.314 (+92.7%)	0.379 (+59.9%)	0.440 (+37.8%)	0.510 (+18.8%)	0.405 (+49.5%)	0.530 (+14.2%)	0.535 (+13.2%)	0.606
5	0.250 (+179.6%)	0.387 (+80.4%)	0.380 (+83.7%)	0.458 (+52.4%)	0.535 (+30.5%)	0.590 (+18.3%)	0.490 (+42.4%)	0.614 (+13.7%)	0.621 (+12.3%)	0.698
6	0.298 (+161.1%)	0.469 (+65.7%)	0.444 (+75.3%)	0.533 (+46.0%)	0.624 (+24.7%)	0.659 (+18.0%)	0.571 (+36.3%)	0.687 (+13.1%)	0.695 (+11.8%)	0.778
7	0.348 (+141.1%)	0.547 (+53.4%)	0.506 (+65.9%)	0.603 (+39.0%)	0.700 (+19.8%)	0.719 (+16.7%)	0.641 (+30.8%)	0.748 (+12.1%)	0.757 (+10.8%)	0.839
8	0.397 (+123.0%)	0.621 (+42.5%)	0.565 (+56.6%)	0.670 (+32.1%)	0.770 (+14.9%)	0.768 (+15.2%)	0.707 (+25.1%)	0.801 (+10.4%)	0.807 (+9.61%)	0.885
9	0.445 (+106.8%)	0.683 (+34.8%)	0.622 (+47.9%)	0.729 (+26.2%)	0.827 (+11.2%)	0.809 (+13.7%)	0.765 (+20.2%)	0.845 (+8.81%)	0.851 (+8.06%)	0.920
10	0.493 (+91.6%)	0.736 (+28.4%)	0.677 (+39.4%)	0.784 (+20.5%)	0.873 (+8.26%)	0.845 (+11.8%)	0.816 (+15.8%)	0.883 (+6.98%)	0.886 (+6.67%)	0.945

Note. STL, DL-CNN, RTM-G, RTM-GH, and BRTM-SEP all use the same negative instances as those in BRTM-Sample for fair comparisons.

Compared with BRTM-SEP, which assumes separate topic spaces for guest reviews, host reviews, and item descriptions, the HR improvement of BRTM-Sample ranges between 6.67% and 14.0%. This means that our topic structure containing shared and corpus-specific topics contributes a lot in predicting transaction success. BRTM-Sample is also 6.98%~15.3% better than RTM-GH, which treats bilateral reviews as unilateral reviews, and this improvement shows the value of designing a specific method to handle bilateral reviews. Compared with the RTM only using guest reviews (RTM-G), the relative increase in hit rate ranges from 15.8% to 85.4%. Because most prior research has only considered guest reviews, this huge improvement demonstrates the great value of leveraging bilateral reviews. Because DL-CNN does not use any topic structure, the results mean that our proposed topic structure contributes a lot in predicting transaction success. For HR, it is natural for improvement to decrease as N increases because bigger N is more tolerant for errors. Ranking the positive listing at the first spot is treated the same as ranking it at the Nth spot.

5.4.2. Topic Investigation and Logistic Regression Results. Our proposed model not only achieves superior predictive performance but also extracts coherent and comprehensive topics. To measure the coherence of the 149 topics discovered from all the three types of documents, we recruited eight master workers from Amazon MTurk to perform the word intrusion task designed by Chang et al. (2009). The coherence values generated by this study demonstrate that most of the topics are coherent. The details of this study are provided in Online Appendix J, which also report the top key words, coherence value, and the percentage of documents covered for each of the 149 topics. We also named each topic based on its top key words. Moreover, topics learned by our proposed model align well with the motivation of introducing shared and corpus-specific topics. After examining the representative words of both shared and corpus-specific topics discovered, we are able to identify shared topics that reflect common talking points embedded in both listing descriptions provided by hosts and reviews left by guests, as well as shared topics showing similar content between guest reviews and host reviews. Meanwhile, we also identified topics which are specific for different types of text inputs from the results. The example topics further illustrate the validity of introducing shared and corpus-specific topics that can provide a more comprehensive understanding of the topic structure embedded in the different types of text inputs. More details with examples are provided in Online Appendix K.

