1-1-2010

DESIGNING RANKING SYSTEMS FOR HOTELS ON TRAVEL SEARCH ENGINES TO ENHANCE USER EXPERIENCE

Anindya Ghose
New York University, aghose@stern.nyu.edu

Panagiotis G. Ipeirotis
New York University, panos@stern.nyu.edu

Beibei Li
New York University, bli@stern.nyu.edu

DESIGNING RANKING SYSTEMS FOR HOTELS ON TRAVEL SEARCH ENGINES TO ENHANCE USER EXPERIENCE

Completed Research Paper

Anindya Ghose
Stern School of Business, New York University
44 West 4th Street, New York, NY 10012
aghose@stern.nyu.edu

Panagiotis G. Ipeirotis
Stern School of Business, New York University
44 West 4th Street, New York, NY 10012
panos@stern.nyu.edu

Beibei Li
Stern School of Business, New York University
44 West 4th Street, New York, NY 10012
bli@stern.nyu.edu

Abstract

Information seeking in an online shopping environment is different from classical relevance-based information retrieval. In this paper, we focus on understanding how humans seek information and make economic decisions, when interacting with an array of choices in an online shopping environment. Our study is instantiated on a unique dataset of US hotel reservations from Travelocity.com. Current travel search engines display results with only rudimentary interfaces by using a single ranking criterion. This largely ignores consumers' multi-dimensional preferences and is constrained by human cognitive limitations towards multiple and diverse online data sources. This paper proposes to improve the search interface using an inter-disciplinary approach. It combines text mining, image classification, social geo-tagging and field experiments with structural modeling techniques, and designs an interface whereby each hotel of interest is ranked according to its average consumer surplus on the travel search site. This system would display the hotels with the "best value for money" for a given consumer at the top of the computer screen during the travel search process in response to a query. The final outcome based on several user studies shows that our proposed ranking system outperforms the existing baselines. This suggests that our inter-disciplinary approach has the potential to enhance user search experience for information seeking in the context of economic decision-making.

Keywords: Travel search, Information Seeking, User Interface Design, Social Media, Ranking Systems, Text Mining, Image classification, Econometrics

Introduction

Computational systems are at the very heart of every information-seeking activity today. When humans seek information, this activity is typically mediated today by a computer (e.g., Marchionini 1997, Ford et al. 2002). How can such computer-based information systems be properly designed to acquire and present effective information to humans? How can the system interfaces be implemented in a way to facilitate the information seeking behavior of humans? These questions have motivated both academics and practitioners across diverse communities of information systems, computer science, psychology and cognitive science. Over the last few years, a tremendous amount of research focused on how to optimize retrieval of relevant documents from the Web, mainly as a response to a keyword query (e.g., Lavrenko & Croft 2001, Agichtein et al. 2001). Increasingly, though, the focus shifts towards understanding the broader context in which humans seek information: Information seeking is rarely the end goal; rather it is the means for making a decision (e.g., Snipes et al. 2005, Lopatoska 2007).

More recently, with the growing pervasiveness of electronic commerce, special attention has been drawn on information seeking where there is an economic benefit or a cost for (some of) the parties that participate in the information exchanges (e.g., Jiang & Benbasat 2007, Jiang et al. 2010, Lopatovska & Mokros 2007). It becomes crucial for electronic marketplace designers and users (e.g., business firms) to understand how humans seek information and make economic decisions, when interacting with an array of choices in an online shopping environment. One typical context for this is the online search for products. Customers try to identify products of high quality with specific desired characteristics, but without compromising on their associated prices. However, given the proliferation of available information from the Web, customers normally do not have the sophistication or time to conduct exhaustive searches to seek the quality or price information in order to compare for the "best value" product.

Unfortunately, today's travel search engines provide only rudimentary ranking facilities, typically using a single ranking criterion. For example, on the Travelocity website customers can only choose to sort hotels by name, price, or star rating, etc. Furthermore, although these travel search engines have access to a lot of information, including product specification, price, or buyer experience from various types of user-generated content (such as textual feedback), they simply lack the ability to incorporate the plethora of information into the rankings. As a result, the current ranking mechanism has quite a few shortcomings. First, it ignores the multidimensional preferences of consumers. Second, it fails to leverage the textual information generated by the online communities beyond the numerical ratings. Third, it hardly takes into account the heterogeneity of consumers. These drawbacks highly necessitate a new recommendation strategy for travel search engines that can better perceive consumers' underlying demand, and facilitate the information seeking activities for consumers to make efficient economic decisions.

In this paper, we focus on how consumers search and make decisions when purchasing a product. By using a combination of structural demand estimation, image classification, and text mining techniques, we propose a novel way for ranking results for product search, based on a "consumer utility maximization" framework that takes into account both social and cognitive factors involved in human decision-making processes. Based on the utility theory, we propose to design a new ranking system for recommendation that leverages economic modeling. The outcome of our study provides a set of products ranked based on providing the "best value for money" to consumers. Such a ranking system can be displayed in response to a user search query on travel search engines.

We instantiate our study in the context of demand for hotel rooms by using a unique dataset of hotel reservations from Travelocity.com. The dataset contains complete information on transactions conducted over a 3-month period from 11/2008 to 1/2009 for 1497 hotels in the United States (US). We have data on user-generated content from three sources: (i) user-generated hotel reviews from two well-known travel search engines, Travelocity.com and Tripadvisor.com, (ii) social-geo tags generated by users identifying different geographic attributes of hotels from Geonames.org, and (iii) user-contributed opinions on the most important hotel characteristics using on-demand surveys and social annotations from users on Amazon Mechanical Turk. Moreover, since some location-based characteristics, such as proximity to the beach, are not directly measurable based on user-generated content, we use image classification techniques to infer such features from the satellite images of the area. These different data sources are then merged to create one comprehensive dataset summarizing the location and service characteristics of all the hotels. Our empirical modeling and analyses enables us to compute the "value for money" of a particular hotel based on the estimation of price elasticities and consumer surplus. Thereafter, we aim to generate hotel rankings that are superior to existing ranking techniques seen in travel search engines.

Our work involves four steps:

1) Identify the important hotel location and service characteristics that influence hotel demand.
2) Estimate how these hotel characteristics influence demand and quantify their marginal effects.
3) Improve search for hotels by incorporating the economic impacts of the hotel characteristics.
4) Evaluate system performance based on user studies.

More specifically, in the first step, we determine the particular hotel characteristics that are most highly valued by customers, and thus, contribute to the aggregate room prices of the hotels. Beyond the directly observable characteristics, such as the "number of stars," provided by most third-party travel websites, many users also tend to value location characteristics, such as proximity to the beach, or proximity to downtown shopping areas. In our work, we incorporate satellite image classification techniques and use both human and computer intelligence (in the form of social geo-tagging and text mining of reviews) to infer these location features. In the second step, we use demand estimation techniques (BLP 1995, Berry and Pakes 2007, Song 2010) and estimate the economic value associated with each hotel characteristic. This enables us to quantitatively analyze how each feature influences demand and estimate its importance relative to the other features. In the third step, we incorporate the derived economic value into designing a ranking system based on the consumer surplus from a given hotel. By doing so, we can provide customers with the "best-value" hotels early on. In the last step, we evaluate our approach based on user studies. The results show that our system performs significantly better compared to the existing baselines, therefore it can improve the quality of online travel search.

Our key contributions can be summarized as follows. First, we demonstrate the value of incorporating multiple and diverse online data sources towards enhancing the experience of user online search and decision-making. Customers today make their decisions in an environment with the plethora of available data. Take the hotel market for example, it is possible that some consumers check the characteristics of the hotels using tourist guides and mapping applications, or consult online review sites to determine the quality of the hotel and its amenities. In order to replicate this decision-making environment, we construct an exhaustive dataset, collecting as much information as possible about the hotels in our data, using a variety of data sources, and a variety of methodologies such as text mining, image classification and social geo-tagging. The implications of the new design can be great: Not only in directly integrating information accumulated by other humans in the decision-making process of an individual, but also in ways to provide potential to personalize the interactions between search interfaces and decision makers in future. Naturally, the systems and algorithms built in this research will be available on the web, allowing everyone to take advantage of the improved interfaces for accessing information and making better decisions.

Furthermore, our research sheds lights on the potential of exploring the economic value of social media data, and incorporating that into electronic marketplace design. We achieve this by introducing a structural model for demand estimation that builds on the work of classical consumer economic choice. Facilitated by text mining techniques, we are able to bridge the gap between the qualitative nature of social media data and the quantitative nature of economic choice models. Our empirical estimates enable us to propose a new ranking system for travel search based on the consumer surplus. The key notion is that in response to a consumer search query, the system would recommend and rank those products higher that provide a higher "value for money" by taking into account consumers' multi-dimensional preferences. By using several different experiments conducted on AMT, we show that our proposed ranking performs significantly better than several baseline-ranking systems that are being currently used. Besides providing direct economic gains, such a system can lead to non-trivial reduction in consumer search costs for the quality and price information of differentiated goods. This will in the mean time increase the usage of travel search engines, facilitate electronic marketplaces as intermediaries between buyers and sellers, and improve the economic efficiency.

Prior Literature

We initiate our research focus on the following two questions: (a) How can we efficiently seek information about product attributes expressed in diverse online social media data, such as the textual reviews and social tags? (b) How can we properly structure and present the acquired information in the context of the overarching decision-making process by using the right economic decision model? To examine these questions, our paper draws from multiple streams of work.

Information Seeking

Our work is related to human's behavior of information seeking, particularly in an electronic environment. Recent

studies have shown that the way that knowledge is organized and made available affects the way that information seekers are able to access this knowledge and thus their information seeking performance (Case 2007). The online search system can support human information seeking by structuring knowledge and constraining access (Marchionini 1997). A variety of researches thereafter focused on improving the design and implementation of the knowledge interface on the Internet (e.g., Agarwal & Venkatesh 2002, Jiang & Benbasat 2007). One common goal in these studies is to form an effective representation of information and to enhance the interactions between the information systems and the end-users. More recent studies started to associate system interface design with users' economic behavior. For example, Jiang et al. (2010) used a laboratory experiment approach to show that website interactivity can impact purchase intention through website involvement. In general, our work contributes to this field in achieving the same goal to enhance human information seeking for decision-making. Moreover, our inter-disciplinary approach is the first to combine economic decision models with text mining, image classification and social geo-tagging techniques, to improve the quality of the knowledge interface and the knowledge body.

Economic Impact of User-Generated Content

Our work is related to estimating the economic impact of online user-generated content. One key challenge when incorporating various types of social media data into the information seeking process is to bridge the gap between the textual and qualitative nature of review content and the quantitative nature of economic choice models. With the rapid growth and popularity of the user-generated content on the Web, a new area of research applying text mining technique to product reviews has emerged. The first stream of this research has focused on the sentiment analysis of product reviews (Hu & Liu 2004, Pang & Lee 2004, Das & Chen 2007). This stimulated additional research on identifying product features in which consumers expressed their opinions (Hu & Liu 2004, Scaffidi et al. 2007, Snyder & Barzilay 2007). The automated extraction of product attributes has also received attention in the recent marketing literature (Lee & Bradlow 2007).

The hypothesis that product reviews affect human economic decisions has received strong support in prior empirical studies (for example, Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Dellarocas et al. 2007, Duan et al. 2008, Forman et al. 2008, Moe 2009). However, these studies focus only on numeric review ratings (e.g., the valence and volume of reviews) in their empirical analysis. Only a handful of empirical studies have formally tested whether the textual information embedded in online user-generated content can have an economic impact (Ghose et al. 2006, Eliashberg et al. 2007, Archak et al. 2008, Ghose 2009, Ghose and Ipeirotis 2010). However, these studies do not focus on estimating the impact of user-generated reviews in influencing sales beyond the effect of numeric review ratings. In addition, researchers using only numeric ratings have to deal with issues like self-selection bias (Li and Hitt 2008) and bimodal distribution of reviews (Hu et al. 2008). More importantly, the matching of consumers to hotels in numerical rating systems is not random. A consumer only rates the hotel that she frequents (i.e. the one that maximizes her utility). Consequently, the average star rating for each hotel need not reflect the population average utility. Due to the above drawbacks, the average numerical star rating assigned to a product may not convey a lot of information to a prospective buyer. Therefore, a key objective of this paper is to analyze the extent to which textual content and linguistic style of user-generated reviews can help us understand human economic decisions.

Demand Estimation

Our work is related to models of demand estimation. One model that has made a significant contribution to the field is the random coefficient logit model, or BLP 1995 (Berry et al. 1995). Due to the limitations of the product-level "taste shock" in logit models, a new model based on pure product characteristics has been proposed recently (Berry & Pakes 2007). The pure characteristic model (hereafter, PCM) differs from the BLP model in the sense that it does not contain the product-level "taste shock." It describes the consumer heterogeneity, purely based on their different tastes towards individual product characteristics, without considerations on the tastes of certain products as a whole (i.e., brand preference). However, the PCM model represents an ideal case. In reality, the product-level idiosyncratic "tastes" of different consumers do exist in many markets. As pointed out in Song (2010), whether or not one should introduce the product-level "taste shock" should depend on the context of the market. Our model is motivated directly by Song (2010). Keeping in mind the two levels of consumer heterogeneity introduced by (1) different travel categories (i.e., family trip, romance, or business trip) and (2) different hotel characteristics, we propose a random coefficient structural model to identify the latent weight distribution that consumers assign to each hotel characteristic. The outcome of our analysis enables us to compute the expected utility gain from each hotel and rank them accordingly on a travel search engine.

Recommender Systems

Our work is related to online recommender systems. By generating a novel ranking approach for hotels, we aim to improve the recommendation strategy for travel search engines and provide customers with the "best-value" hotels early on in the search process. In the marketing literature, several model-based recommendation systems have been proposed to predict preferences for recommended items (Ansari et al. 2000, Ying et al. 2005, Bodapati 2008). A more recent trend along this line is Adaptive Personalization Systems (Ansari and Mela 2003, Rust and Chung 2006, Chung et al. 2009).

The rest of the paper is organized as follows. The next section discusses the work related to the data preparation, including the methods used to identify important hotel characteristics, the steps undertaken to conduct the surveys on Amazon Mechanical Turk to elicit user opinions, and the text mining techniques used to parse user-generated reviews. Then, in the following two sections, we provide an overview of our econometric approach, and discuss empirical results, respectively. After that, we will discuss how to design a value-based ranking system to improve online travel search. Finally, we conclude with a summary of potential insights and future directions.

Data

Our dataset consisted of observations from 1479 hotels in the US. We collected data from various sources to conduct our study. We had 3 months of hotel transaction data from Travelocity.com from 2008/11 to 2009/1, which contained the average transaction price per room per night and the total number of rooms sold per transaction. Our work leveraged three types of user-generated content data:

- On-demand user-contributed opinions through Amazon Mechanical Turk
- Location description based on user-generated geo-tagging and image classification
- Service description based on user-generated product reviews

We first discuss how we leverage Amazon Mechanical Turk to collect information on user preferences for different hotel characteristics. Their responses suggest that these characteristics can be lumped into two groups: location and service characteristics. Once we identify the set of consumer preferences, we use other kinds of user-generated content to infer the external location characteristics, the internal service characteristics, and the textual characteristics of hotel reviews that can influence consumer purchases.

Extraction of Hotel Characteristics using Amazon Mechanical Turk (AMT)

Our analysis first requires knowing what aspects of a product are important for consumers, as these are the ones that may collectively determine consumers' choices. To do so, we performed a survey using Amazon Mechanical Turk (AMT) service to collect user preferences. AMT is an online marketplace, used for "crowd-sourcing" on-demand micro-tasks that require human intervention (i.e., cannot be fully automated using machine learning tools)[1]. For our survey, we conducted a small pilot study. We asked 100 anonymous AMT users for characteristics that would influence their choice of booking a hotel. Our analysis identified two broad categories of characteristics: Location-based characteristics (such as "near the beach") and Service-based characteristics (such as "internal amenities"). Since some important characteristics were extracted from online word-of-mouth (e.g., reviewer rating), to control for the quality and the perceived usefulness of word-of-mouth, we further investigated additional characteristics related to the online review and reviewer information. Next, we describe how we use user-generated content to collect information about the variables that are either too difficult to collect otherwise (e.g., density of shops around the hotel), or are likely to be very subjective (e.g., "quality of service").

Extraction of Location Characteristics using Social Geo-tagging

For the location-based characteristics, we combine user-generated content with automatic techniques, to be able to scale our data collection and generate data sets that were comprehensive at the national and even international level

---

[1] To ensure that the participants are a representative sample of the overall Internet population, we constructed a survey, asking AMT workers about their place of origin and residence, gender, age, education, income, marital status, etc. We found that the AMT population is generally representative of the overall US Internet population.

(i.e., tens or even hundreds of thousands of hotels). A first, automatic approach is to use a service like the Microsoft Virtual Earth Interactive SDK, which enables us to compute characteristics like "Near restaurants and shops" for a given location on a map. Using the automatic API from the Microsoft, we can automatically perform such local search queries. However, the presence of a characteristic like "Near a beach," or "Near downtown" cannot be retrieved by existing mapping services. To measure such characteristics, we use a combination of user-generated *geo-tagging* and automatic classification of satellite images of areas near each hotel in our dataset. In our study, we extracted the location characteristics "Near public transportation," "Near a beach" and "Near the downtown" via the site Geonames.org. For the characteristics "Near a lake/river" and "Near the interstate highway," we extracted them using social annotations from AMT. Such geo-tagging and social annotations enable us to generate a richer description of the location around each hotel, using features that are not available through existing mapping services.