The coefficients of the topic-related features generated by logistic regression can help us further understand the drivers behind transaction success and failure. Among all the topic-related features, the vast majority (96.1%) are significant demonstrating the identified topics are contributing to the transaction prediction. We present some representative features with large coefficients and decent coherence in Table 8. From the guest side, whether the listing is in a "great location," "clean and nice," and providing "breakfast," and whether the host is "helpful and

Table 8. Logistic Regression Coefficients of Representative Features

Feature	Estimate	Standard error
"great location," corpus-specific, from guest review documents	81.1***	3.50
"helpful and comfortable," shared by $\mathcal{A}$ and $\mathcal{B}$, from guest review documents	70.0***	3.39
"clean and nice," shared by $\mathcal{D}$ and $\mathcal{A}$, from guest review documents	59.7***	3.49
"breakfast," corpus-specific, from guest review documents	56.4***	3.19
"great experience," shared by $\mathcal{A}$ and $\mathcal{B}$, from guest review documents	46.4***	3.41
"bar and restaurant," corpus-specific, from guest review documents	37.9***	3.51
"rooftop view," corpus-specific, from guest review documents	35.3***	3.45
"unique and great place," shared by $\mathcal{D}$ and $\mathcal{A}$, from guest review documents	32.5***	2.83
"cancellation before arrival," corpus-specific, from guest review documents	–6.12·	3.51
"friendly," corpus-specific, from host review documents	81.5***	3.53
"delightful," corpus-specific, from host review documents	65.8***	3.47
"family and friend," corpus-specific, from host review documents	63.3***	3.25
"quiet," corpus-specific, from host review documents	51.5***	3.47
"communicative," corpus-specific, from host review documents	50.3***	3.26
"smooth communication," shared by $\mathcal{A}$ and $\mathcal{B}$, from host review documents	49.9***	3.51
"fun talking," corpus-specific, from host review documents	49.5***	3.44
"considerate and responsive," corpus-specific, from host review documents	31.9***	2.24
"follow house rules," corpus-specific, from host review documents	31.6***	3.51

Note. Each feature is described by its related topic; complete results of the logistic regression are available in Online Appendix L.

$^{\cdot}p < 0.1$; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

comfortable" are among guests' main concerns before making booking requests. From the host side, whether the guest is "friendly," "delightful," "quiet," and "communicative" are considered by hosts when reviewing requests from guests.

Both significantly positive and negative features can help the platform understand influential factors for transaction success and failure. Taking the feature related to "clean and nice" from guest review documents as an example, its significantly positive coefficient means that guests who prefer clean and nice listings are more likely to request to book listings that are considered clean and nice by historical guests. On the contrary, if the listing's characteristics do not match the guest's preferences, the potential booking request would not be initiated. The same logic applies to the host side. The topic-related features derived from host review documents are designed to measure similarities between topic vectors representing host preferences and guest performances. Taking "friendly" from host review documents as an example, if a host who prefers friendly guests receives a booking request from a guest who is considered friendly by other hosts, the host is very likely to accept this booking request. If not so, this potential transaction is less likely to come through. Very few (0.6%) of the topic-related features are significantly negative. For example, the feature related to "cancellation before arrival" from guest review documents means that guests tend not to propose booking requests to hosts who canceled transactions before guests arrive in the past. If the host has good historical evaluations without any cancellation before arrival, the potential transaction might be successful.

Because we look into the content of bilateral reviews to reveal important factors imbedded in text for transaction success and failure, the additional insights obtained can complement some earlier research on investigating the reasons why booking requests from guests are rejected by hosts on Airbnb. Based on Airbnb's internal data, Fradkin (2017) reports three types of rejections: host screening, stale vacancy (unavailable listings were mistakenly marked as available), and congestion (more than one guest was trying to book the same dates), and host screening counts about a third of all the rejections. Because Fradkin (2017) did not examine the review contents, why guest booking requests were rejected during the host screening process was not clear. Our discussions about how some of the key factors affecting both guests' and hosts' decision-making process help shed light on the specific reasons for host screening rejection. In Table 8, the example topics derived from host reviews show that if prior host reviews evaluating a guest do not have information such as the guest being "friendly," "delightful," "quiet," or "communicative," the chance of the guest being rejected by host will be higher.

5.4.3. Robustness Check. Furthermore, we conduct sensitivity analysis on the number of shared topics, the review history window size, and the hyperparameter η used to generate word distributions, and the results demonstrate the robustness of our proposed model. Detailed analysis is provided in Online Appendix M. In addition, we lowered the number of topics for each individual type of text inputs from 60 to 30, and the conclusions remain the same as demonstrated in Online Appendix I (under Section I.3). We also conduct experiments after excluding all the listings with instant booking. About 20% of the transactions in our data set are filtered out. Following the same design as before, the comparisons show that our BRTM model still provides much better prediction performance than other models on the data containing no instant bookings. Detailed results are available in Online Appendix N. Moreover, we conduct experiments on another Airbnb data set about London and a data set from a boat sharing platform called Boatsetter for robustness check. Please see Online Appendix O for more details.

6. Applicability on Platforms with Bilateral Reviews

In addition to Airbnb, our BRTM framework can be adapted to other platforms where bilateral reviews with sufficient quality are available to assist different users to make more informed decisions. In recent years, more and more platforms offering bilateral reviews have emerged. In addition to the examples we discussed in the introduction, we listed many more such platforms in Online Appendix P.

The platforms that offer bilateral reviews often have one or several of the following characteristics. (1) Transaction success highly depends on trust between buyers and sellers. In the case of Airbnb, without bilateral reviews, hosts will not be able to trust the guests enough to let them stay in their homes. Past records of the guests in the form of host reviews will assure them that their listings will be taken good care of. Similarly, for the meal sharing platform Eatwith, a lot of hosts offer the service inside their homes, and some of them eat with the guests at the same table. The hosts would want to know whether the guests are pleasant enough based on their interaction from other hosts. (2) The product shared or service offered involves valuable properties. Boatsetter is a platform where boat owners can rent out their boats to people. Because of the high cost of potential boat damage, boat owners would be more comfortable if they can learn more about the renters' use behavior from reviews the renters received from previous boat owners. (3) The taskers (e.g., freelance professionals) need to interact with the consumers (people who request the service) enough during the task so they care about how well the consumers worked with other taskers before committing a long period of time on the projects. Platforms where people go to find taskers such as Fiverr, Handy, and TaskRabbit are all good examples. (4) Detailed product/service characteristics and buyer performance are contained in qualitative information instead of numeric ratings. Most platforms offering numeric bilateral ratings also allow users to write more detailed reviews so that they can better match their preferences. This is a common characteristic of platforms with bilateral reviews. (5) Almost all the platforms offering detailed bilateral reviews have a two-stage decision-making process, where the seller needs to decide whether to accept the buyer's request after reading the reviews for the buyer. Because of this two-stage decision process, instantaneous transactions are not common unless the seller is willing to take the risk by skipping the screening process (e.g., instant bookings on Airbnb). Even though most sharing economy platforms offer bilateral reviews, a few without the characteristics discussed above do not use bilateral reviews. For example, even though ride-sharing platforms such as Uber allows rider and drivers to evaluate each other, most reviews are numeric ratings, and reviews are not normally used for decision making because ride-hailing decisions are often made in a timely fashion.

The model proposed in this paper can be helpful to application problems where bilateral reviews are available to improve the matching mechanism between buyers and products/sellers (e.g., settings that can benefit from a recommender system, targeted marketing, direct marketing, etc.). For example, when a buyer is searching for a specific type of product on the platform (e.g., a two-bedroom apartment in New York City within a certain price range for a certain time period on Airbnb), the platform can present a list of products with the highest chance of generating a transaction within the buyer's search criteria. This is the main problem setting in this paper. In addition, the platform could monitor when a listing becomes more appealing to a guest as both sides accumulate reviews over time and then send target marketing emails to the guest. Moreover, the platform can design programs for the hosts to help them improve their item descriptions to cover topics that are appealing to their target guests who meet the hosts' requirements. With the increasing number of platforms allowing bilateral reviews, our model, especially the idea of leveraging bilateral reviews, can be adapted for more use cases to improve business and user processes.