However, no matter how comprehensive the tagging is, there can be locations that are not yet tagged by users. Therefore, we need ways to leverage the tag database, and allow for the automatic tagging of areas that lack tags. For this, we use automatic image classification techniques of satellite images to automatically tag location features that can influence hotel demand.

Extraction of Location Characteristics using Image Processing

Image Data Retrieving: Consider, for example, the case where we are trying to automatically identify whether a hotel is located "Near a beach," or "Near downtown." For this, we extracted hybrid satellite images (sized 256×256 pixels) using the Visual Earth Tile System[2], for each of the (thousands) of hotel venues located in the US, with four different zoom levels for each. These 4 x 1497 images were used to extract information about the surroundings of the hotel, through image classification and human inspection using AMT.

Image Classification: In order to automatically tag satellite images, we first needed to train our classification model. As a "training set," we used information on areas that have been either tagged by users of social tagging sites such as Geonames.org or annotated by users on AMT. We built the image classifiers as follows: First, we randomly selected a set of 121 hotels and requested five AMT users to label each example according to its corresponding satellite images from four different zoom levels. The labelers answered whether there is a beach in the image, or whether the area is a downtown area. We applied a simple majority voting method to make the final decision from the multi-labels of the example. Second, we trained an SVM classifier on this dataset and used the trained SVM classifier to classify the images that corresponded to the remaining hotels. It has been shown in prior work that non-parametric classifiers, such as Neural Networks, Decision Trees, and Support Vector Machines (SVM) provide better results than parametric classifiers in complex landscapes (Lu and Weng 2007). Therefore, we tested various non-parametric classification techniques. These include (i) Decision Trees, which are widely used for training and classification of remotely sensed image data (due to their ability to generate human interpretable decision rules and its speed in training and classification), and (ii) Support Vector Machines (SVM), that are highly accurate and perform well for a wide variety of classification tasks (Fukuda and Hirosawa 2001).

We conducted a small study to examine the performance of the classifier out-of-sample data. We classified the out of sample images using AMT; our results illustrated that our SVM classifier had an accuracy of 91.2% for the "beach" image classification and 80.7% for the "downtown" image classification. We also used the C4.5 algorithm for the classification, and the results were found consistent. The main reason for this is that "beach" images often contain a "sand strip," together with an "ocean margin" well distributed in density. This typically provides more stable and distinct textural information for the "beach" images, thus making them easier to distinguish.

Extraction of Service Characteristics using Consumer Reviews

There are two broad characteristics in the category of service-based characteristics: hotel class and number of internal amenities. "Hotel class" is an internationally accepted standard ranging from 1-5 stars, representing low to high hotel grades. "Number of internal amenities" is the aggregation of hotel internal amenities, such as "bed quality," "hotel staff," "food quality," "bathroom amenities" and "parking facility." We extracted this information from the Tripadvisor.com website using fully automated parsing. Since hotel amenities are not directly listed on the Tripadvisor.com website, we retrieved them by following the link provided on the hotel web page, which randomly directs the user to one of its cooperating partner websites (i.e., Travelocity.com, Orbitz.com, or Expedia.com).

---

[2] http://msdn2.microsoft.com/en-us/library/bb259689.aspx

Extraction of Textual Content of Reviews

We collected customer reviews from Travelocity.com. In order to consider the indirect influence of "word-of-mouth," we also collected reviews from a neutral, third party site - the Tripadvisor.com website, which is the world's largest online travel community. The online reviews and reviewers' information were collected on a daily basis up to January 31, 2009 (the last date of transactions in our database).

Table 1. Summary of Different Methods for Extracting Hotel Characteristics

Category	Hotel Characteristics		Methods
Transaction Data	Transaction Price (per room per night) Number of Rooms sold (per night)		*Travelocity*
Service-based	Hotel Class Hotel Amenities		*Tripadvisor*
Review-based	Number of Customer Reviews Overall Reviewer Rating Disclosure of Reviewer Identity Information		*Travelocity* and *Tripadvisor*
	Subjectivity	Mean Probability Std. Dev. Of Probability	Text Analysis
	Readability	Number of Characters Number of Syllables Number of Spelling Errors Average Length of Sentence SMOG Index	
	(Additional)	Breakfast Hotel Staff Bathroom Bed Parking	
Location-based	Near the Beach Near Downtown		Image Classification, Tags from *Geonames.org* and Social Annotations from *Amazon Mechanical Turk*
	External Amenities (Number of restaurants) Number of Local Competitors		*Microsoft Virtual Earth Geo-Mapping* Search SDK
	Near Public Transportation		Tags from *Geonames.org*
	Near the Interstate Highway Near the Lake/River		Social Annotations from *Amazon Mechanical Turk*
	City Annual Crime Rate		FBI online statistics

In addition to the total number of reviews and the numeric reviewer rating, we looked into two text style features: "subjectivity" and "readability." Both of them can influence consumers' purchase decisions (Ghose and Ipeirotis 2010). For higher precision, we used a multiple-item method for subjectivity and readability. We included 2 sub-features for subjectivity and 5 sub-features for readability, each of which measures the review text-style from an independent point of view. In order to decide the probability of subjectivity for review text, we trained a classifier using as "objective" documents the hotel descriptions of each of the hotels in our data set. We randomly retrieved 1000 reviews to construct the "subjective" examples in the training set. We conducted the training process by using a 4-gram Dynamic Language Model classifier provided by the LingPipe toolkit. Thus, we were able to acquire a subjectivity confidence score for each sentence in a review, thereby deriving the mean and standard deviation of this score, which represent the probability of the review being subjective. Finally, previous research suggested that the prevalence of reviewer disclosure of identity information is associated with changes in subsequent online product sales. Therefore, we decide to include one particular characteristic capturing the level of reviewers' disclosure of identity information on these websites – "real name or location."

Table 2. Definitions and Summary Statistics of Variables

Variable	Definition	Mean	Std. Dev.	Min	Max
PRICE	Transaction price per room per night	126.59	79.47	12	978
CHARACTERS	Average number of characters	766.54	167.13	121	2187
COMPLEXITY	Average sentence length	16.41	3.95	2	44.
SYLLABLES	Average number of syllables	245.48	53.77	37	700
SMOG	SMOG index	9.91	.63	3	19.80
SPELLERR	Average number of spelling errors	1.10	.37	0	3.33
SUB	Subjectivity - mean	.99	.03	.05	1
SUBDEV	Subjectivity - standard deviation	.02	.02	0	.25
ID	Disclosure of reviewer identity	.77	.14	0	1
CLASS	Hotel class	3.02	.93	1	5
COMPETITOR	Number of local competitors within 2 miles	1.77	2.80	0	20
CRIME	City annual crime rate	195.09	123.11	3	1310
AMENITYCNT	Total number of hotel amenities	16.38	3.21	2	23
EXTAMENITY	Number of external amenities within 1 mile, i.e., restaurants, shopping malls, or bars	4.95	7.37	0	27
BEACH	Beachfront within 0.6 miles	.24	.43	0	1
LAKE	Lake or river within 0.6 miles	.23	.42	0	1
TRANS	Public transportation within 0.6 miles	.11	.31	0	1
HIGHWAY	Highway exits within 0.6 miles	.68	.47	0	1
DOWNTOWN	Downtown area within 0.6 miles	.69	.46	0	1
TA_REVIEWCNT	Total number of reviews (Tripadvisor)	127.81	164.22	0	999
TA_RATING	Overall reviewer rating (Tripadvisor)	3.49	.59	1	5
TL_REVIEWCNT	Total number of reviews (Travelocity)	25.26	29.77	0	202
TL_RATING	Overall reviewer rating (Travelocity)	3.87	.74	1	5
Number of Observations: 8099		**Time Period: 11/1/2008-1/31/2009**			

In sum, there are 5 broad types of characteristics in this category: (i) total number of reviews, (ii) overall review rating, (iii) review subjectivity (mean and variance), (iv) review readability (the number of characters, syllables, and spelling errors, complexity and SMOG Index), and (v) the disclosure identity information by the reviewer.

For a better understanding of the variables in our setting, we present the data sources, definitions, and summary statistics of all variables in Tables 1 and 2.

Random Coefficients-Based Structural Model

In this section, we will discuss how we develop our random coefficient structural model and describe how we apply it to empirically estimate the distribution of consumer preferences towards different hotel characteristics.