7. Conclusions and Future Work

This paper proposes an integrated relational topic modeling framework to analyze item descriptions, bilateral reviews, and transaction results to improve matching efficiency on platforms with bilateral reviews. We define the transaction prediction problem as a two-stage decision process. A topic structure with both shared and corpus-specific topics is used to better handle topics from different types of documents which share some common topics. We adopt a variational EM method to learn our proposed model. Comprehensive experiments are conducted on real-world datasets collected from Airbnb and Boatsetter. Compared with existing methods, our model obtains significant and robust performance improvements for the transaction prediction problem. The shared and corpus-specific topics discovered further support the logic behind our modeling structure and the topic-related features reveal significant factors affecting transaction results.

The managerial implications of our study can be summarized from several different perspectives. From the buyer's perspective, our model will present the buyer with items she is more interested in booking, and thus improve the chance the buyer making a booking request. This is achieved through the first stage of our model. Without the buyer first interested in the items presented, there is no chance for the transaction to happen. Therefore, the listings presented to the buyer, first, have to be of interest to the buyer. For items that do not offer instant booking, the buyer must wait for the seller's acceptance once she makes the booking request in the first stage. Our model will lower the chance that the buyer's booking request is declined by the seller. The underlying reason is that when we generated the list presented to the buyer, we considered whether the buyer performance matches the seller preference through analyzing bilateral reviews. This will lower the chance that a buyer requests an item owned by a seller who does not like the buyer and thus turns down the booking request. This can lower the buyer's search cost on the platform because the buyer has to start new searches after her booking request is turned down. Avoiding the frustrations created by declined booking request can also increase the buyer's satisfaction and improve the overall buyer experience on the platform.

From the seller's perspective, our model will present the seller's items to buyers who are more likely to be liked by the seller. This is achieved through matching seller preferences and buyer performances. Both seller preferences and buyer performances are embedded in bilateral reviews and item descriptions. Our model is designed to process all the information to match the sellers and buyers better. After the seller receives a booking request from a buyer, she will read the reviews for the buyer from other sellers. If the seller does not like what she sees, she will decline the booking request. This will waste the seller's time. More importantly, if this happens more often, it will decrease the seller's satisfaction with the platform and even drive the seller away from the platform. By minimizing the chance of this happening, our model can effectively reduce the seller's transaction cost and improve the seller's overall experience on the platform.

From the platform's perspective, our integrated relational topic model incorporating item descriptions, bilateral reviews, and transaction results aims at generating more suitable items for buyers to increase the chance of a successful transaction. This can significantly improve system efficiency for platforms by lowering the chance of buyers' booking requests being denied and providing sellers with guests they are more likely to transact with. This improvement of the overall efficiency can increase transaction volume for the platform, which can drive up profit. It also increases user satisfaction with the platform, which helps the platform retain and attract more users. In addition, the results of our model can also help the platform understand some of the factors that lead to transaction success and failure and further realize the importance of matching preferences of both sides. The platform may even consider implementing a mechanism to provide guidance to service providers about how to optimize their item descriptions. Moreover, our research sheds light on the critical role of bilateral reviews in affecting transaction success on platforms with bilateral reviews. Because the final transaction decision can be influenced by reviews from both sides, merely using reviews written by buyers to only consider buyers' preferences is not enough. The platforms with bilateral reviews should realize the potential of the information generated by their reciprocal review systems. Instead of mixing bilateral reviews together to process them as if they are unilateral reviews or just using one-sided reviews, platforms should investigate bilateral reviews more to better understand the preferences of both buyers and sellers in order to match them more effectively.