Our model is motivated directly by the model in Song (2010), where the author proposed a hybrid discrete choice model of differentiated product demand. While Song (2010) had one random coefficient on price, we have multiple random coefficients on prices as well as hotel characteristics. Note that this hybrid model is a combination of the BLP (1995) and the PCM (2007) approaches. It is called a hybrid model because it resembles the random coefficient logit demand model in describing a brand choice (BLP 1995) and the pure characteristics demand model in describing a within-brand product choice (PCM 2007). This is basically a discrete choice model of differentiated product demand in which product groups are horizontally differentiated while products within a given group are vertically differentiated conditional on product characteristics. These two types of differentiation are distinguished by a group-level "taste shock," which is assumed to be distributed i.i.d. with a Type I extreme-value distribution.

This taste shock represents each consumer's specific preference towards a product group that is not captured by observed or unobserved product characteristics. Song (2010) refers to a product group that contains vertically differentiated products a "brand." This hybrid model identifies preference for product characteristics in a similar way as the PCM. The main difference that the hybrid model compares products of each brand on the quality ladder separately, while the PCM compares all products on it at the same time. Hence, the quality space is much less crowded in the hybrid model.

In our context, a hotel "travel category" represents a "brand" and the hotels within each "travel category" represent "products." In particular, the market share function of hotel j^k within travel category k can be written as the product of the probability that travel category k is chosen and the probability that hotel j^k is chosen given that travel category k is chosen. The former probability is similar to the choice probability in BLP, and the latter to that of the PCM..

We define a consumer's decision-making behavior as follows. A consumer needs to locate the hotel whose location and service characteristics best matches her travel purpose. For instance, if a consumer wants to go on a romantic trip with a partner, she would be interested in the set of hotels that are located close to a beach, downtown with amenities like nightclubs, restaurants, etc. She is also aware that hotels specializing in the "romance" category are more likely to satisfy such location and service needs. Each hotel can belong to one of the following eight types of ``travel categories:'' *Family Trip, Business Trip, Romantic Trip, Tourist Trip, Trip with Kids, Trip with Seniors, Pet Friendly, and Disability Friendly*.[3] To capture the heterogeneity in consumers' travel purpose, we introduce an idiosyncratic "taste shock" at the travel category level. This is similar to the product-level "taste shock" in the BLP (1995) model.

Each travel category has a hotel that maximizes a consumer's utility in that category. We refer to this as the "best" hotel in that category. To find the "best" hotel within each travel category, we use the pure characteristic model (PCM) proposed by Berry and Pakes (2007). The PCM approach is able to capture the vertical differentiation amongst hotels within the same travel category. A rational consumer chooses a travel category if and only if her utility from the best hotel in that category exceeds her utility from the best hotel in any other travel category. Thus, in our model, the utility for consumer i from choosing hotel j with category type k in market t can be represented as illustrated in Equation (1):

$$u_{ij^k_t} = X_{j^k_t}\beta_i - \alpha_i P_{j^k_t} + \xi_{j^k_t} + \varepsilon_{it}^k, \tag{1}$$

Where: i represents a consumer, j^k represents hotel j with travel category type k ($1 \le k \le 7$), and t represents a hotel market. In this model, β_i and α_i are random coefficients that capture consumers' heterogeneous tastes towards different observed hotel characteristics, X, and towards the average price per night, P, **respectively**. ξ represents hotel characteristics unobservable to the econometrician. ε_{it}^k with a superscript k represents a travel category-level "taste shock." Note that in our model the travel category- level shock is independently and identically distributed across consumers and travel categories, consistent with Song (2010).

We define a "market" as the combination of "city-night." Correspondingly, the market share for each hotel is calculated based on the revenues for that hotel in that city on that night divided by the total revenue from all hotels in that city on that night. We also tried the combination of "city-week." We also tried alternate definitions of market size. For example, we applied similar ideas as in Song (2007), by increasing or decreasing the total revenue for each market by 20%. We found that in our data different market sizes yield qualitatively the same results.

Our main dataset comes from two major sources: Travelocity-generated transaction data and TripAdvisor review data. The dataset we used in our analysis is the set of hotels at the intersection of the two sources. This means that the hotel choice set for each market includes those hotels that not only have a transaction generated via Travelocity, but also have available information on user-generated reviews on TripAdvisor. Since not every hotel that has a

---

[3]Each travel category is defined and chosen according to the information gleaned from the website of TripAdvisor. TripAdvisor allows reviewers to specify their main trip purpose (travel category) while posting a review. We have data on all the hotel reviews posted by users for a given hotel right from the time the first review was posted till the last date of our transaction dataset (February 2009). A hotel is classified into a specific travel category based on the most frequently mentioned travel purpose by the reviewers for that hotel. Hence, each hotel belongs exclusively to a travel category.

Travelocity-generated transaction is listed on the TripAdvisor website, we define our "outside good" as the set of hotels that are listed in the original Travelocity transaction data, but not listed on the TripAdvisor website.

Due to the computational complexity and data restriction, estimating a unique set of weights for each consumer is intractable. To make this model tractable, we made some further assumptions about β_i and α_i One is to assume that these weights are normally distributed among consumers, i.e., $\beta_i \sim (\beta_i \mid \overline{\beta}, \sigma_{\beta i})$ and $\alpha_i \sim (\alpha_i \mid \overline{\alpha}, \sigma_{\alpha i})$. Our goal is then to estimate the means $(\overline{\beta}, \overline{\alpha})$ and the standard deviations $(\sigma_{\beta i}, \sigma_{\alpha i})$ of these two distributions. The means correspond to the set of coefficients on hotel characteristics and on hotel price, which measures the average weight placed by the consumers. The standard deviations provide a measure of the extent of consumer heterogeneity in those weights.

Furthermore, we notice that these heterogeneities result from particular demographic attributes of consumers. For example, the variance in the price coefficient is very likely a result of differences in incomes among the consumers. Therefore, we make additional assumptions about the standard deviations: $\sigma_{\alpha i} \sim \alpha_I I_i$, where I_i represents the income whose distribution can be learned from the consumer demographics; $\sigma_{\beta i} \sim \beta_v v_i$, where $v_i \sim N(0,1)$ represents some random factor that will influence people's preferences towards individual hotel characteristics.

Therefore, we have the following two forms for the consumer-specific coefficients α_i and β_i:

$$\alpha_i = \overline{\alpha} + \alpha_I I_i \qquad \text{and} \qquad \beta_i = \overline{\beta} + \beta_v v_i.$$

We then rewrite our model as follows:

$$u_{ij^k{}_t} = \delta_{j^k{}_t} + X_{j^k{}_t} \beta_v v_i - \alpha_I I_i P_{j^k{}_t} + \varepsilon_{it}^k, \tag{2}$$

Where: $\delta_{j^k{}_t} = X_{j^k{}_t} \overline{\beta} - \overline{\alpha} P_{j^k{}_t} + \xi_{j^k{}_t}$, represents the mean utility of hotel j with category type k in market t. β_v and α_I are the parameters to be estimated.

Estimation

Our goal here is to estimate the mean and variance of β_i and α_i. We apply estimation methods similar to those used in Berry & Pakes (2007) and Song (2010). This problem can be essentially reduced to a procedure of solving a system of nonlinear equations. In general, with a given starting value of $\theta_0 = (\alpha_I^0, \beta_v^0)$, we look for the mean utility δ such that the model predicted market share equates the observed market share. From there, we form a GMM objective function using the moment conditions such that the mean of unobserved characteristics is uncorrelated with instrumental variables. Based on this, we identify a new value of $\theta_1 = (\alpha_I^1, \beta_v^1)$, which is used as the starting point for the next round iteration. This procedure is repeated until the algorithm finds the optimal value of θ that minimizes the GMM objective function. To find a solution, we applied the Newton-Raphson method suggested by Song (2010), where this method was shown to work well when the number of products per market is up to 20. To guarantee the robustness of the results when the number of products is larger than 20, we tried different initial values in the iteration. The final solution was consistent across different initial values. In practice, this approach locates the closest solution for our settings, while the iteration procedure provides a very close form to locate the roots rapidly and stably.

Empirical Analysis and Results

Note that a consumer who is searching for hotel reviews on Travelocity or Tripadvisor gets to see a different number of reviews on the pages of each website. While Travelocity.com displays the first five reviews on a page, Tripadvisor.com lists 10 reviews per page. To minimize the potential bias caused by webpage design, since some customers may only read the reviews on the first page, we decided to consider two more alternatives besides our main dataset: Dataset (II) with hotels that have at least five reviews, and Dataset (III) with hotels that have at least 10 reviews. Meanwhile, in order to control for other unobserved factors, we included 9 major hotel brand dummies in the test: Accor, Best western, Cendant, Choice, Hilton, Hyatt, Intercontinental, Marriott, and Starwood. The estimation results from these three datasets are illustrated in Table 3 columns 2-4.

Moreover, we also estimated the model with additional textual features (Dataset (ADD)). We further exploited the information about hotel service characteristics from the natural language text of the consumer reviews. To do so, we built on the work of Hu and Liu (2004), Popescu and Etzioni (2005), Archak et al. (2008). We first identified the top 5 frequently mentioned hotel service features from the review textual body: *hotel staff, food quality, bathroom, parking facilities,* and *bed quality*. Next, we extracted all the evaluation phrases (adjectives and adverbs) that are being used to evaluate the individual service features (for example, for the feature "hotel staff" we extract phrases like "helpful", "smiling", "rude", "responsive", etc) . Then we used the similar methodology of Archak et al. (2008) to create the "external" scores for the corresponding evaluation-product phrase pair using AMT. The estimation results are illustrated in Table 3 columns 5.

Estimation Results

We first start with the location-based characteristics. There are five location-based characteristics, which have a positive impact on hotel demand: "External Amenities," "Beach," "Public Transportation," "Highway," and "Downtown." These characteristics strongly imply that the location and geographical convenience for a hotel can make a big difference in attracting consumers. Hotels providing easy access to public transportation (such as a subway or bus stations), highway exits, restaurants and shops, or easy access to a downtown area, can have a much higher demand. "Beach," also showed a positive impact on demand. It turns out that most beach-based hotels in our dataset were located in the south where the weather typically stays warm even in winter. Therefore, the desirability of a "walkable" beachfront did not reduce even in the winter season (which is the time of our data).

Two location-based characteristics have a negative impact on hotel demand: "Annual Crime Rate" and "Lake/River." The higher the average crime rate reported in a local area, the lower the desirability of consumers for staying in a hotel located in that area. This indicates that neighborhood safety plays an important role in the hotel industry. The second location-based characteristic that illustrates a negative impact is the presence of a water body like a lake or a river. This is interesting because one would expect people to choose a hotel near a lake or by a riverside. However, most waterfront-based hotels in our dataset were located in places where the weather becomes extremely cold in the winter months of November to January. Due to the low temperatures, it is likely that a lake or riverfront becomes less desirable for travelers. [4]

To further examine the impact of lake, we collected weather data from the National Oceanic and Atmospheric Administration (NOAA) on the average temperature from 2008/11 to 2009/1 for all cities. Then, we defined 2 dummy variables: "High Temp" which equals to 1 if the average temperature is higher than 50 degree, and "Low Temp" which equals to 1 if the average temperature is lower than 40 degree.[5] We interacted "High Temp" and "Low Temp" separately with "Lake" in our model. The results showed that the interaction of "Low Temp" with "Lake" has a significantly negative effect. This supports our earlier argument. While not statistically significant (p value = 0.2), the interaction of "High Temp" with "Lake" showed a positive effect, weakly suggesting that warmer weather may help the lake area to attract more visitors. As a robustness check, we did the similar analysis for "Beach" conditional on high and low temperatures. The results showed similar trend. Moreover, the coefficients are both statistically significant in the case of "Beach."

For Service-based characteristics, we notice that both "Class" and "Amenity Count" has a positive impact on hotel demand. Hotels with a higher number of amenities and higher star-levels have higher demand, controlling for price. "Reviewer Rating" also has a positive association with hotel demand. With regard to the "Number of Reviews," we find a positive sign for its linear form while a negative sign for its quadratic form. This indicates that the economic impact from the customer reviews is increasing in the volume of reviews but at a decreasing rate, as one would expect.

The textual quality and style of reviews demonstrated a statistically significant association with demand. All the readability and subjectivity characteristics had a statistically significant association with hotel demand. Among all the readability sub-features, "Complexity," "Syllables" and "Spelling Errors" had a negative sign and, therefore, have a negative association with hotel room demand. This implies that reviews higher readability characteristics

---

[4]In addition, some traveler reviews commented on the presence of mosquitoes in areas near a lake.

[5]We tried other combinations to classify High vs. Low temperatures (>=70 degrees as High and <=30 degree as Low (ii) >=60 degrees as High and <=20 degrees as Low) but they all yielded qualitatively similar results.

(short sentences and less complex words), and reviews with fewer spelling errors have a positive association with demand. On the other hand, the sign of the coefficients on "Characters" and "SMOG index" is positive, implying that longer reviews that are easier to read have a positive association with demand. [6] These results indicate that consumers can form an image about the quality of a hotel by judging the quality of the (user-generated) reviews.

For the subjectivity sub-features, both "Mean Subjectivity" and "Subjectivity Standard Deviation" turned out to have a negative association with demand. This implies that consumers prefer reviews that contain objective information (such as factual descriptions of rooms) relative to subjective information. With respect to the "Subjectivity Standard Deviation," our findings suggest that people prefer a "consistent objective style" from online customer reviews compared to a mix of objective and subjective sentences. The last review-based characteristic was "Disclosure of Reviewer Identity." This variable demonstrated a positive association with hotel demand. This result is consistent with previous work (Forman et al. 2008), which suggested that the identity information about reviewers in the online travel community can shape positively community members' judgment towards hotels. "Price" has a negative sign, which is as expected.

The results with additional textual variables stay consistent with our main results. In addition, we found that the three features that have a positive and statistically significant impact on demand are "Food Quality," "Hotel Staff" and "Parking Facilities." In contrast, "Bedroom Quality" had a negative impact on demand. While this negative sign is surprising, this can happen if consumers use bedroom quality as a cue for price, especially given that quality in our data is a proxy for the number of beds and size of the room (full, queen, king, etc). This is possible because sometimes prices are obfuscated on the main results page and are only available just before checkout. However, this is only one possible explanation.

Besides the above qualitative implications, we also quantitatively assess the economic value of different hotel characteristics. More specifically, we examined the magnitude of marginal effects on hotel demand for the location-, service- and review-based hotel characteristics. The presence of a beach near the hotel increases demand by 14.59% on an average. In contrast, a location near a lake or river decreases demand by 10.47%. Meanwhile, easy access to transportations and highway exits increase demand by 18.44% and 10.23%, separately. Presence of a hotel near downtown increases demand by 8.35%. With regard to service-based characteristics, a 1-star improvement in hotel class leads to an increase in demand by 5.30% on average. Moreover, the presence of one more internal or external amenity increases demand by 0.06% or 0.07%, respectively. Demand decreases by 0.33% if the local crime rate increases by one unit.

With regard to the review-based characteristics, we found that the SMOG index (which represents the readability of the review text), has the highest marginal influence on demand on an average. A one level increase in SMOG index of reviews is associated with an increase in hotel demand by 7.42% on an average. A one unit increase in the number of characters is associated with an increase in hotel demand by 0.11%, whereas a one unit increase in the number of spelling errors, syllables or in complexity is associated with a decrease in hotel demand by 1.36%, 0.27%, and 0.59%, respectively. In terms of review subjectivity, a 10 percent increase in the average subjectivity level is associated with a decrease in hotel demand by 2.62%; a 10 percent increase in the standard deviation of subjectivity will reduce demand by 3.33%. Finally, a 10 percent increase in the reviewer identity-disclosure levels is associated with an increase in hotel demand by 1.42%.

Note that the estimation results from the three datasets are highly consistent. In general, all the coefficients illustrate a statistical significance with a p-value equal to or below the 5% level across all three datasets. Moreover, a large majority of variables present a high significance with a p-value below the 0.1% level.

Robustness Checks

To assess the robustness of our estimation model and results, we conducted three robustness tests:

<u>Robustness Test I: Use the same model based on alternative sample splits.</u> We considered three alternative datasets: Dataset (IV) containing hotels with at least one review from Tripadvisor.com, Dataset (V) containing hotels with at least one review from Travelocity.com, and Dataset (VI) containing hotels with at least one review from both. We

---

[6]To alleviate any possible concerns with multi-collinearity between *SMOG* and *Syllables*, we re-estimate our model after excluding the SMOG index variable. There was no change in the qualitative nature of the results across the different datasets.

found that the coefficients from the estimations are qualitatively very similar to our main results. Moreover, similar to those in the main results, most variables in the robustness tests also illustrate statistical significance at or below the 5% level.