Our research has several limitations that can be explored in future research. We use heuristic strategies to determine the number of shared and corpus-specific topics. In the future, we can explore to build a nonparametric model to learn the number of shared and corpus-specific topics automatically. In addition, we use the element-wise product, the same treatment as in the original RTM, to transform topic vectors into features used in logistic regression. Going forward, other functions and methods can be investigated to generate features based on topic vectors. Related to information systems research and business research in general, we plan to explore and identify other business applications that can benefit from our modeling framework handling multiple text sources and multiple decision-making stages.

Acknowledgments

The authors thank the senior editor, the associate editor, and three anonymous reviewers for insightful comments and suggestions throughout the review process. Hongyan Liu is the corresponding author of this paper.

References

Abrahams AS, Jiao J, Wang GA, Fan W (2012) Vehicle defect discovery from social media. *Decision Support Systems* 54(1):87–97.

Ansari A, Li Y, Zhang JZ (2018) Probabilistic topic model for hybrid recommender systems: A stochastic variational Bayesian approach. *Marketing Sci.* 37(6):987–1008.

Bao Y, Datta A (2014) Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Sci.* 60(6): 1371–1391.

Blei DM, Jordan MI (2006) Variational inference for Dirichlet process mixtures. *Bayesian Anal.* 1(1):121–143.

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J. Machine Learning Res.* 3(Jan):993–1022.

Bridges J, Vásquez C (2018) If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues Tourism* 21(18):2057–2075.

Büschken J, Allenby GM (2016) Sentence-based text analysis for customer reviews. *Marketing Sci.* 35(6):953–975.

Chang J, Blei DM (2010) Hierarchical relational models for document networks. *Ann. Appl. Statist.* 4(1):124–150.

Chang J, Gerrish S, Wang C, Boyd-Graber JL, Blei DM (2009) Reading tea leaves: How humans interpret topic models. *Adv. Neural Inform. Processing Systems* 22:288–296.

Chattopadhyay M, Mitra SK (2019) Do Airbnb host listing attributes influence room pricing homogenously? *Internat. J. Hospitality Management* 81:54–64.

Chen N, Zhu J, Xia F, Zhang B (2015) Discriminative relational topic models. *IEEE Trans. Pattern Anal. Machine Intelligence* 37(5):973–986.

Cui R, Li J, Zhang DJ (2020) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on Airbnb. *Management Sci.* 66(3):1071–1094.

Fradkin A (2017) Search, matching, and the role of digital marketplace design in enabling trade: Evidence from Airbnb. MIT Sloan Working Paper, Boston, MA.

Fradkin A, Grewal E, Holtz D, Pearson M (2015) Bias and reciprocity in online reviews: Evidence from field experiments on Airbnb. *Proc. 16th ACM Conf. Econom. Comput.*, 641–682.

Fu Y, Liu G, Teng M, Aggarwal C (2017) Unsupervised P2P rental recommendations via integer programming. *Proc. 23rd ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining*, 165–173.

Gao W, Li P, Darwish K (2012) Joint topic modeling for event summarization across news and social media streams. *Proc. 21st ACM Internat. Conf. Inform. Knowledge Management (CIKM)*, 1173–1182.

Ghoshal A, Menon S, Sarkar S (2015) Recommendations using information from multiple association rules: A probabilistic approach. *Inform. Systems Res.* 26(3):532–551.

Gong J, Abhishek V, Li B (2018) Examining the impact of keyword ambiguity on search advertising performance: A topic model approach. *Management Inform. Systems Quart.* 42(3):805–829.

He H, Garcia EA (2008) Learning from imbalanced data. *IEEE Trans. Knowledge Data Engrg.* 9:1263–1284.

He X, Zhang H, Kan MY, Chua TS (2016) Fast matrix factorization for online recommendation with implicit feedback. *Proc. 39th Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval*, 549–558.

Hofmann T (1999) Probabilistic latent semantic indexing. *Proc. 22nd Annu. Internat. ACM SIGIR Conf. Res. and Development in Inform. Retrieval*, 50–57.

Hong L, Dom B, Gurumurthy S, Tsioutsiouliklis K (2011) A time-dependent topic model for multiple text streams. *Proc. 17th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining*, 832–840.