Table 3.   Main Estimation Results

Variable	Coef. (Std. Err)[I]	Coef. (Std. Err)[II]	Coef. (Std. Err)[III]	Coef. (Std. Err)[ADD]
Price[(L)]	-.131*** (.025)	-.130*** (.025)	-.127*** (.026)	-.133***(.025)
CHARACTERS[(L)]	.009*** (.002)	.008*** (.002)	.009*** (.002)	.008***(.002)
COMPLEXITY	-.005* (.002)	-.006* (.003)	-.007* (.003)	-.005* (.002)
SYLLABLES[(L)]	-.024*** (.006)	-.025*** (.006)	-.032*** (.006)	-.026***(.006)
SMOG	.063* (.024)	.060* (.028)	.058* (.026)	.063* (.026)
SPELLERR[(L)]	-.121*** (.039)	-.120*** (.036)	-.128*** (.039)	-.124*** (.037)
SUB	-.223*** (.071)	-.219** (.072)	-.212** (.072)	-.225***(.071)
SUBDEV	-.283*** (.078)	-.282*** (.079)	-.280*** (.079)	-.261***(.079)
ID	.121*** (.026)	.128*** (.026)	.130*** (.026)	.108*** (.025)
CLASS	.045*** (.009)	.042*** (.009)	.039*** (.008)	.038*** (.010)
CRIME	-.028* (.014)	-.026* (.014)	-.030* (.015)	-.035* (.014)
AMENITYCNT[(L)]	.005* (.002)	.004* (.002)	.004* (.002)	—
EXTAMENITY[(L)]	.006*** (.002)	.005** (.002)	.007*** (.002)	.006*** (.002)
BEACH	.124*** (.031)	.123*** (.031)	.125*** (.031)	.134*** (.031)
LAKE	-.089*** (.026)	-.091*** (.026)	-.087*** (.026)	-.076** (.026)
TRANS	.159*** (.035)	.159*** (.036)	.164*** (.036)	.167*** (.035)
HIGHWAY	.087*** (.024)	.088*** (.025)	.091*** (.025)	.087*** (.025)
DOWNTOWN	.071** (.025)	.079** (.026)	.075** (.026)	.072** (.025)
TA_RATING	.010* (.004)	.011* (.004)	.014* (.006)	.015† (.008)
TL_RATING	.017* (.006)	.018* (.006)	.018* (.006)	.017† (.008)
TA_REVIEWCNT[(L)]	.142** (.049)	.146** (.047)	.142** (.044)	.137***(.014)
TA_REVIEWCNT^2[(L)]	-.034*** (.006)	-.035*** (.008)	-.032*** (.006)	-.034***(.006)
TL_REVIEWCNT[(L)]	.010** (.003)	.012** (.004)	.015*** (.003)	.015***(.003)
TL_REVIEWCNT^2[(L)]	-.013** (.005)	-.012** (.005)	-.014** (.005)	-.012** (.005)
Constant	.217 (.254)	.232 (.246)	.211 (.249)	. 264 (.238)
Brand Control	Yes	Yes	Yes	Yes
BREAKFAST	—	—	—	.139**(.049)
STAFF	—	—	—	.078* (.049)
BATHROOM	—	—	—	.016 (.042)
BED	—	—	—	-.026† (.018)
PARKING	—	—	—	.021* (.009)

***	Significant at 0.1% level.	**	Significant at 1% level.
*	Significant at 5% level.	†	Significant at 10% level.
I	Main dataset (at least 1 review from either TA or TL).		
II	Main dataset with reviews >=5.	III	Main dataset with reviews >=10.
ADD	Main dataset with additional textual features.	(L)	Logarithm of the variable.

Robustness Test II: Use an alternative model based on the same datasets. To examine the robustness of the results from our hybrid model, we conducted another group of tests using an alternative model that has been widely used in the industrial organization and marketing literature, the random coefficient logit model, or BLP (Berry et al. 1995). We find that the estimation results are directionally very consistent with our main estimation results using the hybrid model. Specifically, the coefficients from the BLP estimation demonstrate three trends: (i) they have the same signs compared to our main results, which means that the economic effects are consistent in direction, (ii) they exhibit lower levels of statistical significance, compared to our main results, (iii) the scale of BLP coefficients is generally higher compared to our main results. More importantly, these three trends are consistent with the findings in Song (2010). This suggests that our analysis, based on the hybrid structural model, is not only consistent, but also efficient. In the next subsection, our model validation results further confirm this finding.

Robustness Test III: Use the same model based on additional variables. Since "Airport" and "Convention center" are often considered important especially in business travel, we decided to include these two location characteristics for testing as well. Moreover, to understand consumers' personal experience at a more precise level, we further considered 7 individual service ratings from TripAdvisor, aside from the overall customer review rating. These 7 individual ratings are "Value", "Room", "Location", "Cleanliness", "Service", "Check-in" and "Business Service". We found that the results remain qualitatively the same with our main results.

Furthermore, besides using the average price of the "same-star rating" hotels in the other markets as an instrument for price (e.g., Hausman et al. 1994), we have tried four other sets of instruments. First we follow Villas-Boas and Winer (1999) and use lagged prices as instruments in conjunction with Google Trends data. The lagged price may not be an ideal instrument since it is possible to have common demand shocks that are correlated over time. Nevertheless, common demand shocks that are correlated through time are essentially trends. Our control for trends using search volume data thus should alleviate most, if not all, such concerns. Second, we have used employee wage data from BLS as a "cost side" instrument using the category of "Accommodation." The assumption here, like in other papers that use such cost-side instruments is that hotel employee wages are correlated with hotel room prices but uncorrelated with factors that are reflected in the unobserved characteristics term (see for example, Chintagunta et al. 2005). Third, we have also tried region dummies as proxies for the cost (e.g., the cost of transportation, labor, etc.) as suggested by Nevo (2001). Fourth, we have used BLP-style instruments. Specifically, we have used the average characteristics of the same-star rating hotel in the other markets. All these alternate estimations yielded very similar results. We did an F-test in the first stage for each of the four sets of instruments. In each case, the F-test value was well over 10, suggesting that our instruments are valid (i.e., the instruments are not weak). In addition, the Hansen's J-Test could not reject the null hypothesis of valid overidentifying restrictions.

Ranking

After we estimate the parameters, we propose to design a new ranking tool for differentiated products based on the average consumer surplus from transactions in that product. As discussed in the model section, to capture the consumer heterogeneity, we represent the utility from a product for a consumer as consisting of two parts: the mean and the standard deviation. The mean utility provides a good estimation of how much consumers can benefit from choosing this particular product, while the standard deviation of utility describes the variance of this benefit.

In the case of hotels, we are interested in knowing what the excess utility or consumer surplus is for consumers on an aggregate level from choosing a certain hotel. Therefore, we define the consumer surplus from hotel j with travel category type k as the sum of its mean utility $\overline{\mu}_{ij^k t}$ divided by the mean price elasticity $\overline{\alpha}$ over all markets:

$$CS_{j^k} = \sum_t \frac{1}{\overline{\alpha}} \overline{\mu}_{ij^k t} \qquad (3)$$

We thereby propose a new ranking approach for hotels based on the consumer surplus of each hotel for consumers on an aggregate level. This ranking idea is based on how much "extra value" consumers can obtain after paying for that hotel. If a hotel provides a comparably higher surplus for consumers on an aggregate level, then it would appear on the top of our ranking list.

User Study

To evaluate the quality of our ranking technique, we conducted a user study using AMT. We generated different rankings for the top-10 hotels in accordance with 5 current existing baseline criteria: *Price low to high*, *Price high to*

low, *Hotel class*, *Hotel size (number of rooms)*, and *Number of internal amenities*. We also considered 4 other benchmark criteria based on user-generated content: *Customer rating from Tripadvisor.com*, *Customer rating from Travelocity.com, Maximum online review count* and *Most booked*.

We computed the consumer surplus for each hotel from our parameter estimates, and ranked the hotels in each city according to their surplus. Then, we performed pair-wise blind tests by asking 200 anonymous AMT users to compare pairs of rankings and tell us which of the hotel ranking lists they preferred. We tested the results for a few large cities, like New York City, and the results were encouraging. A large majority of customers preferred our ranking, when listed side-by-side with the other competing baseline techniques (p = 0.05, sign test). We also conducted the same comparisons for other cities of different sizes, such as Los Angeles, San Francisco, Orlando, New Orleans and Salt Lake City. The results showed very similar trends. Table 4 shows how often users preferred our own ranking scheme when presented side-by-side with an alternative.

Table 4. Ranking User Study Results

Baselines / Cities	Rating Tripadvisor	Rating Travelocity	Most Booked	Price Low to high	Price High to Low	Hotel Class	# of Reviews	# of Rooms	# of Amenities
New York	77%	63%	61%	57%	71%	88%	76%	89%	60%
Los Angeles	72%	58%	71%	59%	84%	89%	87%	86%	69%
San Francisco	79%	57%	65%	62%	70%	82%	68%	79%	79%
Orlando	83%	81%	62%	63%	73%	79%	73%	79%	61%
New Orleans	61%	69%	60%	78%	69%	80%	72%	91%	58%
Salt Lake City	61%	80%	69%	66%	79%	83%	73%	70%	76%
Significance Level			P=0.05 $\geq$ 62%	P=0.01 $\geq$ 66%	P=0.001 $\geq$ 72%	(Sign Test, N=200)			
[*]Mixed Rating Strategy: (i) Average of Tripadvisor rating and Travelocity rating when both are available; (ii) Equal to one of the two ratings if the other one is missing; (iii) Zero when both ratings are missing.									