Hu Y, Koren Y, Volinsky C (2008) Collaborative filtering for implicit feedback datasets. *Proc. 8th IEEE Internat. Conf. Data Mining (ICDM)*, 263–272.

Huang AH, Lehavy R, Zang AY, Zheng R (2017) Analyst information discovery and interpretation roles: A topic modeling approach. *Management Sci.* 64(6):2833–2855.

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Processing Systems* 25:1097–1105.

Kutty S, Nayak R, Chen L (2014) A people-to-people matching system using graph mining techniques. *World Wide Web* 17(3):311–349.

Lawani A, Reed MR, Mark T, Zheng Y (2019) Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Sci. Urban Econom.* 75:22–34.

Li L, Li T (2012) MEET: A generalized framework for reciprocal recommender systems. *Proc. 21st ACM Internat. Conf. Inform. Knowledge Management (CIKM)*, 35–44.

Li Z, Fang X, Bai X, Sheng ORL (2016) Utility-based link recommendation for online social networks. *Management Sci.* 63(6):1938–1952.

Luo Y, Tang RL (2019) Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). *Internat. J. Hospitality Management* 80:144–154.

McAuley J, Leskovec J (2013) Hidden factors and hidden topics: Understanding rating dimensions with review text. *Proc. 7th ACM Conf. Recommender Systems (RecSys)*, 165–172.

Niu Z, Hua G, Gao X, Tian Q (2014) Semi-supervised relational topic model for weakly annotated image recognition in social media. *Proc. 2014 IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, 4233–4240.

Pan R, Scholz M (2009) Mind the gaps: Weighting the unknown in large-scale one-class collaborative filtering. *Proc. 15th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining*, 667–676.

Pan R, Zhou Y, Cao B, Liu NN, Lukose R, Scholz M, Yang Q (2008) One-class collaborative filtering. *Proc. 8th IEEE Internat. Conf. Data Mining*, 502–511.

Pant G, Sheng OR (2015) Web footprints of firms: Using online isomorphism for competitor identification. *Inform. Systems Res.* 26(1):188–209.

Paul M, Girju R (2009) Cross-cultural analysis of blogs and forums with mixed-collection topic models. *Proc. 2009 Conf. Empirical Methods Natural Language Processing*, 1408–1417.

Puranam D, Narayan V, Kadiyali V (2017) The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors. *Marketing Sci.* 36(5):726–746.

Quattrone G, Nicolazzo S, Nocera A, Quercia D, Capra L (2018) Is the sharing economy about sharing at all? A linguistic analysis of Airbnb reviews. *Proc. 20th Internat. AAAI Conf. Web Social Media*.

Quattrone G, Proserpio D, Quercia D, Capra L, Musolesi M (2016) Who benefits from the "Sharing" economy of Airbnb? *Proc. 25th Internat. Conf. World Wide Web*, 1385–1394.

Quora (2012) What percent of Airbnb hosts leave reviews for their guests? Accessed December 19, 2020. https://www.quora.com/What-percent-of-Airbnb-hosts-leave-reviews-for-their-guests#.

Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L (2009) BPR: Bayesian personalized ranking from implicit feedback. *Proc. 25th Conf. Uncertainty Artificial Intelligence*, (AUAI Press, Corvallis, OR), 452–461.

Sanchez-Vazquez R, Silva J, Santos RL (2017) Exploiting socio-economic models for lodging recommendation in the sharing economy. *Proc. 11th ACM Conf. Recommender Systems (RecSys)*, (ACM, New York, NY), 260–268.

Sarwar B, Karypis G, Konstan J, Riedl J (2001) Item-based collaborative filtering recommendation algorithms. *Proc. 10th Internat. Conf. World Wide Web (WWW)*, (ACM, New York, NY), 285–295.

Wang D, Nicolau JL (2017) Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *Internat. J. Hospitality Management* 62:120–131.

Zhai C, Velivelli A, Yu B (2004) A cross-collection mixture model for comparative text mining. *Proc. 10th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining*, (ACM, New York, NY), 743–748.