As part of this study, we also asked consumers why they chose a particular ranking. This was done to better understand how users interpret the consumer surplus-based ranking. The majority of users indicated that our consumer surplus-based ranking promoted the idea that price was not the main factor in rating the quality of hotels. Instead, a good ranking recommendation is one that could satisfy customers' multidimensional preferences for hotels. Moreover, users strongly preferred the diversity of the retrieved results, given that the list consisted of a mix of hotels cutting across several prices and quality ranges. In contrast, the other ranking approaches tended to list hotel of only one type (e.g., very expensive ones).

Based on the qualitative opinions of these users, it appears that *diversity* in product choices is indeed an important factor that improves user satisfaction from different ranking systems. An economic approach to ranking introduces diversity naturally. This result seems intuitive: if a specific segment of the market systematically appeared to be underpriced, then market forces would modify the prices for the whole segment accordingly. Thus, these results dovetail well with our empirical estimations, which suggests that our consumer surplus-based ranking model can capture consumers' true purchase motivations.

Conclusions

In this paper, we propose a new way to present and rank results for travel search engines to facilitate efficient decision-making in electronic marketplaces, with a focus on travel search engines. Specifically, we propose a "consumer utility maximization" framework that takes into account both cognitive and social factors affecting human information seeking behavior. We show how to enhance existing economic models using modern text mining and image classification algorithms from computer science. Using a unique dataset from Travelocity and Tripadvisor, we empirically estimate the economic value of different characteristics for hotels using a hybrid random coefficient structural model, taking into consideration the two-level consumer heterogeneities introduced by the different travel contexts and different hotel characteristics. Our research enables us to not only quantify the economic impact of product characteristics, but also, by reversing the logic of this analysis, enables us to identify the characteristics that most influenced the demand for a particular product. After inferring the economic significance of

each characteristic, we then incorporate them into a new ranking system based on the derived consumer surplus. By doing so, we can provide customers with the "best-value" products early on in their search process. We conducted blind tests using real users, recruited through AMT to examine how well our ranking system performs in comparison with existing alternatives. We found that our ranking performs significantly better than several baseline-ranking systems that are being currently used.

By examining travel search through the "economic lens" of consumer surplus, we leverage and integrate theories of relevance from information retrieval and micro-economic theory. Our inter-disciplinary approach can improve the quality of information displayed by any travel search engine on the Internet. It has the potential to enhance the efficiency of human-computer interaction and the quality of user search experience in an electronic marketplace.

On a broader note, the objective of this paper was to illustrate how multiple and diverse online data sources can be mined and incorporated into electronic marketplace design, to enhance human information seeking experience in the context of economic decision-making. Besides providing direct economic gains, our proposed system can lead to non-trivial reduction in user search and cognitive costs for accessing and processing the product information. This can encourage effective interactions between the search engines and the end-users. Meanwhile, given the ability to manage and analyze a diversity of information sources, the proposed interface can be easily customized and integrated into any other electronic platforms with multiple and flexible data formats, to improve the real-time and ubiquitous information accessibility. For example, such applications include a search interface for mobile content downloading for images or music; or a GPS automobile system that searches for the best route between two locations regarding the driving time, distance, distribution of gas stations and fast food restaurants. One advantage from our research is that, based on the various types of social media content available for public access both locally and remotely, individual decision-making process can effectively interact with the "wisdom-of-crowd" in many electronic environments.

Our research can also have implications for website design. Human attention has a limited processing capacity that can be allocated in varying amounts to different locations in a visual field (Pashler 1998). Our approach takes a comprehensive perspective and explores how humans process information and make decisions when exposed to multiple ranked lists of choices on the same computer screen. Our ranking interface design optimizes the allocation of user attention on a webpage based on the potential economic satisfaction for the user gained from each item on that webpage. Such a design serves not only as a mechanism for information representation, but also as a stimulus for users to discover "best fit" items, thereby eliciting efficient information transmission and decision-making.

Furthermore, our research has the potential to support personalization and adaptive interfaces in any business decision-making systems. Specifically, by incorporating more individual level user demographics (such as income, age, gender, education level, etc.) and personal context information (such as motivation, purpose or attitude towards the decision) at the time of information seeking during travel search, one can extend our techniques to infer the personalized utility, and design a personalized search interface for each user correspondingly. Such personalized knowledge interfaces can induce a positive impact on users' perception and learning of information. With richer individual level information, we are able to conduct traditional collaborative filtering or content-based algorithms, hence comparing our personalization results with theirs.

Our work has several limitations some of which can serve as fruitful areas for future research. One can further break down the textual content of user-generated reviews in order to extract multiple service related dimensions of every single product and examine the economic impact of each dimension. This can be done by conducting auto topic extraction techniques from text mining, combining with sentimental analysis to evaluate the subjectivity level of each interesting topic. This will enable us to better recover customers' multi-dimensional heterogeneous tastes towards different product characteristics. Meanwhile, in order to better understand the antecedents of consumer's decisions, future work can look not only at transaction data but also into their browsing history and learning behavior. This will help us to explore the dynamics of user search experience and create interfaces with more interactivity.

References

Agarwal, R. and Venkatesh, V. 2002. "Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability," *Information Systems Research*, (13) 2, pp. 168-186.

Agichtein, E., Lawrence, S., and Gravano, L., 2001. "Learning search engine specific query transformations for question answering," in *Proceedings of the 10th international conference on World Wide Web*, pp.169-178.

Ansari, A., and Mela, C.F. 2003. "E-customization," *Journal of Marketing Research*, 40(2) 131--145.

Ansari, A., Essegaier, S. and Kohli, R. 2000. "Internet Recommendation Systems," *Journal of Marketing Research*, v37, pp. 363-375.

Archak, N., Ghose, A. and Ipeirotis, P.G. 2008. "Deriving The Pricing Power Of Product Features By Mining Consumer Reviews," NYU CeDER Working paper.

Arnaud, D. B., Liechty, J., Huizingh, E. and Lilien, G. 2008. "Offering Online Recommendations with Minimum Customer Input Through Conjoint-Based Decision Aids," *Marketing Science*, 27(3), pp. 443-460.

Athey, S. and Imbens, G. 2007. "Discrete Choice Models with Multiple Unobserved Choice Characteristics," *International Economic Review*, 48(4), pp. 1159-1192.

Berndt, E. 1996. *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley Publishing Company, Reading, MA.

Berry, S. 1994. "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics* (25). pp. 242-262.

Berry, S., Levinsohn, J. and Pakes, A. 1995. "Automobile Prices in Market Equilibrium," *Econometrica* (63). pp. 841-890.

Berry, S. and Waldfogel, J. 2001. "Do Mergers Increase Product Variety? Evidence from Radio Broadcasting," *Quarterly Journal of Economics* (116). pp. 969-1007.

Berry, S., Linton, O. and Pakes, A. 2004. "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *Review of Economic Studies* (71). Pp. 613-654.

Berry, S. and Pakes, A. 2007. "The Pure Characteristics Demand Model," *International Economic Review* (48). pp. 1193-1225.

Bodapati, A. 2008. "Recommender Systems With Purchase Data," *Journal of Marketing Research*, 45 (1), 77--93.

Bowen, J. T. and Sparks, B.A. 1998. "Hospitality Marketing Research: A Content Analysis And Implications For Future Research," *International Journal of Hospitality Management* (17:2), pp. 125-144.

Bull, A. 1998. "The Effects of Location and Other Attributes on the Price of Products Which are Place-sensitive in Demand," phd thesis, Griffith University, 1998.

Case, Donald. 2007. *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Second Edition. Academic Press.

Chevalier, J. and Mayzlin, D. 2006. "The Effect Of Word Of Mouth On Sales: Online Book Reviews," *Journal of Marketing Research*. 43(3) 345-354.

Chintagunta 2001. "Research Note Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data," *Marketing Science*, 20(4), 442-456.

Chintagunta, P., Dubé, J.P. and Goh, K.Y. 2005. "Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Models," *Management Science* Vol. 51, No. 5, pp. 832-84.

Chung, T.S., Rust, R. and Wedel, M. 2009. "My Mobile Music: An Adaptive Personalization System For Digital Audio Players," *Marketing Science*, 28 (1), 52--68.

Das, S. and Chen, M. 2007. "Yahoo! For Amazon: Sentiment extraction from small talk on the web," *Management Science* 53(9), pp. 1375–1388.

Dellarocas, C., Awad, N. and Zhang, M. 2007. "Exploring The Value Of Online Product Reviews In Forecasting Sales: The Case Of Motion Pictures," *Journal of Interactive Marketing* 21(4), pp. 23-45.

Duan, W., Gu, B. and Whinston, A. B. 2008. "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Decision Support Systems* 45(4), pp. 1007-1016.

Dube, J.P., Fox, J. and Su, L. 2009. "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," Working Paper, University of Chicago.

Eliashberg, J., Hui, S. K. and Zhang, Z. J. 2007. "From Story Line To Box Office: A New Approach For Green-Lighting Movie Scripts," *Management Science*. 53(6), pp. 881–893.

Fader, P. S. and Hardie, B.G.S. 1996. "Modeling Consumer Choice Among SKUs," *Journal of Marketing Research*, 33 (November), pp. 442-452.

Fellbaum, C. 1998. *Wordnet: An Electronic Lexical Database*. MIT Press.

Ford, N., Wilson, T.D., Foster, A., Ellis, D. and Spink, A. 2002. "Information seeking and mediated searching. Part 4. Cognitive styles in information seeking," *Journal of the American Society for Information Science and Technology*, 53(9), pp. 728–735.

Forman, C., Ghose, A. and Wisenfeld, B. 2008. "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, 19(3), pp. 291-313.

Franklin, S. and Wulder, M. 2002. "Remote Sensing Methods In Medium Spatial Resolution Satellite Data Land Cover Classification Of Large Areas," *Progress in Physical Geography*, (26), pp. 173-205.

Fukuda, S. and Hirosawa, H. 2001. "Support Vector Machine Classification Of Land Cover: Application To Polarimetric SAR Data," in *Proceedings of the IEEE International Geoscience & Remote Sensing Symposium*, pp. 187-189.

Garbers, J., Niemann, M. and Mochol, M. 2006. "A Personalized Hotel Selection Engine," in *Proceedings of the third European Semantic Web Conference,* Budva, Montenegro.

Ghose, A., Ipeirotis, P. and Sundararajan, A. 2006. "The Dimensions of Reputation in Electronic Markets," NYU CeDER Working paper.

Ghose, A. 2009. "Internet Exchanges for Used Goods: An Empirical Analysis of Trade Patterns and Adverse Selection," *MIS Quarterly*, 33(2), pp. 263-291.

Ghose, A. and Ipeirotis, P.G. 2010. "Estimating The Helpfulness And Economic Impact Of Product Reviews: Mining Text And Reviewer Characteristics," Forthcoming, *IEEE Transactions On Knowledge And Data Engineering*.

Godes, D. and Mayzlin, D. 2004. "Using Online Conversations To Study Word Of Mouth Communication," *Marketing Science*. 23(4), pp. 545-560.

Hausman, J., Leonard G. and Zona, D. 1994. "Competitive Analysis With Differentiated Products," *Annales d'Economie Et De Statistique*. Vol. 34, pp. 159-180.

Hu, M. and Liu, B. 2004. "Mining And Summarizing Customer Reviews," *Proceedings Of The Tenth ACM SIGKDD International Conference On Knowledge Discovery And Data Mining*. pp. 168–177.

Hu, N., Pavlou, P.A. and Zhang, J. 2006. "Can Online Reviews Reveal A Product's True Quality? Empirical Findings And Analytical Modeling Of Online Word-Of-Mouth Communication," *Proceedings Of The Seventh ACM Conference On Electronic Commerce*, pp. 324-330.

Janiszewski, C. 1998. "The Influence of Display Characteristics on Visual Exploratory Search Behavior," *Journal of Consumer Research*, 25(3), pp. 290-301.

Jiang, Z. and Benbasat, I. 2007. "Investigating the Influence of the Functional Mechanisms of Online Product Presentations," *Information Systems Research* (18) 4, pp. 454-470.

Jiang, Z., Chan, J., Tan, B. and Chua. W. S. 2010. "Effects of Interactivity on Website Involvement and Purchase Intention," *Journal of the Association for Information Systems*. 11(1), pp. 34-59.

Kelley, C. T. 2003. "Solving Nonlinear Equations with Newton's Method," *SIAM*, Philadelphia, PA.

Knittel, C. R., and Metaxoglou, K. 2008. "Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings," Working Paper, U.C. Davis.

Lancaster, K. 1966. "A New Approach To Consumer Theory," *Journal Of Political Economy* (74), pp. 132-157.

Lancaster, K. 1971. *Consumer Demand: A New Approach*. Columbia University Press, NY.

Lavrenko, V. and Croft, W.B.. 2001. "Relevance based language models," in *Proceedings of the 24th annual international ACM SIGIR conferencee on Research and development in information retrieval,* pp. 120-127.

Lee, T. and Bradlow, E. 2007. "Automatic Construction Of Conjoint Attributes And Levels From Online Customer Reviews," University Of Pennsylvania, The Wharton School Working Paper.

Li, X. and Hitt, L. 2008. "Self-Selection and Information Role Of Online Product Reviews," *Information Systems Research* 19(4).

Liu, Y. 2006. "Word-Of-Mouth For Movies: Its Dynamics And Impact On Box Office Revenue*," Journal Of Marketing*, 70 (3), pp. 74-89.

Liu, B. Hu, M. and Cheng, J. 2005. "Opinion Observer: Analyzing And Comparing Opinions On The Web," *Proceedings Of The 14th International World Wide Web Conference* (WWW 2005). 342–351.

Lopatovska, I., 2007. "Decision making framework for thinking about information seeking behavior," in *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology*, 44.

Lopatovska, I. and Mokros, H. 2007. "Willingness to pay and experienced utility as measures of affective value in the information retrieval objects: users' accounts," *Information Processing and Management*, 44(1), 92-104.

Manning, C. and Schutze, H. 1999. *Foundations of Statistical Natural Language Processing*, The MIT Press.

Marchionini, Gary. 1997. *Information Seeking in Electronic Environments*. Cambridge University Press.

Mcfadden, D. 1973. "Conditional Logit Analysis of Quantitative Choice Behavior," *Frontiers In Econometrics*, Ed. By P. Zarambka. New York: Academic Press.

Mcfadden, D. 2001. "Economic Choices," *The American Economic Review,* 91(3). pp. 351-37.

Moe, W. 2009. "How Much Does A Good Product Rating Help A Bad Product? Modeling The Role Of Product Quality In The Relationship Between Online Consumer Ratings And Sales," Working Paper, University Of Maryland College Park.

Nair, H., Dube, J.P. and Chintagunta, P. 2005. "Accounting for Primary and Secondary Demand Effects with Aggregate Data," *Marketing Science*. 24(3), 444-460.

Nevo, A., 2000. "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9(4), 513-548.

Nevo, A., 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, Vol. 69, No. 2. pp. 307-342.

Nevo, A. and Rosen, A. 2008. "Identification with Imperfect Instruments," Working Paper, SSRN.

Nocedal, J. and Wright, S. J. 2006. *Numerical Optimization*. Springer, New York, NY.

Pang, B. and Lee, L. 2004. "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based On Minimum Cuts," in *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pp. 271-278.

Pashler, H.E. 1998. *The Psychology of Attention*. Cambridge, MA. MIT Press.

Petrin, A. 2002. "Quantifying The Benefits of New Products: The Case Of The Minivan," *Journal of Political Economy*.

Popescu, A. and Etzioni, O. 2005. "Extracting product features and opinions from reviews," *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pp. 339–346.

Rosen, S. 1974. "Hedonic Prices And Implicit Markets: Product Differentiation In Pure Competition," *Journal of Political Economy* (82:1), pp. 34-55.

Rust, R.T. and Chung, T.S. 2006. "Marketing Models of Service and Relationships," *Marketing Science*. 25 (6) 560-580.

Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H. and Jin, C. 2007. "Red Opal: Product-Feature Scoring From Reviews," *Proceedings of the 8th ACM conference on Electronic commerce* (EC'07). 182–191.

Shy, Oz. 1996. *Industrial Organization: Theory and Applications*, The MIT Press.

Snipes, R. L., Ingram, R. and Jiang, P. 2005. "Information search in health care decision-making: A study of word-of-mouth and internet information users," *Journal of Hospital Marketing & Public Relations*, 15(2), 33-53.

Song, M. 2007. "Measuring consumer welfare in the CPU market: an application of the pure-characteristics demand model," *RAND Journal of Economics*, 38(2), 429-446.

Song, M. 2010. "A Hybrid Discrete Choice Model of Differentiated Product Demand with an Application to Personal Computers," Simon School Working Paper No. FR 08-09. Available at SSRN: http://ssrn.com/abstract=1315271.

Steckel, J. and Vanhonacker, W. 1993. "Cross-Validating Regression Models in Marketing Research," *Marketing Science*, 12(4) pp. 415-427.

Villas-Boas, J. Miguel and Winer, Russell S. 1999. "Endogeneity in Brand Choice Models," *Management Science*, 45(10) 1324–1338.

White, P. 1998. "Site And Situation Determinants of Hotel Room Rates," Master's thesis.

Ying, Y., Feinberg, F. and Wedel, M. 2006. "Leveraging Missing Ratings To Improve Online Recommendation Systems," *Journal of Marketing Research*, 43(3) 355-365.